

**Record Number:** 13170  
**Author, Monographic:** Rasmussen, P. F.//Fagherazzi, L.//Duquette, R.//Legendre, P.//Perreault, L.  
**Author Role:**  
**Title, Monographic:** Détermination de séries mensuelles types pour le modèle Minerve - Phase 1  
**Translated Title:**  
**Reprint Status:**  
**Edition:**  
**Author, Subsidiary:**  
**Author Role:**  
**Place of Publication:** Québec  
**Publisher Name:** INRS-Eau  
**Date of Publication:** 1996  
**Original Publication Date:** Janvier 1996  
**Volume Identification:**  
**Extent of Work:** iii, 65  
**Packaging Method:** pages  
**Series Editor:**  
**Series Editor Role:**  
**Series Title:** INRS-Eau, rapport de recherche  
**Series Volume ID:** 480  
**Location/URL:**  
**ISBN:** 2-89146-420-6  
**Notes:** Rapport annuel 1996-1997  
**Abstract:** Rapport de la Chaire en hydrologie statistique.  
**Call Number:** R000480  
**Keywords:** rapport/ ok/ dl

**DÉTERMINATION DE SÉRIES  
MENSUELLES TYPES POUR LE  
MODÈLE MINERVE - PHASE 1**

**DÉTERMINATION DE SÉRIES MENSUELLES TYPES  
POUR LE MODÈLE MINERVE - PHASE 1**

Rapport préparé par

**Peter F. Rasmussen  
Laura Fagherazzi  
Robin Duquette  
Pierre Legendre  
Luc Perreault**

**Chaire en Hydrologie statistique  
Institut national de la recherche scientifique, INRS-Eau  
2800, rue Einstein, Case postale 7500, SAINTE-FOY (Québec), G1V 4C7**

**Rapport de recherche No R-480**

**Janvier 1996**



# Table des matières

1	Introduction.....	1
2	Description sommaire du modèle MINERVE.....	3
2.1	Problématique .....	3
2.2	Approche de résolution .....	4
2.3	Prise en compte de la stochasticité des apports .....	5
3	Génération de séries synthétiques.....	7
3.1	Introduction .....	7
3.2	Étapes de la modélisation stochastique.....	8
3.2.1	Choix de variables.....	8
3.2.2	Transformation des données mensuelles.....	8
3.2.3	Application d'un modèle stochastique.....	11
3.2.4	Validation d'un modèle stochastique.....	12
3.3	Modèles stochastiques de simulation.....	12
3.3.1	Modèles de désagrégation .....	13
3.3.2	Modèles PARMA.....	17
4	Classification des séries synthétiques.....	23
4.1	Introduction .....	23
4.2	Choix de séries agrégées représentatives.....	24
4.3	Revue de certaines caractéristiques statistiques de séries synthétiques.....	26
4.3.1	Caractéristiques reliées au volume.....	26
4.3.2	Caractéristiques reliées à la distribution .....	26
4.4	Analyse fréquentielle des caractéristiques statistiques.....	31
4.4.1	Analyse en composantes principales .....	32
4.4.2	Choix de scénarios .....	35
5	Exemple d'application: Grande Baleine.....	39
5.1	Description du système hydrique de la rivière Grande Baleine.....	39
5.2	Génération de séries synthétiques à l'aide du modèle SPIGOT.....	40
5.2.1	Application de SPIGOT .....	40
5.2.2	Validation .....	41
5.3	Identification de scénarios types .....	42
5.3.1	Séries agrégées.....	42
5.3.2	Sélection de critères .....	43
5.3.3	Résultats pour les séries d'apports intermédiaires agrégés.....	43
5.3.4	Résultats pour la série énergétique.....	49
5.4	Discussion des résultats .....	52
6	Conclusions .....	55
7	Bibliographie .....	57

# Liste des figures

3.1 Options de configurations dans SPIGOT.....	17
3.2 Corrélation d'ordre 1 pour le site I93891 du complexe Grande Baleine.....	19
4.1 Définitions des sommes partielles et du domaine de variation .....	28
4.2 Définitions des sommes partielles ajustées et du domaine de variation ajusté.....	29
4.3 Définition des propriétés d'une séquence .....	31
4.4 Étapes de la méthodologie proposée .....	38
5.1 Apport intermédiaires du complexe Grande Baleine .....	39
5.2 Position des critères dans l'espace des trois premiers facteurs (apport intermédiaire).....	47
5.3 Position des séries synthétiques d'apports intermédiaires agrégés dans l'espace des deux premiers facteurs .....	48
5.4 Position des critères dans l'espace de deux facteurs (série énergétique).....	51
5.5 Position des séries énergétiques synthétiques dans l'espace des deux premiers facteurs .....	52
5.6 Somme partielles ajustées avec la moyenne à long terme pour neuf scénarios types .....	54

# Liste des tableaux

3.1	Caractéristiques statistiques que doit reproduire un modèle stochastique.....	13
5.1	Caractéristiques statistiques des séries d'apports intermédiaires historiques et synthétiques pour le complexe Grande Baleine.....	41
5.2	Hauteur de chute pour les centrales du complexe Grande Baleine.....	42
5.3	Caractéristiques statistiques des critères pour les séries synthétiques d'apports intermédiaires.....	44
5.4	Matrice de corrélation des critères (séries d'apports intermédiaires).....	44
5.5	Coordonnées des critères sur les axes factoriels (séries d'apports intermédiaires).....	46
5.6	Coefficients pour calculer les facteurs (séries d'apports intermédiaires).....	46
5.7	Caractéristiques statistiques des critères pour les séries énergétiques synthétiques.....	49
5.8	Matrice de corrélation des critères (séries énergétiques).....	50
5.9	Coordonnées des critères sur les axes factoriels (séries énergétiques). a) tous les critères; b) tous les critères moins l'écart-type S.....	50
5.10	Coefficients pour calculer les facteurs (séries énergétiques).....	51

# 1 Introduction

---

Le dimensionnement d'un nouvel aménagement hydroélectrique est effectué généralement en plusieurs étapes dont chacune a ses propres objectifs. Ainsi, l'étude préliminaire a pour but d'évaluer les schémas d'aménagement avantageux et d'établir la préfaisabilité technique, économique et environnementale d'un projet. L'étude d'avant-projet a comme objectif d'effectuer une évaluation méthodique des variantes d'aménagement et de confirmer la faisabilité technique, économique et environnementale du projet. Le but est, à l'aide d'analyses techniques, statistiques et économiques, de comparer les bénéfices et les coûts entre les diverses variantes de l'étude. La maximisation du rapport bénéfices-coûts est habituellement recherchée. Les bénéfices sont obtenus à partir des calculs de la production effectués à l'aide des variables de débit et de hauteur de chute. Les coûts sont évalués sur la base de ces mêmes variables. Les données requises pour l'analyse bénéfices-coûts sont sensiblement les mêmes à chaque étape, seul le niveau de détail diffère d'une étape à l'autre.

Ce qui distingue de façon particulière les différentes étapes d'une étude, ce sont les méthodes d'analyse utilisées. Il existe deux approches de base distinctes pour la solution du problème de maximisation du rapport bénéfice/coût: la simulation et l'optimisation. Ainsi, pour une étude préliminaire, ce sont les modèles mathématiques d'optimisation qui sont utilisés, alors que pour les avant-projets, les modèles de simulations sont employés.

La simulation est basée sur la méthode essai et erreur pour la recherche d'une solution près de l'optimum. La difficulté de cette procédure est qu'il existe un vaste champ de solutions possibles difficiles à cerner par cette approche. Par contre, les modèles de simulations peuvent résoudre des problèmes non linéaires avec un niveau de détail très précis, ce qui est rarement possible avec les modèles d'optimisation. L'avantage du modèle d'optimisation réside dans le fait qu'il peut réduire considérablement le nombre de variantes et converger plus rapidement vers des solutions près de l'optimum. La principale difficulté avec ce type de modèle est la résolution d'un système d'équations non linéaires.

MINERVE est un modèle d'optimisation qui permet de résoudre des systèmes d'équations non linéaires. Toutefois, le modèle est déterministe et il utilise des séries d'apports connues à l'avance. Généralement, on ne considère que les séries d'apports historiques dans les modèles déterministes. L'aménagement identifié comme étant le plus apte à répondre à une certaine demande est basé sur ces séries. On se pose alors la question si l'aménagement sélectionné est toujours optimal lorsque les apports futurs sont très différents de ceux observés dans le passé. L'analyse stochastique des apports peut aider à répondre à cette question. En effet, l'utilisation dans le modèle MINERVE de scénarios d'apports synthétiques vise à évaluer la rentabilité de différents schémas d'aménagement sur une base plus éclairée.

La complexité actuelle du modèle MINERVE empêche l'introduction de l'aspect stochastique dans le modèle lui-même. Une approche novatrice doit être utilisée; ce travail s'inscrit dans le cadre d'un projet de recherche ayant pour but le développement d'une méthodologie qui tient compte du caractère aléatoire des apports naturels dans un contexte d'optimisation non linéaire.

## **2 Description sommaire du modèle MINERVE**

### **2.1 Problématique**

L'aménagement optimal d'un bassin pour fin de production d'électricité à Hydro-Québec consiste:

- à recenser d'abord les sites où il sera possible de construire des centrales et/ou des réservoirs;
- à déterminer les différentes tailles que pourront avoir ces installations avec les coûts de construction associés; et
- à déterminer ensuite un schéma d'aménagement, c'est-à-dire un choix de centrales et de réservoirs à construire et leurs tailles, de façon à satisfaire une demande prioritaire en électricité à un coût minimal.

Le problème lié à la troisième étape, c'est-à-dire le choix d'un schéma d'aménagement optimal, peut s'exprimer comme un problème de programmation non linéaire en variables mixtes. Les variables entières représentent les réservoirs, les centrales et les dérivations à construire alors que les variables continues représentent la taille des diverses installations (la hauteur des barrages, la capacité de turbiner des centrales, etc.). Le problème est non linéaire principalement parce que les fonctions de coûts de construction des installations sont généralement non linéaires et que la production d'une centrale est une fonction non séparable de la hauteur de chute et de l'eau turbinée. Ainsi, lorsque le choix des installations est fixé, le problème résultant est non convexe et l'emploi d'une méthode standard pour résoudre le problème de programmation non linéaire n'assure pas une solution optimale globale par rapport à tout le domaine réalisable. En effet, la solution trouvée ne sera optimale que dans un certain voisinage.

Ce problème est présentement résolu à Hydro-Québec par un programme appelé MINERVE (cf. Duquette et Lafond, 1992) qui est une version améliorée d'ATHENA. Le logiciel ATHENA a permis de réaliser une économie de 50 M\$ sur l'aménagement d'une seule rivière. Ces deux programmes ont été développés à Hydro-Québec.

## 2.2 Approche de résolution

L'approche de résolution utilisée présentement par MINERVE pour déterminer le schéma d'un aménagement optimal d'un bassin est composée des deux étapes suivantes:

1. La première étape élimine rapidement certains choix d'installation pour lesquels la demande prioritaire en énergie ne pourrait être satisfaite. Par exemple, on peut écarter *a priori* tous les choix d'installation dont la capacité installée maximale n'est pas suffisante pour satisfaire les périodes de pointe. Cette étape est nécessaire parce que le nombre de choix possibles d'installation peut être aussi grand que 124 416.
2. La seconde étape trouve, pour chaque choix d'installation retenu à l'étape précédente, les dimensions des installations de manière à minimiser la somme des coûts de construction, plus la somme des coûts de délestage moins les revenus de vente d'énergie excédentaire, ceci sujet à des contraintes de design et d'exploitation. Les contraintes sont essentiellement les suivantes:
  - Pour chaque réservoir, le niveau maximal d'exploitation doit être inférieur ou égal au niveau maximal de restitution des installations situées en amont immédiat.
  - Pour chaque centrale, le niveau minimal de restitution doit être supérieur ou égal au niveau minimal d'exploitation du réservoir situé en aval immédiat.
  - Pour chaque réservoir lié à une centrale, la zone comprise entre les niveaux minimal et maximal d'exploitation doit être inférieure ou égale à une fraction de la hauteur de chute maximale associée à la centrale en aval immédiat.
  - Pour chaque réservoir, la zone de marnage est bornée supérieurement par une constante pour des raisons écologiques.
  - Pour chaque centrale, la hauteur de chute maximale est égale à la différence entre le niveau minimal de restitution et le niveau maximal d'exploitation du réservoir en amont immédiat.
  - Pour chaque réservoir et chaque période, une contrainte de conservation de l'eau.
  - Pour chaque période, une contrainte de demande.

Mathématiquement, le problème résolu à la deuxième étape s'exprime comme suit:

$$\min_{x \in X} \text{Coût}(x) + \sum_{k \in P} c D_k - r E_k \quad (1)$$

sujet à

$$G(x) \leq 0$$

$$S_{i,k} = S_{i,k-1} + \tau_k \sum_j Q_{j,k} + \tau_k \sum_i R_{i,k} + \tau_k a_{i,k} - \tau_k R_{i,k} - \tau_k Q_{j,k}$$

$$S_{i,0} = S_{i,P1}$$

$$\phi_k \sum_j \eta_j Q_{j,k} H_{j,k} + D_k - E_k = d_k$$

où

- La fonction Coût(x) représente le coût de construction (M\$); elle est non linéaire et non convexe.
- $D_k$  et  $E_k$  représentent respectivement le déficit de la période k (GWh) et le surplus produit lors de la période k (GWh); c et r sont les coefficients associés respectivement à  $D_k$  et  $E_k$  (M\$/GWh).
- Le premier ensemble de contraintes (i.e.,  $G(x) \leq 0$ ) représente les restrictions sur les dimensions des installations, quelques contraintes d'exploitation et les bornes sur les variables.
- Le deuxième ensemble de contraintes représente les contraintes de conservation de l'eau où pour la période k,  $S_{i,k}$  est le volume du réservoir ( $hm^3$ ),  $R_{i,k}$  est le débit déversé du réservoir i ( $m^3/s$ ),  $Q_{j,k}$  est l'eau turbinée à la centrale j ( $m^3/s$ ) et  $a_{i,k}$  est l'apport naturel en eau ( $m^3/s$ ).  $\tau_k$  est un facteur de conversion.
- Le troisième ensemble de contraintes impose pour chaque réservoir que le contenu initial soit égal au contenu final. Ces contraintes assurent que l'électricité ne sera produite qu'à partir des apports naturels.
- Finalement, le dernier ensemble de contraintes exige de satisfaire la demande prioritaire à chaque période.  $D_k$  représente le déficit de la période k (cette variable est définie uniquement pour des raisons de faisabilité) et  $E_k$  représente les surplus de la période k.

### 2.3 Prise en compte de la stochasticité des apports

Le modèle (1) est déterministe et ne tient pas compte du caractère aléatoire des apports naturels. En effet, l'aménagement optimal d'un bassin est déterminé uniquement à partir d'un scénario d'apports : les apports observés historiquement. Une façon de rendre le modèle moins déterministe

et de prendre en compte d'une certaine manière la stochasticité des apports est de considérer plus d'un scénario à la fois. On dira alors que l'aménagement déterminé sera optimal en moyenne.

Prendre en compte plus d'un scénario exige de modifier la fonction objectif définie en (1) ainsi que les contraintes de conservation de l'eau. Plus précisément, soit  $S$  l'ensemble des indices des scénarios et soit  $\xi_s = \text{Coût}(x) + \sum_{k \in P} [c D_{k,s} - r E_{k,s}]$  pour tout  $s \in S$ . La fonction objectif devient maintenant une *variable* aléatoire qui prend la valeur  $\xi_s$  avec une probabilité d'occurrence  $p_s$  ( $\sum_{s \in S} p_s = 1$ ). Le modèle mathématique résultant est:

$$\min_{x \in X'} \sum_S p_s \xi_s + \lambda \sum_S p_s (\xi_s - \sum_{S'} p_{S'} \xi_{S'})^2$$

sujet à

$$G'(x) \leq 0$$

$$S_{i,k,s} = S_{i,k-1,s} + \tau_k \sum_j Q_{j,k,s} + \tau_k \sum_i R_{i,k,s} + \tau_k a_{i,k,s} - \tau_k R_{i,k,s} - \tau_k Q_{j,k,s}$$

$$S_{i,0} = S_{i,0,s} = S_{i,|P|,s} = S_{i,|P|}$$

$$\varphi_k \sum_j \eta_j Q_{j,k,s} H_{j,k,s} + D_{k,s} - E_{k,s} = d_k$$

Le premier terme dans la fonction objectif,  $\sum_{s \in S} p_s \xi_s$ , correspond, en optimisation stochastique, à la valeur moyenne de la fonction objectif définie en (1). Le second terme permet d'obtenir une solution robuste, c'est-à-dire une solution qui n'est pas trop sensible aux différents scénarios. On notera que nous avons maintenant une contrainte de conservation de l'eau par réservoir, par période et par scénario. À cause de cela, on ne pourra pas considérer une grande quantité de scénarios.

L'objectif du présent document est donc de définir une méthode pour générer une quantité raisonnable de scénarios qui seront utilisés par la suite pour déterminer l'aménagement optimal d'un bassin.

## **3 Génération de séries synthétiques**

---

### **3.1 Introduction**

Tel que décrit dans la section précédente, l'application de la nouvelle version du logiciel MINERVE vise à tenir compte du caractère aléatoire des apports en considérant plusieurs scénarios. Il est important que ces scénarios soient réalistes du point de vue physique et statistique, c'est-à-dire qu'ils possèdent des caractéristiques que l'on retrouverait dans les séries d'apports futurs.

Afin d'obtenir des scénarios d'apports réalistes qui tiennent compte de caractéristiques complexes, en particulier la cohérence temporelle et spatiale des apports, on doit effectuer une modélisation stochastique. Les hydrologues font souvent appel à des modèles de simulation stochastique pour étudier différents aspects d'un système hydrique. À titre d'exemple, on peut mentionner l'évaluation de la gestion des réservoirs, de la production d'hydroélectricité et de la fiabilité des ouvrages hydrauliques. La philosophie souvent adoptée est que les séries historiques d'apports d'un système hydrique représentent un scénario moyen. On cherche alors à construire un modèle de simulation stochastique de façon à ce qu'il reproduise en moyenne certaines caractéristiques de séries observées. On peut ainsi étudier la variation des scénarios simulés autour de la moyenne. Bien que ce principe soit courant, il n'est pas toujours approprié. Ceci est particulièrement vrai dans le cas de simulation de scénarios futurs en présence de non-stationnarité (changement de moyenne par exemple). Dans ces cas, il faut être prudent et accorder un intérêt particulier à la nature des non-stationnarités car une mauvaise modélisation de ces phénomènes peut avoir un impact très négatif sur les conclusions tirées des scénarios simulés. Il est hors du cadre de cette étude d'analyser les différents types de non-stationnarité et leur modélisation. Nous nous limitons ici à souligner l'importance d'une bonne modélisation des non-stationnarités et considérons dans ce qui suit uniquement le cas où les séries annuelles d'apports sont stationnaires.

Dans ce chapitre, les différentes étapes de la modélisation stochastique sont décrites. Nous présentons également une brève revue des modèles stochastiques les plus utilisés en hydrologie.

## 3.2 Étapes de la modélisation stochastique

### 3.2.1 Choix de variables

La première étape d'une modélisation stochastique consiste à identifier les variables à représenter. Essentiellement, il s'agit ici de déterminer le type d'agrégation des apports (temporel et/ou spatial). Le choix d'agréger les apports dans le temps et/ou dans l'espace doit être fait en fonction du problème étudié et selon l'utilisation éventuelle des séries générées. Souvent, les séries générées sont utilisées dans d'autres logiciels et doivent alors respecter certaines contraintes. Le logiciel MINERVE considère des apports mensuels. Aussi, la modélisation stochastique doit-elle être basée sur l'étude des apports mensuels observés.

En ce qui a trait à l'agrégation spatiale des données, le problème considéré conditionne également souvent le choix de variables. Dans le cas de la détermination de l'aménagement optimal d'un nouveau bassin, tel que formulé dans le modèle MINERVE, on a généralement une idée *a priori* de l'emplacement des centrales et des réservoirs. On doit alors s'intéresser aux apports intermédiaires aux différents réservoirs.

On doit noter, qu'assez fréquemment, le nombre de données à certains sites d'un système hydrique est très restreint. Ceci est particulièrement vrai pour les bassins où un aménagement hydroélectrique n'était pas prévu il y a vingt ans. Pour le calage adéquat d'un modèle stochastique, on doit disposer d'un certain nombre d'années d'observations concomitantes. Il est donc souvent nécessaire de reconstituer les données aux sites cibles où les données sont insuffisantes ou inexistantes. Il importe de souligner que la qualité des données utilisées pour le calage du modèle stochastique diminue lorsqu'elles sont obtenues par reconstitution.

En résumé, pour la modélisation stochastique des systèmes hydriques dans un optique d'aménagement optimal d'un nouveau bassin, on considère les apports mensuels intermédiaires aux différents réservoirs du système.

### 3.2.2 Transformation des données mensuelles

La majorité des modèles stochastiques de simulation repose sur l'hypothèse que les données saisonnières (mensuelles) sont distribuées selon une loi normale, alors qu'en réalité ce n'est pas toujours le cas. Bien qu'il existe certains modèles permettant de modéliser des données non normales (Fernandez et Salas, 1986, 1990; Sim, 1987), l'approche habituelle consiste à effectuer

une transformation des données saisonnières. En général, cette procédure facilite la modélisation et permet de considérer une gamme plus large de modèles. Les transformations peuvent avoir un impact important sur les simulations, notamment si l'on s'intéresse aux valeurs extrêmes, c'est-à-dire à l'extrapolation de la distribution marginale des données. Il est alors important d'effectuer une analyse fréquentielle détaillée des données saisonnières afin de choisir les transformations les plus adéquates. Dans ce qui suit, nous présentons les types de transformations les plus classiques.

### *Transformation LN2*

La transformation log-normale à deux paramètres (LN2) est très souvent utilisée. Cette transformation est appropriée si les données saisonnières ( $Q$ ) suivent une loi LN2 dont la fonction de densité de probabilité (fdp) est donnée par:

$$f(q) = \frac{1}{q\beta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\log q - \alpha}{\beta}\right]^2\right) \quad (2)$$

En effet, si  $Q$  suit une loi LN2 alors la nouvelle variable  $X = \log(Q)$  est distribuée selon une loi normale de moyenne  $\alpha$  et de variance  $\beta^2$ .

L'estimation des paramètres  $\alpha$  et  $\beta$  peut être effectuée par la méthode des moments ou la méthode du maximum de vraisemblance. Dans le cas de la méthode des moments, les estimateurs de  $\alpha$  et  $\beta$  sont donnés par:

$$\hat{\alpha} = \log \bar{q} - \frac{1}{2} \log\left(\frac{s_q^2}{\bar{q}^2} + 1\right) \quad \text{et} \quad \hat{\beta}^2 = \log\left(\frac{s_q^2}{\bar{q}^2} + 1\right) \quad (3)$$

Les estimateurs du maximum de vraisemblance sont donnés par:

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \log q_i \quad \text{et} \quad \hat{\beta} = \frac{1}{n} \sum_{i=1}^n (\log q_i - \hat{\alpha})^2 \quad (4)$$

### *Transformation LN3*

La transformation log-normale à trois paramètres est particulièrement utile lorsqu'il y a des valeurs négatives ou nulles dans l'échantillon. La f.d.p. de la loi LN3 est donnée par l'équation suivante:

$$f(q) = \frac{1}{(q-m)\beta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\log(q-m) - \alpha}{\beta}\right]^2\right)$$

La transformation  $X = \log(Q - m)$  permet d'obtenir des données  $X$  distribuées selon une loi normale de moyenne  $\alpha$  et de variance  $\beta^2$  si la variable originale  $Q$  suit une loi LN3.

Les méthodes des moments et du maximum de vraisemblance sont les plus employées pour estimer les paramètres  $\alpha$ ,  $\beta$  et  $m$ . Dans le cas de la méthode des moments, les estimateurs de  $\alpha$ ,  $\beta$  et  $m$  sont:

$$\hat{\alpha} = \frac{s_q}{\sqrt{\hat{\phi}(\hat{\phi} - 1)}}, \quad \hat{\beta} = \sqrt{\log(\hat{\phi})} \quad \text{et} \quad \hat{m} = \bar{q} - \frac{s_q}{\sqrt{\hat{\phi} - 1}}$$

$$\text{où } \hat{\phi} = \left[ 1 + \frac{C_s}{2} (C_s + \sqrt{4 + C_s^2}) \right]^{1/3} + \left[ 1 + \frac{C_s}{2} (C_s - \sqrt{4 + C_s^2}) \right]^{1/3} - 1$$

Dans le cas de la méthode du maximum de vraisemblance, une méthode numérique (Newton-Raphson, par exemple) doit être utilisée pour déterminer les estimateurs.

#### *Transformation gamma*

Il n'existe pas de méthode exacte permettant de transformer une variable suivant une loi gamma en une variable distribuée selon une loi normale. Toutefois, une transformation approximative a été proposée par Wilson et Hilferty (1931). Elle s'exprime de la façon suivante:

$$X = \frac{6}{C_s} \left[ \left( \frac{C_s(Q - \bar{Q})}{2s_Q} + 1 \right)^{1/3} + \frac{C_s^2}{36} - 1 \right] \quad \text{où} \quad \begin{array}{l} Q < -2s_Q / C_s + \bar{Q} \quad \text{si } C_s < 0 \\ Q > -2s_Q / C_s + \bar{Q} \quad \text{si } C_s > 0 \end{array}$$

Cette transformation est valide pour  $|C_s| \leq 3$ .

#### *Transformation de Box-Cox*

Une méthode plus générale permettant de normaliser les données a été proposée par Box et Cox (1964). Ces auteurs ont étudié une généralisation de la famille des transformations monotones des fonctions puissances. Cette famille est indexée par un paramètre  $\lambda$  et elle se définit comme suit:

$$X^{(\lambda)} = \begin{cases} (Q^\lambda - 1) / \lambda & \lambda \neq 0 \\ \log(Q) & \lambda = 0 \end{cases} \quad \text{où } Q > 0$$

Cette famille contient comme cas particulier les transformations usuelles  $\log(Q)$ ,  $\sqrt{Q}$ ,  $1/Q$ ,  $Q^2$ . Le paramètre  $\lambda$  peut être estimé par la méthode du maximum de vraisemblance. Une méthode numérique doit être utilisée pour déterminer la valeur de  $\lambda$  qui minimise la quantité:

$$t = \sum_{i=1}^n (z_i - \bar{z})^2$$

$$\text{où } z_i = x_i^{(\lambda)} / \hat{q}^{\lambda-1} \text{ avec } \hat{q} = \sqrt[n]{\prod_{i=1}^n q_i}$$

### *Statistique de Filliben*

La statistique de Filliben (1975) permet de mesurer la linéarité des données transformées sur un papier de probabilité normale. Elle est donnée par:

$$r = \frac{\sum_{i=1}^n (x_{(i)} - \bar{x})(M_i - \bar{M})}{\left[ \sum_{i=1}^n (x_{(i)} - \bar{x})^2 \sum_{i=1}^n (M_i - \bar{M})^2 \right]^{1/2}}$$

où  $x_{(i)}$  est la  $i$ ème observation de l'échantillon classé en ordre croissant;  
 $M_i$  est l'espérance mathématique de  $x_{(i)}$ .

La valeur de  $M_i$  est déduite de la probabilité empirique  $p_i$  attribuée à l'élément  $x_{(i)}$ . La statistique de Filliben est en réalité un coefficient de corrélation qui mesure la linéarité entre les valeurs  $x_{(i)}$  et  $M_i$ . Plus la statistique de Filliben est voisine de 1 plus la transformation est adéquate.

### **3.2.3 Application d'un modèle stochastique**

Il existe plusieurs modèles stochastiques que l'on peut utiliser pour générer des séries synthétiques. Chacun possède ses propres avantages et lacunes. Bien que certaines comparaisons aient été faites (par exemple Acres International Limited, 1990), aucun modèle n'apparaît systématiquement supérieur aux autres. Il n'est donc pas possible de recommander l'utilisation générale d'un modèle particulier. Aussi, est-il souhaitable de considérer plusieurs modèles et de comparer leur performance selon des critères préalablement établis.

Le choix du modèle stochastique à utiliser doit, bien entendu, être fait en fonction du problème étudié. Par exemple, dans le cas d'une génération d'apports saisonniers, il faut éliminer les modèles

stationnaires puisque les processus mensuels ne sont généralement pas stationnaires. De plus, si on s'intéresse à des réservoirs interannuels, on doit considérer uniquement des modèles qui permettent de modéliser adéquatement la persistance interannuelle.

Ce rapport n'envisage pas une revue exhaustive de tous les modèles stochastiques utilisés en hydrologie. Dans la section 3.3, nous décrivons deux classes de modèles que nous considérons particulièrement adéquates pour la génération de séries saisonnières ayant une persistance interannuelle et qui répondent bien aux objectifs de la présente étude.

### **3.2.4 Validation d'un modèle stochastique**

La validation est une étape très importante de la modélisation stochastique. Dans un premier temps, la validation sert à détecter de véritables erreurs dans l'estimation des paramètres du modèle ou dans la représentation du système hydrique (Stedinger et Taylor, 1982). Ensuite, elle permet de vérifier que le modèle génère des séries synthétiques ayant des caractéristiques hydrologiques réalistes vis-à-vis des données utilisées pour l'estimation du modèle (séries historiques). Pour ce faire, on s'appuie sur le principe de "la loi de la moyenne", c'est-à-dire que l'on exige qu'en moyenne les différentes caractéristiques statistiques des séries générées soient proches de celles des séries observées. En pratique, on se limite souvent à quelques caractéristiques statistiques de base. Le tableau 3.1 en donne une liste que l'on pourrait considérer. Soulignons que cette liste n'est pas exhaustive. Par exemple, dans le cas où la série observée est nonstationnaire (tendances ou sauts de la moyenne), il faut adapter les critères en conséquence.

De façon générale, on doit appliquer plusieurs modèles et choisir le plus adéquat. L'étape de validation a pour objectif non seulement l'évaluation de chaque modèle, mais également leur comparaison. Le modèle qui satisfait le mieux aux différents critères jugés importants doit être privilégié. En pratique, puisqu'il y a toujours plusieurs critères à considérer, on adopte une méthode de prise de décision basée sur l'analyse multicritère.

## **3.3 Modèles stochastiques de simulation**

Dans cette section, nous donnerons une brève description de quelques modèles stochastiques fréquemment utilisés. Notre objectif n'est pas de fournir tous les détails techniques de chacun des

**Tableau 3.1** Caractéristiques statistiques que doit reproduire un modèle stochastique

Séries mensuelles	Séries annuelles
Moyenne périodique	Moyenne
Variance périodique	Variance
Asymétrie périodique	Asymétrie
Autocorrélation périodique	Autocorrélation
Distribution du maximum annuel	
Distribution du minimum annuel	
Corrélation croisée périodique	Corrélation croisée
	Moyenne de la séquence maximale
	Domaine de variation

modèles, mais plutôt de donner une vue globale sur les techniques de simulation les plus utilisées en indiquant les principales références qui y sont associées.

On distingue souvent deux types de modèles: les modèles directs et les modèles indirects. Le premier type de modèle consiste à simuler directement les données saisonnières. Dans le cas des modèles indirects, on simule d'abord les données annuelles que l'on désagrège ensuite entre les différentes saisons (mois).

### **3.3.1 Modèles de désagrégation**

Valencia et Schaake (1973) ont présenté une approche de simulation appelée "désagrégation". Les modèles de désagrégation sont de type indirect. En effet, le principe de base de ces modèles est de simuler les apports agrégés temporellement et éventuellement spatialement. Plus particulièrement, les apports annuels sont simulés à l'aide d'un modèle stationnaire classique et désagrégés ensuite en des apports saisonniers (mensuels) tout en préservant les apports annuels (somme) et certaines caractéristiques liées à la structure de corrélation des données. Cette approche permet de modéliser explicitement la persistance interannuelle ce qui est généralement important dans le cas de la gestion des grands réservoirs et dans le cas de l'aménagement d'un nouveau bassin. En raison de leur simplicité, les modèles de désagrégation ont été largement utilisés en pratique.

Le modèle présenté par Valencia et Schaake (1973) (VS) repose sur l'hypothèse que les apports annuels et mensuels sont distribués selon une loi normale. Ainsi, en pratique, il est généralement nécessaire de transformer les données observées pour en permettre l'utilisation. Si l'on considère le cas d'un seul site, le modèle VS s'écrit de la manière suivante:

$$\mathbf{X}_v = \mathbf{A}Z_v + \mathbf{B}\boldsymbol{\varepsilon}_v \quad (5)$$

où  $\mathbf{X}_v = (X_{v,1}, X_{v,2}, \dots, X_{v,12})^T$  est un vecteur d'apports mensuels pour l'année  $v$ ,  $Z_v$  est l'apport annuel de l'année  $v$ , simulé préalablement à l'aide d'un modèle (stationnaire) quelconque,  $\boldsymbol{\varepsilon}_v$  est un vecteur de douze variables normales indépendantes, et  $\mathbf{A}$  et  $\mathbf{B}$  sont des matrices de paramètres de dimension  $[12 \times 1]$  et  $[12 \times 12]$ . On suppose, sans perte de généralité, que toutes les variables sont de moyenne nulle. Valencia et Schaake (1973) ont démontré que les matrices de paramètres peuvent être estimées à l'aide de la méthode des moments en résolvant le système d'équations suivant:

$$\mathbf{A} = \mathbf{E}[\mathbf{X}_v Z_v] \mathbf{E}[Z_v^2]^{-1} \quad (6)$$

$$\mathbf{B}\mathbf{B}^T = \mathbf{E}[\mathbf{X}_v \mathbf{X}_v^T] - \mathbf{A}\mathbf{E}[Z_v^2]\mathbf{A}^T \quad (7)$$

La matrice  $\mathbf{B}\mathbf{B}^T$  est symétrique, et  $\mathbf{B}$  peut être déterminée, par exemple, à l'aide de la décomposition de Cholesky (Bras et Rodríguez-Iturbe, 1985). Un modèle dont les estimateurs sont obtenus par la méthode des moments préserve explicitement les moments  $\mathbf{E}[\mathbf{X}_v Z_v]$  et  $\mathbf{E}[\mathbf{X}_v \mathbf{X}_v^T]$ , c'est-à-dire les corrélations entre les apports mensuels et l'apport annuel, ainsi que la corrélation entre les apports de tous les mois d'une année donnée. On peut également montrer qu'une relation du type  $Z_v = \mathbf{C}\mathbf{X}_v$  sera préservée par le modèle (Valencia et Schaake, 1973). Plus particulièrement, si l'on considère  $\mathbf{C} = (1, 1, \dots, 1)$ , on notera que le modèle VS préserve les apports annuels totaux (la somme des apports mensuels simulés est égale aux apports annuels).

Le modèle VS ne préserve pas explicitement la corrélation entre les apports mensuels de deux années consécutives. Plusieurs auteurs ont tenté de combler cette lacune dont Mejia et Rousselle (1976), Lane (1979) et Stedinger et Vogel (1984). Cependant, en pratique, on se limite souvent à rechercher le mois de l'année ayant la plus faible corrélation avec le mois qui précède. Au Canada, on considère souvent le mois d'octobre comme le début de l'année hydrologique.

Le modèle VS peut facilement être généralisé au cas multivarié ( $k$  sites). Il suffit de définir le vecteur  $\mathbf{X}_v$  de la manière suivante:

$$\mathbf{X}_v = (\mathbf{X}_{v,1}^{(1)}, \mathbf{X}_{v,2}^{(1)}, \dots, \mathbf{X}_{v,12}^{(1)}, \mathbf{X}_{v,1}^{(2)}, \dots, \mathbf{X}_{v,12}^{(2)}, \dots, \mathbf{X}_{v,1}^{(k)}, \dots, \mathbf{X}_{v,12}^{(k)})^T \quad (8)$$

Le vecteur  $\mathbf{X}_v$  contient alors tous les apports mensuels à tous les sites. Les dimensions des matrices de paramètres changent en conséquence. On peut soit générer un seul apport annuel agrégé qui sera ensuite désagrégé en des apports mensuels aux  $k$  sites, ou soit appliquer un modèle annuel multivarié. La dernière approche conduit à la génération d'un vecteur  $\mathbf{Z}_v$  qui contient les apports annuels aux  $k$  sites. L'estimation des paramètres est effectuée de la même façon que dans le cas univarié. Le modèle VS multivarié préserve toutes les corrélations entre les apports mensuels aux  $k$  sites ainsi que la corrélation entre les apports mensuels et l'apport annuel pour une année donnée. Mentionnons que la préservation de toutes ces caractéristiques nécessite un modèle possédant un nombre très élevé de paramètres.

Lors de l'identification d'un modèle statistique/stochastique on doit chercher à respecter le principe de *parcimonie*, c'est-à-dire l'utilisation du modèle le plus simple représentant de façon adéquate le phénomène étudié. Le modèle VS dans sa forme originale ne respecte pas ce principe. Le grand nombre de paramètres et la conservation de toute la structure de corrélation des données observées laissent peu de degrés de liberté pour la simulation. Deux approches ont été proposées pour pallier ce problème de surparamétrisation du modèle VS:

- les versions condensées du modèle VS; et
- la désagrégation par étapes (*staged disaggregation*).

Les modèles condensés ont pour but de diminuer le nombre de paramètres en réduisant le nombre de corrélations temporelles préservées explicitement. La désagrégation par étapes cherche plutôt à réduire le nombre de paramètres en simplifiant la structure de corrélation spatiale.

#### *Modèle condensé*

Lane (1979) a été le premier à proposer un modèle de désagrégation condensé. En considérant le cas d'un seul site, Stedinger et al. (1985b) montrent que la matrice  $\mathbf{B}$  est une matrice triangulaire inférieure qui peut être obtenue par exemple par la décomposition de Choleski. Le modèle VS original peut ensuite s'écrire comme:

$$\mathbf{X}_{v,1} = \alpha_1 \mathbf{Z}_v + \beta_1 \varepsilon_{v,1} \quad (9a)$$

$$\mathbf{X}_{v,2} = \alpha_2 \mathbf{Z}_v + \gamma_{2,1} \mathbf{X}_{v,1} + \beta_2 \varepsilon_{v,2} \quad (9b)$$

$$\mathbf{X}_{v,\tau} = \alpha_\tau \mathbf{Z}_v + \sum_{i=1}^{\tau-1} \gamma_{\tau,i} \mathbf{X}_{v,i} + \beta_\tau \varepsilon_{v,\tau} \quad \tau = 3, 4, \dots, 12 \quad (9c)$$

Cette formulation fait clairement ressortir la relation linéaire entre l'apport du mois  $\tau$  et les apports des mois précédents de la même année. La relation entre l'apport du mois  $\tau$  et l'apport annuel est également bien apparente. Les paramètres  $\alpha_\tau$ ,  $\beta_v$  et  $\gamma_{\tau,i}$  peuvent être obtenus à partir des matrices **A** et **B** définies précédemment. Pour réduire le nombre de paramètres dans le modèle VS, Stedinger et al. (1985b) proposent de remplacer l'équation 9c par:

$$X_{v,\tau} = \alpha_\tau Z_v + \gamma_\tau X_{v,\tau-1} + \delta_v \sum_{i=1}^{\tau-1} X_{v,i} + \beta_\tau \varepsilon_{v,\tau} \quad \tau = 3, 4, \dots, 12 \quad (10)$$

Soulignons que notre notation est légèrement différente de celle de Stedinger et al. (1985b). Dans cette équation, seule la relation entre les apports de deux mois consécutifs est explicite. Bien entendu, l'apport du mois  $\tau$  est toujours relié à l'apport annuel et un nouveau terme a été introduit, l'apport agrégé des  $\tau-1$  mois qui précèdent. Toutefois, cette formulation du modèle de désagrégation ne garantit pas que la somme des apports mensuels corresponde à l'apport annuel. Stedinger et al. (1985b) proposent différents facteurs de correction qui permettent d'obtenir de bons résultats. Notons d'ailleurs que lorsque  $X_v$  est constitué de données transformées, l'agrégation des apports mensuels réels en apports annuels n'est généralement pas exacte. En raison de la régression de  $X_{v,\tau}$  sur  $X_{v,\tau-1}$ , on peut dire que ce modèle est autorégressif d'ordre 1 avec des paramètres périodiques.

Le modèle VS condensé contient beaucoup moins de paramètres que le modèle VS original. Cette approche est incorporée dans le logiciel SPIGOT qui est fréquemment utilisé par Hydro-Québec (Grygier and Stedinger, 1990, 1991).

### *Désagrégation par étapes*

Le principe de la désagrégation par étapes est bien illustré par un exemple. Pour un système hydrique comprenant plusieurs sites, on peut d'abord générer l'apport annuel pour tout le système.

Ensuite, on peut désagréger l'apport total annuel en apports mensuels totaux et, pour chaque mois, on peut désagréger l'apport mensuel total en apports mensuels aux différents sites du système. Les étapes peuvent être exécutées dans un ordre différent. La figure 3.1 montre les options de configuration disponibles dans le logiciel SPIGOT (Grygier et Stedinger, 1990, 1991). Le schéma I correspond à l'exemple ci-dessus.

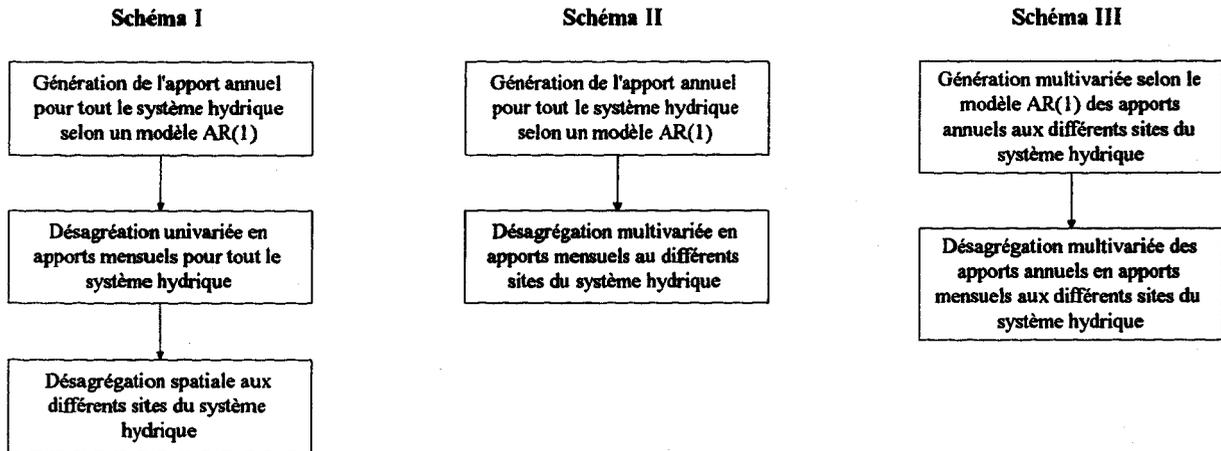


Figure 3.1 Options de configuration dans SPIGOT

Soulignons qu'en utilisant un modèle de désagrégation par étapes, on réduit le nombre de corrélations explicitement préservées. Les modèles de désagrégation font l'objet de recherches actives; on peut en particulier mentionner les travaux de Todini (1980), Grygier et Stedinger (1988), Lin (1990), Lin et Lee (1992), Santos et Salas (1992) et Koutsoyiannis (1992).

### 3.3.2 Modèles PARMA

La transformation et la standardisation des données mensuelles permettent d'éliminer une grande partie de la variation saisonnière, en particulier celle liée à l'échelle des apports. Cependant, comme le montre la figure 3.2, même après une transformation et une standardisation des données mensuelles, une certaine saisonnalité demeure. Elle apparaît dans la corrélation entre les apports de deux mois consécutifs. Ainsi, les modèles classiques de type ARMA, qui reposent sur l'hypothèse de stationnarité, ne sont pas appropriés. Par contre, le modèle PARMA (*periodic autoregressive moving average*), qui est une généralisation du modèle ARMA classique, permet de modéliser explicitement la non-stationnarité due à la variation saisonnière des apports mensuels. Le modèle PARMA(p,q) univarié est défini par la relation suivante (Salas et al. 1980):

$$X_{v,\tau} = \sum_{i=1}^p \phi_{i,\tau} X_{v,\tau-i} + \varepsilon_{v,\tau} - \sum_{j=1}^q \theta_{j,\tau} \varepsilon_{v,\tau-j} \quad (11)$$

où  $X_{v,\tau}$  est l'apport (transformé et standardisé) de l'année  $v$  et du mois  $\tau$ ,  $\phi$  et  $\theta$  sont des paramètres autorégressifs et de moyenne mobile respectivement,  $p$  et  $q$  représentent le nombre de termes autorégressifs et de moyenne mobile respectivement. La variable  $\varepsilon_{v,\tau}$  est un terme d'erreur qui suit une loi normale de moyenne nulle et de variance  $g_\tau$ . Dans ce qui suit, on suppose que l'année est constituée de  $\omega$  périodes (pour des données mensuelles,  $\omega = 12$ ). Notons que le modèle ci-dessus comprend comme cas particulier le modèle de Thomas et Fiering (1962) (PARMA(1,0)).

Le modèle PARMA( $p,q$ ) possède  $\omega(p+q+1)$  paramètres qui doivent être estimés à partir de données observées. La méthode des moindres carrés est la plus facile à utiliser bien que d'autres approches, comme la méthode des moments, pourraient également être considérées (Rasmussen et al. 1995a). Dans le cas de la méthode des moindres carrés, on cherche les paramètres qui minimisent la somme des résidus élevés au carré, c'est-à-dire la fonction suivante:

$$\sum_{v,\tau} \varepsilon_{v,\tau}^2 = \sum_{v,\tau} \left[ X_{v,\tau} - \sum_{i=1}^p \phi_{i,\tau} X_{v,\tau-i} + \sum_{j=1}^q \theta_{j,\tau} \varepsilon_{v,\tau-j} \right]^2 \quad (12)$$

Il existe plusieurs algorithmes pour résoudre ce type de problème d'optimisation; cependant, il est hors du cadre de cette étude d'en discuter. Une fois que les paramètres minimisant  $\sum \varepsilon_{v,\tau}^2$  ont été identifiés, la variance  $g_\tau$  des termes d'erreur peut être estimée directement à partir de la matrice de résidus.

Dans le cas du modèle ARMA( $p,q$ ) stationnaire, il existe plusieurs outils pour identifier l'ordre du modèle, c'est-à-dire les valeurs de  $p$  et de  $q$ . La fonction d'autocorrélation et la fonction d'autocorrélation partielle sont souvent utilisées à cette fin. Cependant, dans le cas des modèles PARMA, ces fonctions sont difficiles à interpréter. On considère plutôt la fonction d'autocorrélation périodique qui est définie de la façon suivante:

$$\rho_{k,\tau} = \frac{\text{Cov}[X_\tau, X_{\tau-k}]}{\sqrt{\text{Var}[X_\tau] \text{Var}[X_{\tau-k}]}} \quad (13)$$

où l'indice  $v$  a été éliminé pour alléger la notation. La fonction d'autocorrélation périodique est donc fonction de deux variables, la période  $\tau$  et le décalage  $k$  (*lag*). On peut en principe utiliser cette fonction pour identifier les valeurs appropriées de  $p$  et  $q$  pour chaque période car théoriquement, il n'est pas nécessaire que  $p$  et  $q$  soient les mêmes à chaque période (Vecchia, 1985). Cependant, on considère généralement des valeurs constantes de  $p$  et  $q$ . En effet, l'approche privilégiée en pratique consiste à fixer d'abord  $p$  et  $q$  *a priori*, à estimer ensuite les

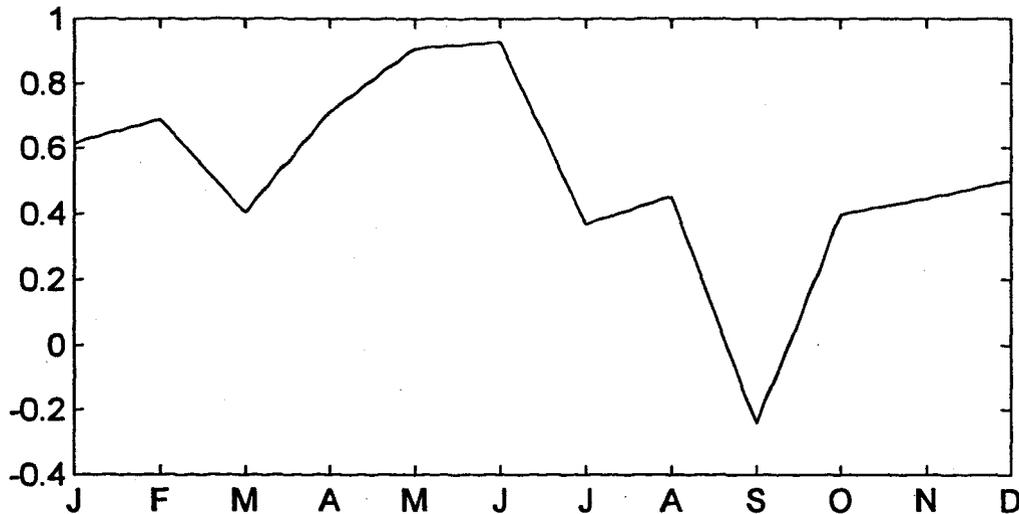


Figure 3.2 Corrélation d'ordre 1 pour site le I93891 du complexe Grande Baleine

paramètres  $\phi$  et  $\theta$  du modèle PARMA( $p,q$ ) et enfin à examiner la performance du modèle en terme de reproduction des caractéristiques de base des données observées.

Pour respecter le principe de parcimonie, on doit utiliser un modèle possédant peu de paramètres et donc se limiter à de faibles valeurs de  $p$  et  $q$ . Les modèles les plus utilisés en hydrologie sont PARMA(1,0), PARMA(2,0) et PARMA(1,1). Récemment, Bartolini et Salas (1993) ont étudié le modèle PARMA(2,2) et ont démontré qu'il préservait mieux la persistance interannuelle des apports que les modèles d'ordres inférieurs. Notons que le modèle PARMA(2,2) a été utilisé dans une récente étude portant sur le système hydrique de la rivière des Outaouais avec de très bons résultats (Rasmussen et al., 1995b).

On peut mentionner qu'il est possible d'évaluer de manière analytique la fonction d'autocorrélation d'un modèle estimé, ce qui permet une évaluation préliminaire de l'adéquation du modèle. En effet, Rasmussen et al. (1995b) ont dérivé les relations suivantes applicables au modèle PARMA(2,2) et à tous les modèles d'ordres inférieurs:

$$\begin{aligned}
 m_{0,\tau} = & \phi_{1,\tau} m_{1,\tau} + \phi_{2,\tau} m_{2,\tau} + g_{\tau} - \theta_{1,\tau} g_{\tau-1} [\phi_{1,\tau} - \theta_{1,\tau}] \\
 & - \theta_{2,\tau} g_{\tau-2} [\phi_{1,\tau} \phi_{1,\tau-1} - \phi_{1,\tau} \theta_{1,\tau-1} + \phi_{2,\tau} - \theta_{2,\tau}]
 \end{aligned}
 \tag{14a}$$

$$m_{1,\tau} = \phi_{1,\tau} m_{0,\tau-1} + \phi_{2,\tau} m_{1,\tau-1} - \theta_{1,\tau} g_{\tau-1} - \theta_{2,\tau} g_{\tau-2} [\phi_{1,\tau-1} - \theta_{1,\tau-1}] \quad (14b)$$

$$m_{2,\tau} = \phi_{1,\tau} m_{1,\tau-1} + \phi_{2,\tau} m_{0,\tau-2} - \theta_{2,\tau} g_{\tau-2} \quad (14c)$$

$$m_{k,\tau} = \phi_{1,\tau} m_{k-1,\tau-1} + \phi_{2,\tau} m_{k-2,\tau-2} \quad k > 2 \quad (14d)$$

où  $m_{k,\tau}$  est la fonction d'autocovariance périodique. En particulier,  $m_{0,\tau}$  est la variance périodique du processus (transformé). Pour un processus standardisé, on s'attend à ce que la variance soit proche de 1. La fonction d'autocovariance périodique est reliée à la fonction d'autocorrélation périodique de la façon suivante:

$$m_{k,\tau} = \text{Cov}[X_\tau, X_{\tau-k}] = \rho_{k,\tau} \sqrt{m_{0,\tau} m_{0,\tau-k}}$$

Une fois le modèle estimé, les paramètres  $\phi_{i,\tau}$  et  $\theta_{i,\tau}$  ainsi que la variance  $g_\tau$  sont connus (une estimation). Le système d'équation (14) est donc linéaire en  $m_{k,\tau}$  qui sont les termes que l'on veut déterminer. Après quelques manipulations sur les équations du système (14), on obtient le système suivant pour chaque période ( $2\omega$  équations):

$$\begin{aligned} -\hat{\phi}_{2,\tau}^2 m_{0,\tau-2} + m_{0,\tau} - \hat{\phi}_{1,\tau} \hat{\phi}_{2,\tau} m_{1,\tau-1} - \hat{\phi}_{1,\tau} m_{1,\tau} \\ = \hat{g}_\tau - \hat{\theta}_{1,\tau} \hat{g}_{\tau-1} [\hat{\phi}_{1,\tau} - \hat{\theta}_{1,\tau}] - \hat{\theta}_{2,\tau} \hat{g}_{\tau-2} [\hat{\phi}_{1,\tau} \hat{\phi}_{1,\tau-1} - \hat{\phi}_{1,\tau} \hat{\theta}_{1,\tau-1} + 2\hat{\phi}_{2,\tau} - \hat{\theta}_{2,\tau}] \end{aligned} \quad (15a)$$

$$-\hat{\phi}_{1,\tau} m_{0,\tau-1} - \hat{\phi}_{2,\tau} m_{1,\tau-1} + m_{1,\tau} = -\hat{\theta}_{1,\tau} \hat{g}_{\tau-1} - \hat{\theta}_{2,\tau} \hat{g}_{\tau-2} [\hat{\phi}_{1,\tau-1} - \hat{\theta}_{1,\tau-1}] \quad (15b)$$

qui peut facilement être résolu pour  $m_{0,\tau}$  et  $m_{1,\tau}$ . À l'aide des relations (14c-d), on peut obtenir la fonction d'autocovariance périodique complète et calculer la fonction d'autocorrélation périodique. La fonction  $m_{k,\tau}$  ainsi obtenue peut être comparée à la fonction d'autocorrélation périodique des données observées. Si le modèle ne reproduit pas de façon adéquate les caractéristiques statistiques des données, on le rejette et on considère d'autres modèles.

Dans l'exposé ci-dessus, nous n'avons considéré que le modèle PARMA univarié. Pour la plupart des cas pratiques, on s'intéresse aux apports à plusieurs sites et il faut alors considérer un modèle multivarié. En général, on peut écrire le modèle PARMA multivarié de la manière suivante:

$$X_{v,\tau} = \sum_{i=1}^p \Phi_{i,\tau} X_{v,\tau-i} + \varepsilon_{v,\tau} - \sum_{j=1}^q \Theta_{j,\tau} \varepsilon_{v,\tau-j} \quad (16)$$

où  $X_{v,\tau}$  est un vecteur de dimension  $[k \times 1]$ ,  $\Phi_{i,\tau}$  et  $\Theta_{j,\tau}$  sont des matrices de paramètres autorégressifs et de moyenne mobile de dimension  $[k \times k]$  et  $\varepsilon_{v,\tau}$  est un vecteur de  $k$  variables

aléatoires distribuées selon une loi normale multivariée de vecteur-moyenne nul et matrice de variances et covariances  $G_\tau$ . Soulignons que les  $\varepsilon_{v,\tau}$  sont indépendants dans le temps. Dans la forme la plus générale du modèle, les matrices de paramètres sont considérées pleines. Cependant, l'estimation de  $\Phi_{i,\tau}$  et  $\Theta_{i,\tau}$  devient délicate et pose quelques problèmes. C'est pourquoi, en pratique, on considère plutôt des matrices diagonales de paramètres. L'avantage d'une telle représentation est que l'estimation des paramètres autorégressifs et de moyenne mobile peut se décomposer en  $k$  problèmes d'estimation univariée (correspondant à chacun des  $k$  sites du système). La corrélation des apports transformés aux  $k$  sites est introduite à partir de la corrélation des résidus, c'est-à-dire à l'aide de la matrice  $G_\tau$ . Ce type de modèle a été appelé *contemporain* puisque seule la corrélation croisée de décalage zéro est modélisée explicitement. L'approche a souvent été utilisée dans le cas des modèles stationnaires (Pegram et James, 1972; Camacho et al., 1985, 1987; Salas et al., 1985; et Stedinger et al., 1985a), mais rarement pour les modèles périodiques (Salas et al., 1980; Salas et al., 1985; Haltiner et Salas, 1988; Bartolini et al., 1988).

Dans le cas de la modélisation multivariée des apports saisonniers à l'aide du modèle PARMA contemporain, le problème est donc réduit à l'estimation de la matrice  $G_\tau$  pour chaque période de l'année. La méthode du maximum de vraisemblance a été largement utilisée pour estimer  $G_\tau$ . Cette méthode est simple et conduit toujours à des résultats. Plus particulièrement, elle consiste à établir la matrice de résidus à chaque site pour la période d'observations concomitantes. Les lignes de ces matrices sont les années et les colonnes sont les mois. Par exemple, pour calculer  $G_1$ , la matrice de variance-covariance des résidus de janvier, on extrait la première colonne de chacune des  $k$  matrices résiduelles et calcule leur variances et covariances. Pour le cas de deux sites, on obtient alors:

$$G_1 = \begin{bmatrix} G_1^{(11)} & G_1^{(12)} \\ G_1^{(21)} & G_1^{(22)} \end{bmatrix} \quad (17)$$

où  $G_1^{(12)}$  est la corrélation entre les résidus du premier mois des deux sites. On peut noter que les éléments de la diagonale correspondent aux variances des résidus déjà obtenues lors de l'estimation univariée.

Tel que démontré par Rasmussen et al. (1995b, 1996), il existe des moyens analytiques permettant d'évaluer la corrélation croisée produite par un modèle particulier. De façon générale, si on définit:

$$M_{k,\tau}^{(ij)} = \text{Cov}[X_\tau^{(i)}, X_{\tau-k}^{(j)}] \quad (18)$$

comme étant la covariance décalée entre les sites  $i$  et  $j$ , on peut déduire le système d'équations suivant:

$$\begin{aligned} & -\hat{\phi}_{2,\tau}^{(i)}\hat{\phi}_{2,\tau}^{(j)} M_{0,\tau-2}^{(ij)} + M_{0,\tau}^{(ij)} - \hat{\phi}_{1,\tau}^{(i)}\hat{\phi}_{2,\tau}^{(j)} M_{1,\tau-1}^{(ij)} - \hat{\phi}_{1,\tau}^{(j)} M_{1,\tau}^{(ij)} \\ & = -\hat{\theta}_{2,\tau}^{(i)}\hat{G}_{\tau-2}^{(ij)}\hat{\phi}_{2,\tau}^{(j)} + \hat{G}_{\tau}^{(ij)} - (\hat{\phi}_{1,\tau}^{(i)} - \hat{\theta}_{1,\tau}^{(i)})\hat{G}_{\tau-1}^{(ij)}\hat{\theta}_{1,\tau}^{(j)} - (\hat{\phi}_{1,\tau}^{(i)}\hat{\phi}_{1,\tau-1}^{(i)} - \hat{\phi}_{1,\tau}^{(i)}\hat{\theta}_{1,\tau-1}^{(i)} + \hat{\phi}_{2,\tau}^{(i)} - \hat{\theta}_{2,\tau}^{(i)})\hat{G}_{\tau-2}^{(ij)}\hat{\theta}_{2,\tau}^{(j)} \end{aligned} \quad (19a)$$

$$\hat{\phi}_{1,\tau}^{(j)} M_{0,\tau-1}^{(ij)} + \hat{\phi}_{2,\tau}^{(j)} M_{1,\tau-1}^{(ij)} - M_{1,\tau}^{(ji)} = \hat{G}_{\tau-1}^{(ij)}\hat{\theta}_{1,\tau}^{(i)} + (\hat{\phi}_{1,\tau-1}^{(i)} - \hat{\theta}_{1,\tau-1}^{(i)})\hat{G}_{\tau-2}^{(ij)}\hat{\theta}_{2,\tau}^{(j)} \quad (19b)$$

Bien que ce système apparaisse assez complexe, il s'agit en réalité d'un système linéaire de  $3\omega$  équations avec le même nombre d'inconnus ( $M_{0,\tau}^{(ij)}=M_{0,\tau}^{(ji)}$ ,  $M_{1,\tau}^{(ij)}$  et  $M_{1,\tau}^{(ji)}$ ) qui peut être résolu sans difficulté numérique. En pratique, on s'intéresse notamment à  $M_{0,\tau}^{(ij)}$ , la corrélation croisée entre les deux stations. La comparaison avec la corrélation croisée des données observées permet d'évaluer rapidement (sans appel à la simulation) la performance d'un modèle multivarié. Notons que les expressions ci-dessus sont valables pour le modèle PARMA(2,2) contemporain et tous les modèles d'ordre inférieur.

Bien que simple, la méthode du maximum de vraisemblance ne donne pas toujours à de bons résultats. En effet, cette méthode conduit souvent à une sous-estimation considérable de la corrélation croisée historique. Une approche alternative à la méthode du maximum de vraisemblance est la méthode des moments qui vise spécifiquement à reproduire exactement la corrélation croisée entre les sites.

Cette méthode a été étudiée par Haltiner et Salas (1988) pour le modèle PARMA(1,1) et par Rasmussen et al. (1996) pour le modèle PARMA(2,2). On peut montrer que l'estimation par la méthode des moments correspond à la résolution d'un système d'équations linéaires pour chaque combinaison de deux stations. Toutefois, l'application de la méthode des moments ne conduit pas toujours à une solution acceptable. En effet, il peut très bien arriver que les matrices  $G_{\tau}$  obtenues par la méthode des moments ne soient pas définies positives, ce qui, pour une matrice des variances et covariances, n'est pas acceptable. Rasmussen et al. (1996) proposent différentes modifications de la méthode des moments dans le cas où une solution exacte n'est pas possible. Notons que ce sujet fait actuellement l'objet d'une recherche active.

## 4 Classification des séries synthétiques

---

### 4.1 Introduction

Tel que indiqué au chapitre 2 de ce rapport, la nouvelle version du modèle MINERVE permet de tenir compte de plusieurs scénarios d'apports dans les contraintes de conservation de l'eau. Cependant, en raison de la complexité du modèle et du temps important d'exécution, seul un nombre réduit de scénarios peut être considéré. On a en pratique jugé réaliste de retenir dix scénarios. Les séries d'apports utilisées dans MINERVE doivent couvrir de façon adéquate la gamme possible de scénarios futurs, c'est-à-dire de scénarios *pessimistes* à scénarios *optimistes*.

Les séries identifiées doivent posséder des caractéristiques hydrologiques réalistes. L'application des modèles stochastiques décrits dans la section précédente a pour objectif la génération de scénarios d'apports réalistes du point de vue hydrologique. Il s'agit alors de générer un grand nombre de séries équiprobables et d'en choisir une dizaine selon des critères à établir. En effet, il est nécessaire de définir plus précisément ce que l'on entend par scénarios pessimistes et optimistes. Le terme *équiprobable* est souvent utilisé pour caractériser la nature probabiliste des séries d'apports générées à l'aide d'un modèle stochastique. Cependant, parce que les apports sont des variables continues et qu'il s'agit d'un problème multivarié, tant dans le temps que dans l'espace, la probabilité qu'un des  $N$  scénarios simulés se produise effectivement dans le futur est nulle. Le terme *équiprobable* peut donc entraîner une certaine confusion; essentiellement, il veut dire que l'on associe *a priori* (avant la simulation) la même probabilité aux différents scénarios. Cependant, si par exemple un scénario généré a un apport moyen annuel cinquante pour cent plus élevé que la moyenne observée, on le considère naturellement moins probable que les séries qui ont un apport annuel moyen proche de la moyenne historique. En faisant cette observation, on a intuitivement introduit comme critère de comparaison la moyenne annuelle des apports. En effet, pour comparer les scénarios générés, on doit nécessairement adopter des critères qui réduisent le caractère multidimensionnel des séries en un nombre faible de dimensions. Plus particulièrement, si l'on veut associer une probabilité non nulle aux scénarios, on doit utiliser des classes discrètes pour les caractéristiques statistiques, car on ne peut associer une probabilité à une variable continue.

Il importe de souligner que les caractéristiques statistiques utilisées pour classer les séries simulées doivent être choisies en fonction du problème étudié. Si l'on s'intéresse aux valeurs extrêmes élevées (ou faibles), on doit choisir par exemple l'apport maximum (ou minimum) de chaque série générée et classer les séries selon ce critère. Pour simplifier l'analyse, on doit éviter l'introduction des critères non pertinents. Dans le cas de l'aménagement d'un nouveau bassin, deux caractéristiques des scénarios semblent particulièrement pertinents: le volume total d'eau et la répartition de l'eau à l'intérieur de la période simulée. En ce qui concerne la répartition de l'eau, on s'intéresse souvent aux séquences d'apports faibles. Plusieurs critères ont été proposés pour décrire différents aspects des séries simulées. Dans ce qui suit, nous passerons en revue les critères les plus courants. Nous montrerons ensuite comment modéliser statistiquement l'interaction des différents critères et, plus particulièrement, comment des méthodes statistiques multivariées permettent de condenser un grand nombre de critères en quelques facteurs principaux.

## 4.2 Choix de séries agrégées représentatives

Avant d'entreprendre la présentation des différentes caractéristiques statistiques utilisées pour discriminer et classer les scénarios simulés, il importe de préciser les variables qui seront considérées. De manière générale, les systèmes hydriques que l'on désire modéliser avec MINERVE comprennent plusieurs sites et les séries simulées à l'aide du modèle stochastique retenu sont alors multivariées. Dans la classification des scénarios simulés, il semble impossible de tenir compte explicitement de l'aspect multidimensionnel à partir de plusieurs critères. C'est pourquoi nous proposons d'établir des séries agrégées spatialement et de baser le classement et le choix des scénarios sur ces séries.

L'agrégation des séries multivariées peut être effectuée de plusieurs façons. On peut envisager au moins trois options:

### *Option 1: Agrégation des apports intermédiaires*

En général, on modélise et simule les apports intermédiaires. On peut obtenir une série agrégée en faisant l'agrégation des tous les apports intermédiaires, c'est-à-dire que pour chaque année et chaque mois on fait la sommation des apports intermédiaires. On peut noter que la série ainsi agrégée correspond à l'apport total du système hydrique ou bien à l'apport que l'on observerait à la centrale située la plus en aval du système.

*Option 2: Agrégation des apports bruts*

L'option 1 ne tient pas compte du fait que l'apport intermédiaire n'a pas la même valeur à chaque site du système. Par exemple, pour le cas de centrales au fil de l'eau, il est souhaitable que l'eau entre dans le système le plus en amont possible, car les centrales situées en aval profiteront également de cette eau. On pourrait ainsi considérer une série qui consiste à agréger les apports bruts. Essentiellement, on procède à un laminage des apports intermédiaires et on calcule la somme des apports réels.

On peut noter que l'utilisation de l'option 2 conduit à une série qui contient l'eau disponible à tous les sites (si l'on ne fait pas de régularisation). Ainsi, la série n'est pas réelle car elle ne peut être effectivement observée. Cependant, on peut très bien effectuer l'analyse que nous présenterons dans les sections suivantes sur une telle série.

*Option 3: Série énergétique*

Ce n'est pas seulement la position d'un sous-bassin dans le système hydrique qui détermine la valeur de l'eau. Les caractéristiques physiographiques des sites où l'on planifie la construction des centrales est également un facteur important. Ainsi, la valeur de l'eau est fortement reliée à la hauteur de chute et on pourrait envisager de véritables séries énergétiques. Ceci nécessitera une connaissance de la hauteur approximative de chute à chaque site. D'autres facteurs pourraient également être pris en considération. De manière générale, il est difficile, voire impossible, de donner une valeur précise de la hauteur de chute à un site, les centrales et réservoirs étant encore fictives (rappelons qu'un des buts de l'application du modèle MINERVE est justement d'estimer la hauteur de chute). Cependant, pour la détermination des séries énergétiques en vue d'identifier les scénarios pessimistes et optimistes, il sera acceptable d'admettre des valeurs assez grossières pour les différents facteurs de production.

On peut mentionner que les critères développés initialement pour les séries d'apports peuvent aussi bien être appliqués à des séries énergétiques.

Soulignons enfin que ce ne sont pas les séries agrégées qui seront utilisées dans le modèle MINERVE. Les séries agrégées serviront uniquement à identifier les scénarios pessimistes, moyens et optimistes. Ce sont les séries d'apports intermédiaires utilisées pour établir les séries agrégées qui seront introduites dans le modèle MINERVE.

### 4.3 Revue de certaines caractéristique statistiques de séries synthétiques

Dans la littérature hydrologique, il existe plusieurs bonnes revues des caractéristiques statistiques de séries synthétiques pertinentes pour le dimensionnement des ouvrages hydrauliques. Le rapport de Pegram et al. (1980) en est une excellente référence. Ici, nous mettons l'accent sur les caractéristiques qui sont susceptibles d'être importantes pour l'étude de l'aménagement d'un nouveau bassin. On considère le cas d'une seule série obtenue tel que décrit dans la section précédente. Le terme *apport* est utilisé de façon générique dans ce qui suit; la variable considérée pourrait aussi bien être une quantité d'énergie (série énergétique). De plus, on considère des séries annuelles en convertissant les séries agrégées saisonnières en séries annuelles. En effet, pour l'aménagement de grands systèmes hydriques où l'évaluation de la capacité de stockage interannuelle est le problème principal, la variation des apports saisonniers à l'intérieur d'une année est rarement un facteur important.

#### 4.3.1 Caractéristiques reliées au volume

Il est évident que les caractéristiques reliées au volume d'eau sont très importantes pour la classification des séries d'apports. On considère généralement l'apport total ou l'apport annuel moyen. On pourrait également considérer le volume printanier moyen seulement, bien que celui-ci soit étroitement relié à l'apport annuel moyen.

#### 4.3.2 Caractéristiques reliées à la distribution

Dans l'analyse générale de stockage, on distingue les réservoirs fini, semi-infini et infini. Le réservoir fini a des bornes inférieure et supérieure fixes. Le réservoir semi-infini a une borne supérieure fixe, mais pas de borne inférieure. Quant au réservoir infini, il n'a aucune borne. C'est ce dernier type de réservoir qui sera considéré dans ce qui suit.

Nous donnerons d'abord quelques définitions qui s'avéreront utiles pour l'étude des réservoirs et des besoins de stockage. Si l'on définit  $X_t$  comme l'apport brut qui entre dans un réservoir au temps  $t$  et  $Y_t$  comme le volume d'eau extrait du réservoir (production d'hydroélectricité, évaporation, irrigation, etc.) dans la même période, on obtient le changement de stockage comme:

$$\Delta V_t = V_t - V_{t-1} = X_t - Y_t \quad (20)$$

On définit le degré d'exploitation,  $\alpha$ , par le rapport entre des espérances mathématiques de  $Y_t$  et de  $X_t$ :

$$\alpha = \frac{E[Y_t]}{E[X_t]} \quad (21)$$

où  $E[Y_t]$  est le volume que l'on espère exploiter pour la production. En pratique, on estime  $E[X_t]$  par la moyenne,  $\bar{X}_t$ .

Considérons une série chronologique,  $\{X_t\} = (X_1, X_2, \dots, X_n)$ . On définit la séquence de sommes partielles de la façon suivante:

$$S_k = S_{k-1} + (X_k - \alpha \bar{X}_n) = \sum_{\tau=1}^k X_\tau - \alpha k \bar{X}_n \quad \tau = 1, 2, \dots, n \quad (22)$$

où  $\bar{X}_n = \frac{1}{n} \sum_{\tau=1}^n X_\tau$ . Le processus  $S_k$  représente alors l'écart cumulé entre les intrants et la moyenne (réservoir infini). Pour caractériser une série simulée, on peut considérer le maximum et le minimum du processus  $S_t$ :

$$M_n = \max(0, S_1, S_2, \dots, S_n) \quad (23)$$

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

Souvent on considère  $\alpha = 1$ , ce qui correspond à une exploitation de toute l'eau. Dans ce cas, la somme partielle générale est nommée somme partielle ajustée (*adjusted partial sum*). Plus précisément, on définit la somme partielle ajustée comme suit:

$$S_k^* = S_{k-1}^* + (X_k - \bar{X}_n) = \sum_{\tau=1}^k X_\tau - k \bar{X}_n \quad k = 1, 2, \dots, n \quad (24)$$

En particulier, on trouve que  $S_n^* = 0$ . Pour caractériser une série donnée, on peut considérer le maximum et le minimum de  $S_k^*$ :

$$M_n^* = \max(0, S_1^*, S_2^*, \dots, S_n^*) \quad (25)$$

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

*Domaine de variation général et ajusté*

Le domaine de variation général (*range*),  $R_n$ , est le domaine de variation des sommes partielles, c'est-à-dire:

$$R_n = \max_{k \in \{0, n\}} S_k - \min_{k \in \{0, n\}} S_k = M_n - m_n \quad (26)$$

La figure 4.1 montre la définition des sommes partielles maximum et minimum ainsi que leur domaine de variation.

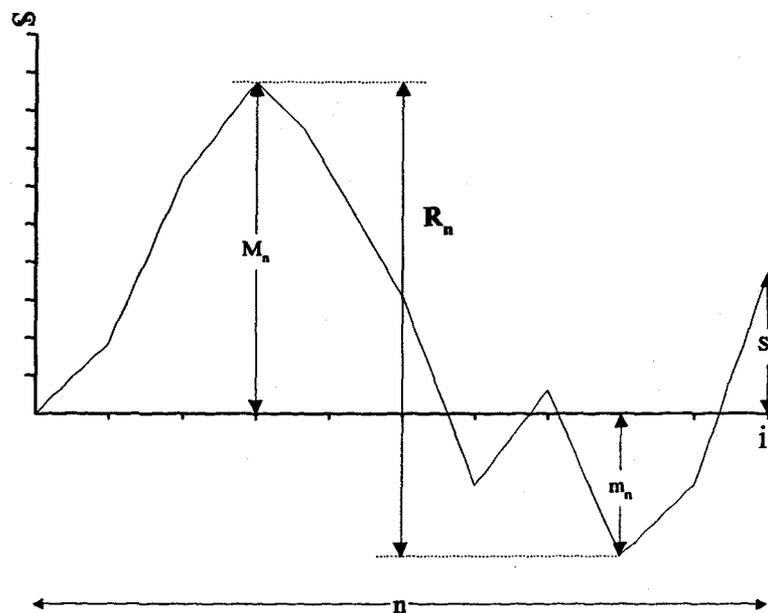


Figure 4.1 Définition des sommes partielles et du domaine de variation

On s'intéresse souvent au domaine de variation ajusté (*adjusted range*) qui est donné par:

$$R_n^* = \max_{k \in \{0, n\}} S_k^* - \min_{k \in \{0, n\}} S_k^* = M_n^* - m_n^* \quad (27)$$

Cette caractéristique statistique peut être interprétée de façon physique. En effet,  $R_n^*$  représente le besoin de stockage, si l'on veut exploiter le volume total d'eau à un taux (*rate*) constant. La figure 4.2 montre la définition de la somme partielle ajustée et son domaine de variation.

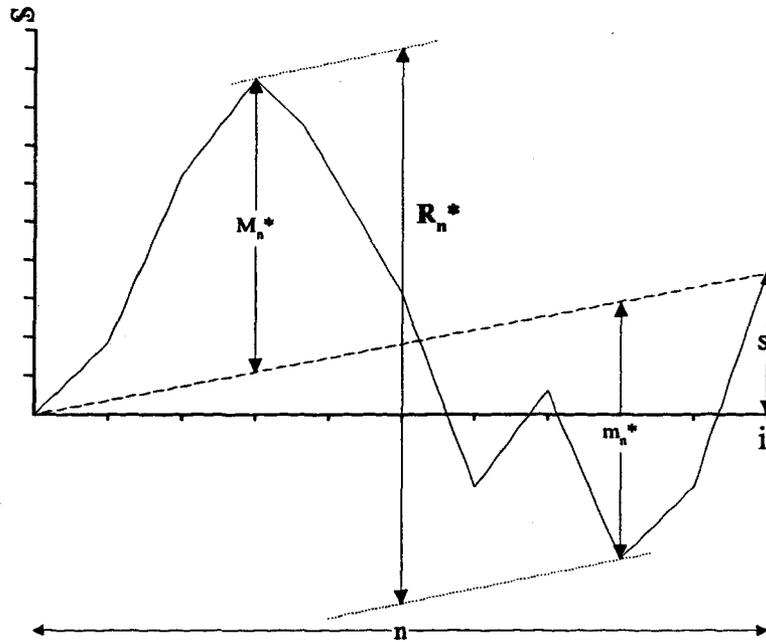


Figure 4.2 Définition des sommes partielles ajustées et du domaine de variation ajusté

#### Domaine de variation ajusté standardisé

Le domaine de variation ajusté standardisé (*rescaled adjusted range*) est défini comme:

$$\bar{R}_n^* = \frac{R_n^*}{\hat{\sigma}_n^2} \quad (28)$$

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

#### Coefficient de Hurst

Suite à de nombreuses études expérimentales, Hurst (1951) a montré l'existence d'une relation linéaire entre  $\bar{R}_n^*$  et  $n$  dans un diagramme log-log. Cette observation l'a conduit à proposer une relation du type:

$$\bar{R}_n^* = \frac{R_n^*}{\hat{\sigma}_n^2} = k n^h \quad (29)$$

où  $k$  et  $h$  sont des constantes. On peut montrer que, pour  $n = 2$ ,  $\bar{R}_n^* = 1$ . On en déduit que  $k = 2^{-h}$ . Pour cette raison, Hurst a proposé la relation suivante

$$\bar{R}_n^* = (n/2)^H \quad (30)$$

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 examinées par Hurst, la moyenne du coefficient  $H$  était égale à 0.73 avec un écart-type de 0.08. Ce paradoxe a été nommé phénomène de Hurst. Feller (1951) a suggéré qu'un modèle markovien pourrait expliquer le phénomène, mais Moran (1959) a pu démontrer que même pour les processus autocorrélés, on obtient une valeur asymptotique  $H = 0.5$ . En raison de sa relation avec le domaine de variation des sommes ajustées et le besoin de stockage, on a traditionnellement accordé beaucoup d'importance au coefficient de Hurst.

En pratique, on estime  $H$  comme suit:

$$H = \frac{\log \bar{R}_n^*}{\log(n/2)} \quad (31)$$

### *Séquences*

Lors de l'évaluation du besoin de stockage (capacité des réservoirs) pour les systèmes hydriques d'envergure, les séquences d'années sèches jouent un rôle important. C'est pourquoi les hydrologues font souvent appel à l'analyse des séquences (*runs analysis*) au-dessous ou au-dessus d'un certain seuil. En général, on considère comme seuil la moyenne à long terme qui peut être estimée à partir de la série historique observée. Dans le cas des séries simulées, on peut également considérer les séquences par rapport à la moyenne de chaque série, au moins lors de la validation d'un modèle stochastique. Pour le classement des séries synthétiques, il semble cependant plus logique d'utiliser la moyenne historique comme seuil. Si l'on considère les séquences par rapport à la moyenne de chaque série, on doit tenir compte simultanément des moyennes.

Dans le cas de l'aménagement d'un nouveau bassin, on doit s'intéresser aux séquences d'étiage, c'est-à-dire au-dessous du seuil. Dracup et al. (1980a,b) ont proposé plusieurs caractéristiques statistiques pour décrire une séquence, notamment la longueur, le déficit accumulé et l'intensité. La longueur d'une séquence (RL - *run length*) est définie comme le nombre d'années consécutives pendant lesquelles l'apport annuel reste au-dessous du seuil. Le déficit accumulé (RS - *run sum*)

est défini comme la différence entre l'apport moyen pendant les RL années et l'apport observé/simulé pendant la même période. Quant à l'intensité d'une séquence (RI - *run intensity*), elle correspond au déficit moyen pendant l'étiage et est reliée à RL et RS par la relation  $RI=RS/RL$ . La figure 4.3 illustre les différentes caractéristiques d'une séquence.

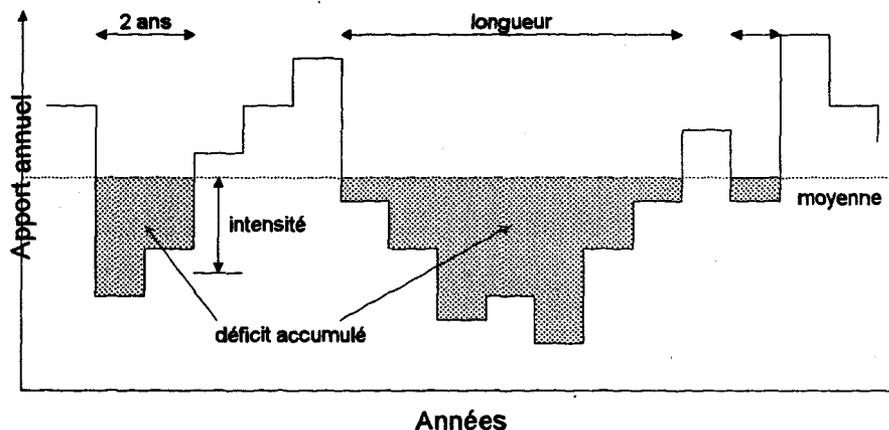


Figure 4.3 Définition des caractéristique d'une séquence

De manière générale, chaque série simulée contient plusieurs séquences au-dessous du seuil et il faut donc réduire la quantité d'information en quelques caractéristiques statistiques principales. Pour une série donnée et pour chacune des variables, RL, RS et RI, on peut calculer la moyenne, l'écart-type et le maximum qui résument les caractéristiques principales des séquences.

Notons qu'il est parfois avantageux de considérer une moyenne mobile des apports quand on calcule les séquences. On évite ainsi qu'une longue séquence d'années généralement sèches soit divisée en deux séquences par une année dont l'apport annuel est légèrement au-dessus du seuil (Frick et al., 1990). On peut, par exemple, calculer toutes les moyennes des valeurs de deux années consécutives et baser l'analyse sur la série ainsi obtenue.

#### 4.4 Analyse fréquentielle des caractéristiques statistiques

Nous avons déjà souligné qu'il est impossible d'associer une probabilité non nulle à la réalisation d'une variable continue telle que les apports simulés à l'aide d'un modèle stochastique et qu'il est nécessaire de baser le classement sur les caractéristiques statistiques des séries d'apports générées.

Ainsi, on peut choisir une gamme de critères parmi ceux présentés dans la section précédente et effectuer une analyse statistique de ceux-ci.

Deux problèmes majeurs se présentent dans l'analyse statistique fréquentielle des critères:

- il n'est pas réaliste de considérer plus de deux critères pour l'analyse fréquentielle; et
- il y a une interaction (corrélation) complexe entre les différents critères.

Si l'on trouve que seulement deux critères parmi ceux exposés dans la section précédente sont d'intérêt, le premier problème ne se pose pas. Cependant, les différents critères caractérisent différents aspects des séries et, de manière générale, il n'est pas facile, *a priori*, d'en isoler deux qui discriminent adéquatement les séries. En ce qui concerne la corrélation entre les critères, celle-ci doit certainement être prise en considération dans l'analyse fréquentielle.

Pour résoudre ces problèmes, nous proposons d'utiliser l'analyse en composantes principales qui est un cas particulier de l'analyse factorielle. Dans la section suivante, nous présentons les aspects théoriques de cette méthode. Un exemple d'application sera présenté dans le chapitre 5.

#### **4.4.1 Analyse en composantes principales (ACP)**

Supposons que l'on observe une variable aléatoire de  $p$  dimensions,  $\mathbf{Y} = (y_1, y_2, \dots, y_p)^T$ . Dans le contexte actuel,  $y_i$  représente des caractéristiques statistiques des séries générées, par exemple la moyenne annuelle ou le domaine de variation ajusté. Sans perte de généralité, on suppose que chaque variable  $y_i$  a été standardisée de façon à ce qu'elle ait une moyenne nulle et une variance unité. En raison de la corrélation entre les éléments de  $\mathbf{Y}$ , la dimension de  $\mathbf{Y}$  peut, en pratique, être réduite à  $m < p$ . Très souvent, on peut obtenir une représentation acceptable de  $\mathbf{Y}$  dans un espace à deux ou trois dimensions. L'analyse en composantes principales a pour but d'identifier cet espace. En effet, l'analyse factorielle a deux objectifs:

- réduire les dimensions du problème étudié en identifiant un sous-espace  $\mathcal{R}^m$  de  $\mathcal{R}^p$ , où  $m < p$ ; et
- identifier les facteurs  $\mathbf{F} = (F_1, F_2, \dots, F_m)^T$ , fonctions des observations, qui génèrent le sous-espace  $\mathcal{R}^m$  et y associer une interprétation.

Dans l'analyse factorielle classique, on suppose l'existence de la relation suivante:

$$\mathbf{Y} = \mathbf{AF} + \boldsymbol{\varepsilon} \tag{32}$$

Plus précisément, on suppose que les valeurs que prennent les éléments de  $Y$  peuvent être expliquées par certains facteurs communs non observés. Le terme  $\varepsilon$  représente les facteurs spécifiques, c'est-à-dire la partie de la variation de  $Y$  qui ne peut être expliquée par les facteurs communs. Dans l'analyse factorielle, dont l'analyse en composante principale est un cas particulier, on fait les hypothèses suivantes:

$$\begin{aligned}
 E[\mathbf{Y}] &= \mathbf{0} & \text{Cov}[\mathbf{Y}] &= \Sigma_Y \\
 E[\mathbf{F}] &= \mathbf{0} & \text{Cov}[\mathbf{F}] &= \mathbf{I} \\
 E[\boldsymbol{\varepsilon}] &= \mathbf{0} & \text{Cov}[\boldsymbol{\varepsilon}] &= \Psi = \begin{bmatrix} \Psi_1 & 0 & & 0 \\ 0 & \Psi_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \Psi_p \end{bmatrix}
 \end{aligned} \tag{33}$$

Notons que puisque les variables ont été standardisées,  $\Sigma_Y$  est une matrice de corrélation et non pas une matrice de variances-covariances.  $\mathbf{I}$  est la matrice d'identité. On fait de plus l'hypothèse que les facteurs communs et spécifiques sont indépendants:

$$\text{Cov}[\mathbf{F}, \boldsymbol{\varepsilon}] = \mathbf{0} \tag{34}$$

Le modèle (32) et les différentes hypothèses qui y sont associées constituent le modèle factoriel. La première étape de l'analyse consiste à déterminer la matrice  $\mathbf{A}$ , dont les éléments sont nommés *poids factoriels* (*factor loadings*). Ce nom est dû au fait que la variable  $y_i$  est une combinaison linéaire des facteurs:

$$y_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + \varepsilon_i \tag{35}$$

Ainsi, l'élément  $a_{ij}$  de  $\mathbf{A}$  est le coefficient du facteur  $j$  pour la variable  $i$  ou, plus précisément, la corrélation entre la variable  $i$  et le facteur  $j$ . À partir de l'expression ci-dessus, on montre facilement que:

$$\begin{aligned}
 \text{Var}[y_i] &= \text{Var}[a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + \varepsilon_i] \\
 &= a_{i1}^2 \text{Var}[F_1] + a_{i2}^2 \text{Var}[F_2] + \dots + a_{im}^2 \text{Var}[F_m] + \Psi_i \\
 &= a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 + \Psi_i
 \end{aligned} \tag{36}$$

De façon plus générale, on peut obtenir la relation reliant la matrice de corrélation de  $Y$  à  $\mathbf{A}$  et  $\Psi$ :

$$\begin{aligned}\Sigma_Y &= \text{Cov}[AF + \varepsilon] \\ &= AA^T + \Psi\end{aligned}\quad (37)$$

Dans l'analyse en composantes principales, on ignore les facteurs spécifiques dans la matrice  $\Psi$  lors de l'estimation des poids factoriels, c'est-à-dire que l'on détermine  $A$  de façon à ce que  $AA^T = \Sigma_Y$ . En pratique,  $\Sigma_Y$  est estimée à partir des observations. Dans l'ACP, on doit calculer les valeurs propres et les vecteurs propres de  $\Sigma_Y$ . Définissons la matrice  $\Lambda$  comme:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \lambda_p \end{bmatrix}\quad (38)$$

où  $\lambda_i$ ,  $i = 1, 2, \dots, p$  sont les valeurs propres de  $\Sigma_Y$ . On suppose que celles-ci sont classées de façon à ce que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . On note par  $v_1, v_2, \dots, v_p$  les vecteurs propres standardisés qui correspondent aux  $p$  valeurs propres et on définit la matrice  $V$  qui a comme colonne  $j$  le vecteur propre  $v_j$ . En utilisant la définition même des valeurs propres, il est trivial de montrer que  $V^T \Sigma_Y V = \Lambda$ . L'orthogonalité de  $V$  (i.e.  $VV^T = I$ ) permet d'obtenir l'expression suivante:

$$\Sigma_Y = V\Lambda V^T = V\Lambda^{1/2}(V\Lambda^{1/2})^T\quad (39)$$

$$\text{où } \Lambda^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & & 0 \\ 0 & \sqrt{\lambda_2} & & 0 \\ & & \ddots & \\ 0 & 0 & & \sqrt{\lambda_p} \end{bmatrix}$$

On pourrait ainsi estimer  $A$  par  $V\Lambda^{1/2}$ . Cependant, la matrice  $V\Lambda^{1/2}$  a les dimensions  $[p \times p]$  alors que  $A$  a les dimensions  $[p \times m]$  ( $m$  est le nombre de facteurs que nous désirons retenir). C'est pourquoi on ne retient que les  $m$  valeurs propres les plus grandes et les vecteurs propres correspondants. Il est important de conserver les valeurs propres les plus grandes qui mènent à la meilleure reproduction de la matrice  $\Sigma_Y$  originale. En effet, la valeur propre  $\lambda_i$  est directement reliée à la variance expliquée par l'axe  $i$ . Si l'on retient par exemple deux facteurs, on obtient alors:

$$\Lambda_*^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \quad \text{et} \quad V_* = (v_1, v_2) = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ \vdots & \vdots \\ v_{p1} & v_{p2} \end{bmatrix}\quad (40)$$

En ce qui concerne le choix du nombre de facteurs à considérer, on peut remarquer que la procédure la plus utilisée consiste à retenir autant de facteurs qu'il y a de valeurs propres supérieures à un. La matrice  $A$  est alors estimée par:

$$A = V_* \Lambda_*^{1/2} \quad (41)$$

et on obtient:

$$AA^T = V_* \Lambda_*^{1/2} (V_* \Lambda_*^{1/2})^T \approx \Sigma_Y \quad (42)$$

La matrice  $\Sigma_Y$  observée n'est pas reproduite exactement par le modèle parce que nous n'avons retenu qu'un nombre limité de facteurs. On voit que les colonnes de  $A$  qui contiennent les poids sur chaque facteur considéré sont proportionnelles aux vecteurs propres, c'est-à-dire que les poids sur le facteur  $j$  sont proportionnels aux coefficients de la  $j$ 'ième composante principale.

Il est commode d'examiner les variables observées dans l'espace factoriel. Par exemple, la variable  $y_i$  a les coordonnées  $(a_{i1}, a_{i2})$  dans l'espace des deux premiers facteurs  $(F_1, F_2)$ . Si deux variables sont proches dans l'espace factoriel, cela veut dire que les deux variables expriment essentiellement la même chose par rapport aux deux facteurs. Il est intéressant d'interpréter les facteurs non observés. Pour cela, il est souvent avantageux d'effectuer une rotation des axes factoriels. Cette opération ne change pas la position des points les uns par rapport aux autres, mais elle change évidemment les poids factoriels.

L'interprétation des facteurs, illustrée par un exemple dans le chapitre suivant, est un problème intéressant en soi, mais l'objectif poursuivi ici est de calculer les valeurs des facteurs pour chaque observation (*factor scores*) qui sont obtenus par:

$$F = A^T \Sigma_Y^{-1} Y \quad (43)$$

où  $A \Sigma_Y^{-1}$  sont des coefficients (*factor score coefficients*). En particulier, on remarque que les facteurs sont des combinaisons linéaires des variables originales.

#### 4.4.2 Choix de scénarios

La méthodologie proposée pour classer les séries synthétiques et pour en choisir quelques-unes est basée sur l'analyse fréquentielle des facteurs. Rappelons qu'on doit également associer une probabilité/poids à chaque scénario retenu. Nous utiliserons le terme *facteur* plutôt que *critère*,

mais en principe, un facteur pourrait être simplement une des caractéristique statistiques (critère) utilisée pour décrire les séries synthétiques. Plus précisément, on peut raisonnablement supposer l'existence de deux facteurs importants dont un est relié au volume ( $F_V$ ) et l'autre à la distribution des apports ( $F_D$ ).

Si nous faisons d'abord l'hypothèse que seul le volume est d'intérêt et que tous les autres aspects des séries sont non pertinents, on peut alors calculer le facteur  $F_V$  pour chaque scénario généré et effectuer une analyse fréquentielle sur les valeurs de  $F_V$ . À l'aide d'un histogramme ou un graphique de probabilité empirique, on peut visuellement examiner la distribution de  $F_V$  et estimer la fonction de répartition du facteur. Plus précisément, celle-ci permet d'associer une probabilité à des *intervalles* de  $F_V$ . Par exemple, on peut dire qu'il y a une probabilité,  $p$ , qu'un scénario synthétique ait une valeur de  $F_V$  inférieure à un certain niveau ou qu'il y a une telle probabilité que  $F_V$  soit compris dans un intervalle donné  $[a,b]$ . La mesure de probabilité pourrait être basée sur les probabilités empiriques ou sur l'ajustement d'une loi paramétrique. Si le modèle paramétrique est adéquat, la différence entre les deux approches sera généralement petite.

Si l'on désire retenir dix scénarios, on peut alors diviser l'axe de  $F_V$  en dix intervalles équiprobables. En pratique, on doit trier les  $N$  observations de  $F_V$  (en notant à quelles séries elles sont associées). La première classe est constituée des  $N/10$  valeurs les plus petites de  $F_V$ , etc. Soulignons qu'il est tout à fait légitime d'associer une probabilité à une classe de valeurs de  $F_V$ .

Il reste à choisir un scénario représentatif de chaque classe. Il semble raisonnable de choisir le scénario dont la valeur de  $F_V$  est la plus proche de la valeur moyenne de la classe.

Il faut souligner que l'on n'est pas tenu à choisir des classes équiprobables. En principe, on peut choisir les classes de façon arbitraire. Cependant, le choix peut avoir un impact important sur les résultats produits par le modèle MINERVE. C'est pourquoi l'utilisateur doit effectuer une analyse de sensibilité en examinant les conséquences du choix des classes. Cet aspect sera abordé dans la deuxième phase du présent projet.

Dans le cas plus général où plusieurs facteurs sont importants, on doit alors effectuer une analyse fréquentielle multivariée. Un des avantages de l'utilisation des facteurs obtenus à l'aide d'une analyse en composantes principales est que les facteurs sont indépendants. Il est donc facile d'établir des classes équiprobables, par exemple, dans l'espace  $(F_V, F_D)$ . Il est évident que pour limiter le nombre de classes à une dizaine, il faut considérer des classes plus grandes pour chaque

facteur. Dans le cas de deux facteurs, on pourrait par exemple considérer trois classes pour chaque facteur ce qui conduit à un total de neuf classes comme le montre le tableau ci-dessous.

		$F_V$			Prob. totale
		classe 1	classe 2	classe 3	
$F_D$	classe 1	1/9	1/9	1/9	1/3
	classe 2	1/9	1/9	1/9	1/3
	classe 3	1/9	1/9	1/9	1/3
Prob. totale		1/3	1/3	1/3	1

Pour la construction de ce tableau, on a fait l'hypothèse que pour chaque facteur, les trois classes sont équiprobables, c'est-à-dire ont une probabilité 1/3. Si les deux facteurs de plus sont indépendants, on peut alors associer la probabilité 1/9 à chacune des classes dans l'espace  $(F_V, F_D)$ . On peut également calculer le nombre relatif de séries tombant dans chaque classe et leur associer une probabilité empirique. On s'attend à ce qu'environ 1/9 des séries simulées tombe dans chaque classe.

Comme dans le cas de l'analyse fréquentielle univariée, il n'est pas nécessaire que les classes soient équiprobables dans l'analyse multivariée. Il n'est pas non plus nécessaire que les facteurs considérés soient divisés en un même nombre de classes. Par exemple, si l'on considère avantageux de couvrir une gamme plus large du facteur  $F_V$ , on pourrait considérer quatre classes pour ce facteur et deux classes seulement pour le facteur  $F_D$ , ce qui conduirait à huit classes dans l'espace  $(F_V, F_D)$ .

Pour le choix des scénarios à retenir, on procède comme dans le cas univarié. Plus précisément, on doit calculer le centre de gravité (moyenne) des points dans l'espace considéré et choisir le scénario qui se trouve le plus proche du centre. Ici, on notera d'ailleurs un autre avantage de l'utilisation des facteurs indépendants et standardisés: la distance entre deux points  $i$  et  $j$  est une simple distance euclidienne:

$$\text{dist}(i, j) = \left[ (F_i - F_j)^T (F_i - F_j) \right]^{1/2} = \sqrt{(F_{D,i} - F_{D,j})^2 + (F_{V,i} - F_{V,j})^2}$$

où  $F_i = (F_{D,i}, F_{V,i})^T$  sont les facteurs. Dans le cas où l'on considère des critères bruts, par exemple la moyenne annuelle  $(\bar{Q})$  et le domaine de variation  $(R)$ , on doit définir le métrique:

$$\text{dist}(i, j) = \left( \begin{bmatrix} \bar{Q}_i \\ \bar{R}_i \end{bmatrix} - \begin{bmatrix} \bar{Q}_j \\ \bar{R}_j \end{bmatrix} \right)^T \Sigma_F^{-1} \left( \begin{bmatrix} \bar{Q}_i \\ \bar{R}_i \end{bmatrix} - \begin{bmatrix} \bar{Q}_j \\ \bar{R}_j \end{bmatrix} \right)$$

où  $\Sigma_F^{-1}$  est la matrice de variance-covariance du vecteur aléatoire  $(\bar{Q}, \bar{R})^T$ . Cette métrique tient compte du fait que  $\bar{Q}$  et  $\bar{R}$  n'ont pas la même échelle et qu'ils sont corrélés.

La figure 4.4 résume les différentes étapes de la méthodologie proposée.

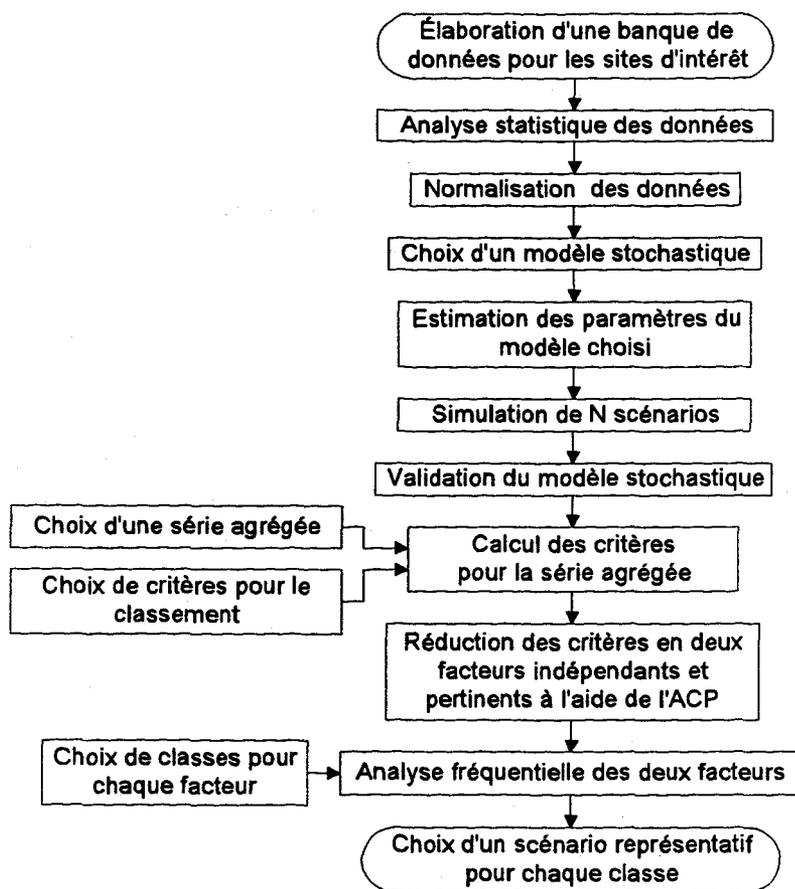


Figure 4.4 Étapes de la méthodologie proposée

## 5 Exemple d'application: Grande Baleine

### 5.1 Description du système hydrique de la rivière Grande Baleine

Le système hydrique de la rivière Grande Baleine représente un des cas où un aménagement hydroélectrique pourrait être envisagé dans un avenir rapproché. En effet, depuis plusieurs années, Hydro-Québec étudie la possibilité de construire trois centrales sur ce bassin versant dont le potentiel est d'environ 3 GW.

Le complexe Grande Baleine est formé de deux cours d'eau principaux, la Grande rivière de la Baleine et la Petite rivière de la Baleine, qui coulent d'est en ouest et se jettent dans le Baie d'Hudson. La superficie totale du bassin versant est d'environ 40,000 km<sup>2</sup>. Un tiers de celui-ci est formé d'un plateau de nombreux lacs et rivières qui se termine par le lac Bienville.

Des données hydrométriques y sont acquises depuis 1960. Certaines valeurs manquantes ont été reconstituées à l'aide du logiciel REMUL développé à l'INRS-Eau (Caron et Guerrere, 1993). Pour l'étude de l'optimisation de l'aménagement des trois centrales du complexe Grande Baleine, cinq apports intermédiaires doivent être considérés. Les stations des sous-bassins ainsi que la localisation des centrales sont illustrées de manière schématique dans la figure 5.1.

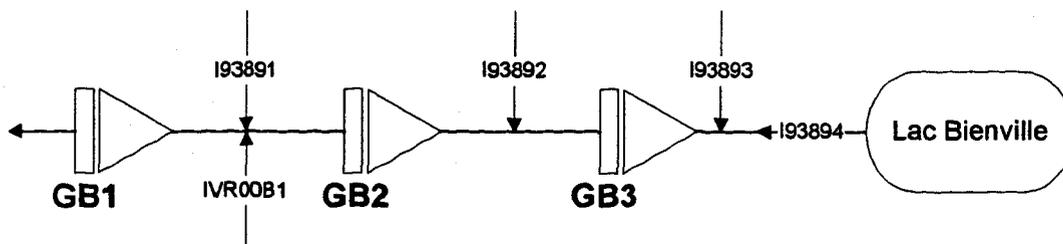


Figure 5.1 Apports intermédiaires du complexe Grande Baleine

Notons que pour l'aménagement d'un système hydrique, on considère généralement des séries mensuelles de 50 ans. Les séries pour le complexe Grande Baleine ont été reconstituées à l'aide d'un modèle pluie-débit. Cependant, dans cet exemple, nous considérons uniquement les données observées (ou reconstituées) pour la période de 1960 à 1991.

## 5.2 Génération de séries synthétiques à l'aide du modèle SPIGOT

Dans cet exemple, nous avons généré des séries synthétiques à l'aide du logiciel SPIGOT qui est basé sur le principe de désagrégation présenté dans la section 3.3.1. Tel que mentionné précédemment, on doit préférablement considérer plusieurs modèles et les comparer pour choisir le meilleur. Le modèle PARMA serait une alternative intéressante, mais cette présentation à titre d'illustration sera limitée à l'utilisation du logiciel SPIGOT<sup>1</sup>.

### 5.2.1 Application du modèle SPIGOT

Dans SPIGOT, on peut spécifier que le type de transformation doit être choisi de façon automatique. Le choix est basé sur la statistique de Filliben, c'est-à-dire le coefficient de corrélation entre les données transformées ordonnées et leur espérance mathématique (*probability plot correlation coefficient*, Filliben, 1975).

Le logiciel SPIGOT offre plusieurs options pour la désagrégation. Nous avons retenu le schéma 3 dans la figure 3.1, qui est la plus adéquate pour le cas actuel. Selon cette procédure, on génère d'abord les apports annuels aux cinq sites à l'aide d'un modèle AR(1) multivarié. SPIGOT permet de tester de façon automatique si l'autocorrélation pour un site donné est significative (critère Akaike). Aucune des cinq séries d'apports intermédiaires ne montre une autocorrélation significative. En conséquence, on utilise un modèle AR(0) multivarié pour la génération des séries annuelles. Notons que SPIGOT permet aussi de choisir par défaut des termes autorégressifs dans le modèle annuel.

Les apports annuels aux cinq sites sont ensuite désagrégés en apports mensuels. Encore là, SPIGOT offre l'option, pour chaque site et pour chaque mois, de choisir automatiquement si un terme autorégressif doit être considéré. Généralement, l'autocorrélation d'ordre un est significative,

---

<sup>1</sup>Actuellement, il n'existe pas de logiciel commercial pour le modèle PARMA multivarié. Le développement d'un logiciel, basé sur la programmation avec MATLAB, est actuellement en cours à l'INRS-Eau.

ce qui conduit au choix d'une structure AR(1) pour les séries mensuelles. Notons cependant que la corrélation entre l'apport du premier mois de l'année et celui du dernier mois de l'année précédente n'est pas préservée par SPIGOT.

À titre d'illustration de l'approche proposée et plus particulièrement pour faciliter l'interprétation de certains graphiques, nous avons décidé de générer cent séries. Dans le cas d'une application réelle, ce nombre serait probablement insuffisant.

### 5.2.2 Validation

Le tableau 5.1 donne quelques caractéristiques statistiques de base des séries d'apports annuels observées et générées.

Tableau 5.1 Caractéristiques statistiques des séries d'apports intermédiaires historiques et synthétiques pour le complexe Grande Baleine

	IVR00B1		I93891		I93892		I93893		I93894	
	HIS	SIM	HIS	SIM	HIS	SIM	HIS	SIM	HIS	SIM
Moyenne	2275	2274	2271	2274	369	368	948	944	2870	2875
Écart type	202	200	262	269	47	47	147	147	321	352
Corr. d'ordre 1	0.13	0	-0.01	0	-0.08	0	0.24	-0.04	0.16	-0.04
Coef. Hurst	0.62	0.63	0.54	0.64	0.51	0.64	0.80	0.62	0.69	0.63
RL moyen	2.25	2.03	1.75	2.12	1.44	1.99	2.66	2.06	2.00	2.51
RS moyen	297	312	376	433	57	75	315	227	445	562
Corr. croisée										
IVR00B1	1.00	1.00								
I93891	0.73	0.72	1.00	1.00						
I93892	0.58	0.60	0.95	0.95	1.00	1.00				
I93893	0.42	0.44	0.13	0.13	0.12	0.12	1.00	1.00		
I93894	0.39	0.45	0.14	0.19	0.09	0.16	0.74	0.72	1.00	1.00

On peut noter que la moyenne et l'écart-type pour les cinq sites sont généralement bien préservés par le modèle. Pour l'autocorrélation d'ordre 1, nous avons utilisé les valeurs théoriques du modèle pour la fin de comparaison (0). On note que l'autocorrélation d'ordre 1 observée est en générale

proche de zéro, ce qui montre que le modèle est adéquat au moins pour cette caractéristique. En ce qui concerne le coefficient de Hurst et les caractéristiques des séquences, on observe certains écarts. Ceci n'est pas surprenant, car ces caractéristiques sont implicites, i.e. aucun effort n'est fait pour les préserver explicitement. Si l'on n'est pas satisfait avec la reproduction de ces caractéristiques, on peut essayer un autre modèle et choisir celui qui donne les meilleurs résultats. Finalement, en ce qui concerne les corrélations croisées, on peut noter que le modèle les reproduit très bien.

En résumé, les résultats obtenus ici pour les apports annuels avec SPIGOT semblent assez bons. On peut effectuer une étude similaire des caractéristiques statistiques des données mensuelles, si celles-ci sont considérées importantes.

## 5.3 Identification de scénarios types

### 5.3.1 Séries agrégées

La première étape de l'identification de scénarios types consiste à choisir le type de série agrégée que l'on veut considérer (voir section 4.2). C'est à partir de ces séries que le choix des scénarios sera fait. Pour évaluer l'impact du choix de série agrégée représentative, nous avons retenu non pas un mais bien deux types de séries. Plus précisément, nous considérons la série des apports intermédiaires agrégés et une série du type énergétique, correspondant respectivement aux options 1 et 3 décrites dans la section 4.2. Notons que nous utilisons le terme *série énergétique* de façon générique. Il ne s'agit pas strictement d'une série énergétique, mais d'une série dont la variation est semblable à celle d'une véritable série énergétique. Plus précisément, la série que nous considérons a été obtenue en multipliant l'apport actuel à chaque centrale avec la hauteur des chutes. Les hauteurs sont données dans le tableau ci-dessous.

**Tableau 5.2** Hauteur de chute pour les centrales du complexe Grande Baleine

Centrale	Hauteur des chutes
GB1	185 m
GB2	91.6 m
GB3	96.1 m

### **5.3.2 Sélection de critères**

Lors du choix de critères à considérer pour le classement des scénarios, on doit se rappeler que la dimension des observations pour chaque série (nombre de critères) sera réduite à deux facteurs par l'analyse en composantes principales. Ainsi, aucun effort n'est fait ici pour identifier des critères redondants. En effet, il peut arriver que deux critères décrivent sensiblement la même caractéristique d'une série. Un des avantages de l'ACP est que les variables redondantes apparaîtront clairement. La philosophie adoptée ici est alors de choisir un nombre assez grand de critères et de les condenser en deux facteurs. Il est évident que les critères doivent être pertinents au problème étudié.

Pour discriminer les séries simulées, nous avons retenu les onze critères suivants:

- Moyenne (E)
- Écart-type (S)
- Longueur maximum des séquences de déficit par rapport à la moyenne historique (MAX\_RL)
- Longueur moyenne des séquences de déficit par rapport à la moyenne historique (E\_RL)
- Écart-type de la longueur des séquences de déficit par rapport à la moyenne historique (S\_RL)
- Maximum des déficits accumulés par rapport à la moyenne historique (MAX\_RS)
- Moyenne des déficits accumulés par rapport à la moyenne historique (E\_RS)
- Écart-type des déficits accumulés par rapport à la moyenne historique (S\_RS)
- Domaine de variation ajusté (AR)
- Domaine de variation ajusté standardisé (RAR)
- Coefficient de Hurst (H)

Une présentation plus détaillée de ces critères a été faite à la section 4.3.

### **5.3.3 Résultats pour les séries d'apports intermédiaires agrégés**

Pour chaque série agrégée d'apports intermédiaires générée, nous avons calculé les caractéristiques mentionnées dans la section précédente. Pour l'analyse suivante, nous avons fait appel au logiciel STATISTICA. Le tableau 5.3 donne quelques statistiques de base de ces caractéristiques. Le

Tableau 5.3 Caractéristiques statistiques des critères pour les séries synthétiques d'apports intermédiaires

STAT. BASIC STATS	Descriptive Statistics (option1.sta)					
	Variable	Mean	Median	Minimum	Maximum	Std.Dev.
MAX_RL	4.54	4.00	2.00	11.00	1.86	1.07
E_RL	2.11	2.00	1.10	4.40	.61	1.04
S_RL	1.32	1.18	.32	3.58	.66	1.11
MAX_RS	2840.6	2633.3	579.07	7250.1	1188.8	1.25
E_RS	1184.6	1131.6	300.81	2254.7	390.46	.67
S_RS	900.50	841.64	215.49	2367.1	409.18	1.24
E	8736.0	8741.6	8423.0	9341.8	163.62	.56
S	767.84	757.39	496.09	1132.6	114.69	.34
AR	4389.9	4217.0	2507.7	7561.6	1157.7	.53
RAR	5.71	5.55	3.33	9.03	1.18	.46
H	.63	.63	.44	.80	.08	-.07

Tableau 5.4 Matrice de corrélation des critères (séries d'apports intermédiaires)

STAT. BASIC STATS	Correlations (option1.sta)										
	Marked correlations are significant at $p < .05000$ N=100 (Casewise deletion of missing data)										
Variable	MAX_RL	E_RL	S_RL	MAX_RS	E_RS	S_RS	E	S	AR	RAR	H
MAX_RL	1.00	.77*	.98*	.79*	.76*	.82*	-.62*	-.22*	.10	.27*	.29*
E_RL	.77*	1.00	.79*	.59*	.87*	.63*	-.67*	-.32*	-.02	.19	.21*
S_RL	.98*	.79*	1.00	.78*	.77*	.82*	-.59*	-.20*	.12	.29*	.30*
MAX_RS	.79*	.59*	.78*	1.00	.77*	.97*	-.64*	-.06	.19	.28*	.29*
E_RS	.76*	.87*	.77*	.77*	1.00	.78*	-.77*	-.13	.08	.19	.21*
S_RS	.82*	.63*	.82*	.97*	.78*	1.00	-.60*	-.05	.20*	.29*	.30*
E	-.62*	-.67*	-.59*	-.64*	-.77*	-.60*	1.00	.49*	.30*	.03	.02
S	-.22*	-.32*	-.20*	-.06	-.13	-.05	.49*	1.00	.59*	.05	.04
AR	.10	-.02	.12	.19	.08	.20*	.30*	.59*	1.00	.83*	.82*
RAR	.27*	.19	.29*	.28*	.19	.29*	.03	.05	.83*	1.00	.99*
H	.29*	.21*	.30*	.29*	.21*	.30*	.02	.04	.82*	.99*	1.00

tableau 5.4 montre la corrélation entre les différents critères. On remarque que certaines corrélations sont assez élevées. Par exemple, la corrélation entre RAR et H est de 0.99 et la corrélation entre MAX\_RL et S\_RL est de 0.98. On note également une corrélation négative assez élevée entre la moyenne E et les statistiques de séquences. Ce résultat n'est pas surprenant

compte tenu du fait que les séquences ont été définies par rapport à la moyenne historique. Il est évident que les onze variables considérées contiennent de l'information redondante.

La matrice de corrélation des critères a trois valeurs propres supérieures à un qui seront toutes considérées dans un premier temps. Les deux et trois premières composantes principales expliquent respectivement 75% et 89% de la variance totale. Le tableau 5.5 montre les coordonnées des onze critères sur les trois premiers axes factoriels. Notons que nous avons effectué une rotation des axes originaux selon le principe *varimax*. Le but de la rotation est de faciliter l'interprétation des facteurs. Pour le premier facteur, les coordonnées des statistiques de séquences et la moyenne sont supérieures à 0.7 (marqué avec un astérisque \*). Ce facteur est clairement relié à l'apport moyen. Sur le deuxième axe factoriel, ce sont les variables AR, RAR et H qui ont des coordonnées élevées. Ce facteur est donc relié au domaine de variation ou au besoin de stockage. En ce qui concerne le troisième facteur, c'est l'écart-type S qui a un poids important.

Aux fins d'interprétation, il est commode d'illustrer graphiquement la position des variables dans les espaces factoriels. La figure 5.2a-c montre les onze points dans l'espace des trois premiers facteurs. Dans les trois graphiques, on remarque le regroupement des statistiques des séquences. Ce regroupement indique que ces caractéristiques mesurent essentiellement la même chose. On voit également comment le facteur deux distingue les critères reliés au domaine de variation des autres critères.

Il semble que les deux premiers facteurs couvrent assez bien les deux aspects des séries qui nous intéressent: la quantité d'eau et la répartition de l'eau. Pour cette raison, nous retenons les deux premiers facteurs qui expliquent environ 75% de la variabilité totale.

Le tableau 5.6 montre les coefficients que l'on doit appliquer aux variables pour calculer les facteurs (*factor score coefficients*) pour les séries synthétiques. On remarque que plus le volume est faible, plus la valeur du premier facteur est élevée et que plus le domaine de variation (*range*) est grand, plus le deuxième facteur est grand. Du point de vue de la production d'hydroélectricité, on souhaite une grande quantité d'eau et un faible besoin de stockage interannuel. Le cas de faibles valeurs de chacun des deux facteurs représente conséquemment un scénario optimiste par rapport à la moyenne. Si pour une série donnée, les deux facteurs sont relativement élevés, on parle alors d'un scénario pessimiste.

Tableau 5.5 Coordonnées des critères sur les axes factoriels (séries d'apports intermédiaires)

STAT. FACTOR ANALYSIS	Factor Loadings (Varimax normalized) (option1.sta) Extraction: Principal components (Marked loadings are > .700000)		
Variable	Factor 1	Factor 2	Factor 3
MAX_RL	.908 *	.159	.116
E_RL	.817 *	.085	.293
S_RL	.907 *	.175	.101
MAX_RS	.908 *	.139	-.100
E_RS	.915 *	.055	.066
S_RS	.922 *	.150	-.112
E	-.753 *	.171	-.434
S	-.125	.084	-.967 *
AR	.046	.838 *	-.530
RAR	.144	.986 *	.012
H	.157	.983 *	.026
Expl.Var	5.458	2.784	1.542
Prp.Totl	.496	.253	.140

Tableau 5.6 Coefficients pour calculer les facteurs (séries d'apports intermédiaires)

STAT. FACTOR ANALYSIS	Factor Score Coefficients (option1.sta) Rotation: Varimax normalized Extraction: Principal components		
Variable	Factor 1	Factor 2	Factor 3
MAX_RL	.1660	.0039	-.0022
E_RL	.1293	.0149	.1336
S_RL	.1664	.0079	-.0107
MAX_RS	.1972	-.0456	-.1741
E_RS	.1841	-.0502	-.0617
S_RS	.2009	-.0445	-.1833
E	-.1205	.0611	-.2035
S	.0958	-.1366	-.7195
AR	.0026	.2510	-.2589
RAR	-.0635	.4078	.1775
H	-.0622	.4081	.1860

À l'aide de MATLAB, nous avons généré la figure 5.3 qui montre les séries dans l'espace factoriel. Les numéros sont les indices des séquences simulées. Par exemple, "6" est la sixième séquence générée de 31 ans. Les lignes horizontales et verticales sont des bornes qui, pour chaque facteur, divise l'espace en trois classes empiriquement équiprobables. On s'attend à observer environ 11 séries dans chaque classe puisque les facteurs sont indépendants.

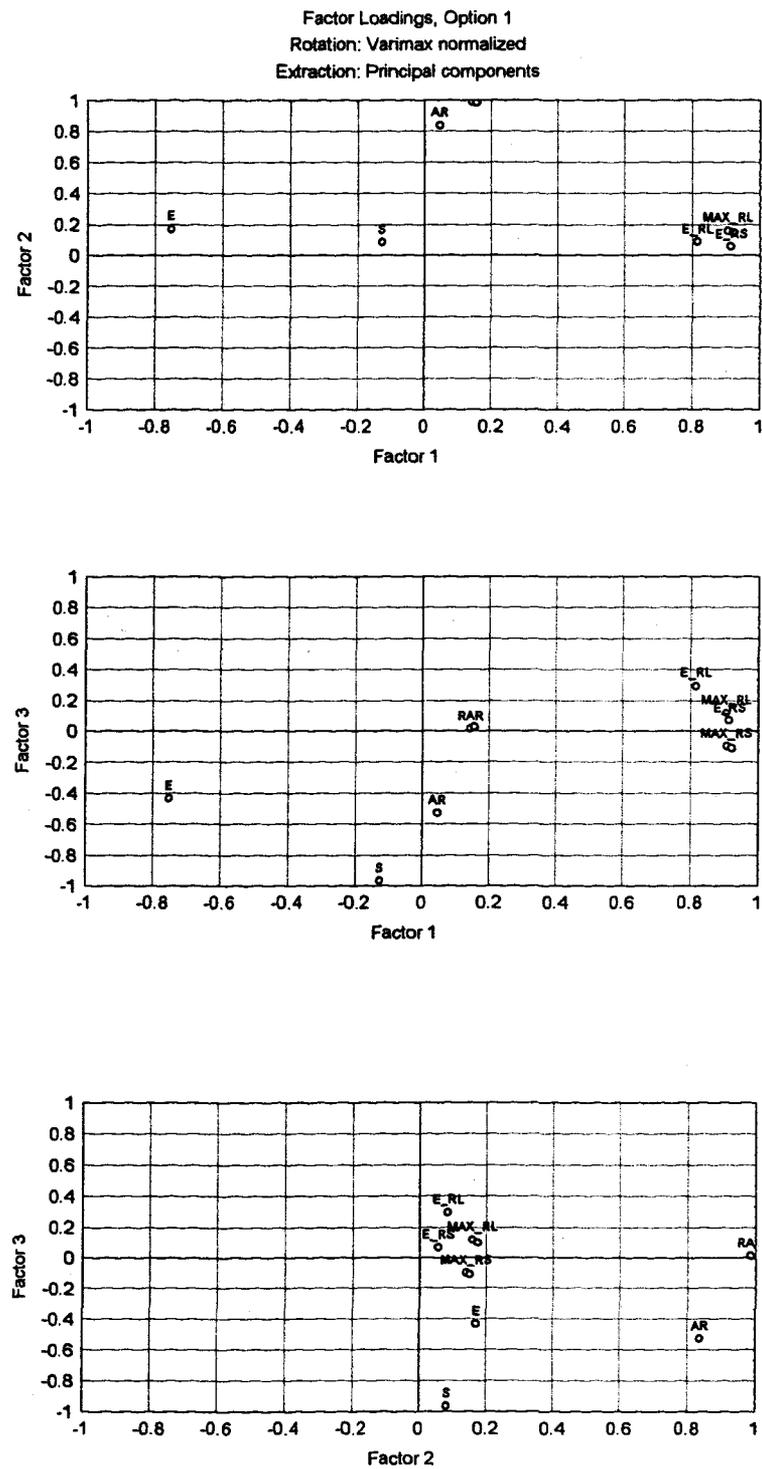


Figure 5.2 Position des critères dans l'espace des trois premiers facteurs (apport intermédiaire)

Tableau 5.8 Matrice de corrélation des critères (séries énergétiques)

STAT. BASIC STATS	Correlations (option3.sta)										
	Marked correlations are significant at $p < .05000$ N=100 (Casewise deletion of missing data)										
Variable	MAX_RL	E_RL	S_RL	MAX_RS	E_RS	S_RS	E	S	AR	RAR	H
MAX_RL	1.00	.71*	.96*	.83*	.66*	.84*	-.45*	-.08	.19	.30*	.30*
E_RL	.71*	1.00	.73*	.60*	.89*	.63*	-.55*	-.18	.10	.29*	.29*
S_RL	.96*	.73*	1.00	.83*	.69*	.88*	-.43*	-.08	.19	.31*	.30*
MAX_RS	.83*	.60*	.83*	1.00	.72*	.95*	-.54*	-.04	.22*	.32*	.33*
E_RS	.66*	.89*	.69*	.72*	1.00	.71*	-.67*	-.13	.11	.26*	.26*
S_RS	.84*	.63*	.88*	.95*	.71*	1.00	-.48*	-.03	.26*	.36*	.35*
E	-.45*	-.55*	-.43*	-.54*	-.67*	-.48*	1.00	.60*	.40*	.03	.02
S	-.08	-.18	-.08	-.04	-.13	-.03	.60*	1.00	.66*	.06	.07
AR	.19	.10	.19	.22*	.11	.26*	.40*	.66*	1.00	.78*	.78*
RAR	.30*	.29*	.31*	.32*	.26*	.36*	.03	.06	.78*	1.00	.99*
H	.30*	.29*	.30*	.33*	.26*	.35*	.02	.07	.78*	.99*	1.00

Tableau 5.9 Coordonnées des critères sur les axes factoriels (séries énergétiques). a) tous les critères; b) tous les critères moins l'écart-type S

STAT. FACTOR ANALYSIS	Factor Loadings (option3.sta)		
	Factor 1	Factor 2	Factor 3
MAX_RL	.918 *	.119	.008
E_RL	.790 *	.168	.256
S_RL	.933 *	.126	.006
MAX_RS	.913 *	.144	.011
E_RS	.832 *	.133	.244
S_RS	.923 *	.176	-.023
E	-.559	.116	-.712 *
S	-.014	.061	-.952 *
AR	.137	.766 *	-.609
RAR	.189	.979 *	-.020
H	.187	.978 *	-.022
Expl.Var	5.119	2.646	1.911
Prp.Totl	.465	.241	.174

STAT. FACTOR ANALYSIS	Factor Loadings (option 3)	
	Factor 1	Factor 2
MAX_RL	.887 *	.188
E_RL	.839 *	.107
S_RL	.900 *	.198
MAX_RS	.888 *	.202
E_RS	.880 *	.072
S_RS	.888 *	.248
E	-.724 *	.344
AR	.002	.935 *
RAR	.200	.930 *
H	.198	.929 *
Expl.Var	5.255	2.913
Prp.Totl	.526	.291

Le tableau 5.10 donne les coefficients que l'on doit appliquer pour calculer les facteurs. La figure 5.5 montre les séries générées dans l'espace factoriel. Il y a une certaine similarité entre les figures 5.3 et 5.5. Les séquences 6, 1, 56, 90, 80, 93, 78, 36, 27 sont choisies comme étant représentatives pour chaque classe.

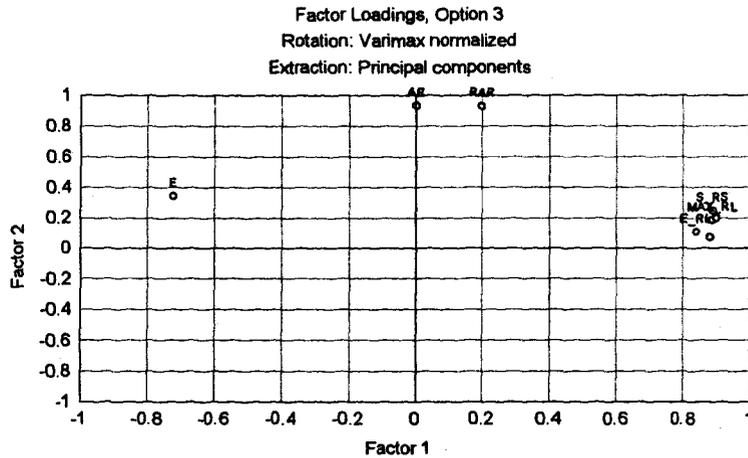


Figure 5.4 Position des critères dans l'espace de deux facteurs (série énergétique)

Tableau 5.10 Coefficients pour calculer les facteurs (séries énergétiques)

STAT. FACTOR ANALYSIS	Factor Score Coefficients (option3.sta)	
	Rotation: Varimax normalized Extraction: Principal components	
Variable	Factor 1	Factor 2
MAX_RL	.1677	.0059
E_RL	.1636	-.0207
S_RL	.1696	.0085
MAX_RS	.1669	.0108
E_RS	.1745	-.0363
S_RS	.1635	.0280
E	-.1724	.1783
AR	-.0665	.3442
RAR	-.0256	.3283
H	-.0260	.3278

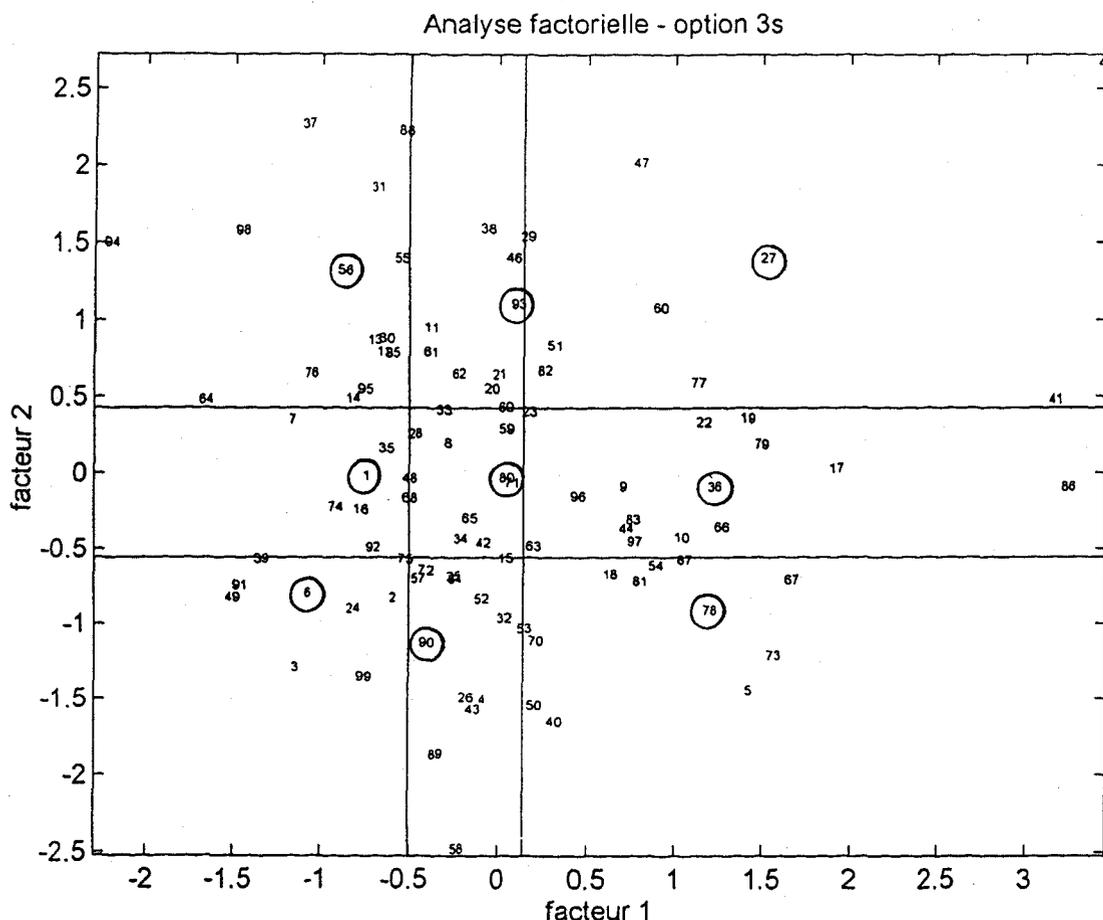


Figure 5.5 Position des séries énergétiques synthétiques dans l'espace des deux premiers facteurs

## 5.4 Discussion des résultats

L'utilisation des séries d'apports intermédiaires agrégés et des séries énergétiques ne conduit pas au même choix de scénarios types. Les séquences 6 (scénario optimiste) et 80 (scénario moyen) se répètent dans les deux cas (et pour les mêmes raisons), mais les autres sont différentes. Il est donc important de faire un choix entre les deux types de séries. Dans cet exemple, nous retiendrons les résultats basés sur les séries énergétiques. Ce type de série nous semble plus approprié compte tenu de l'objectif de l'étude. Toutefois, le choix est subjectif et demeure la responsabilité de l'utilisateur.

Il serait intéressant d'examiner de façon visuelle le caractère des scénarios retenus. La figure 5.6 montre la séquence de sommes partielles ajustées avec la moyenne à long terme (moyenne de toutes les simulations) pour chacune des séries énergétiques retenues. Ce type de graphique a été

choisi parce qu'il permet de bien présenter les caractéristiques générales des séries. On remarque que les séquences 1, 6 et 56 représentent des scénarios à moyenne élevée, les séquences 80, 93 et 90, des scénarios dont la moyenne est relativement proche de la moyenne historique, et les séquences 36, 27 et 78, des scénarios à moyenne faible. Bien que la séquence 56 ait une moyenne qui est supérieure à celle de la séquence 6, cette dernière est considérée comme plus optimiste parce que la répartition de l'énergie sur toute la période est plus homogène. En effet, pour la séquence 56, on observe quelques années au début de la série pendant lesquelles l'apport est significativement au-dessus de la moyenne historique. Ces années influencent fortement la moyenne générale de la séquence. Il n'est guère rentable d'emmagasiner tout le volume de cette période humide. On peut noter que la moyenne sur les vingt dernières années est comparable à la moyenne historique. En réalité, la séquence 56 pourrait être moins favorable qu'elle n'apparaît, car pendant plusieurs années, l'apport est inférieur à la moyenne historique. On pourrait envisager l'utilisation de critères plus élaborés qui tiennent compte, par exemple, d'une limite supérieure approximative du stockage (réservoir semi-infini ou fini).

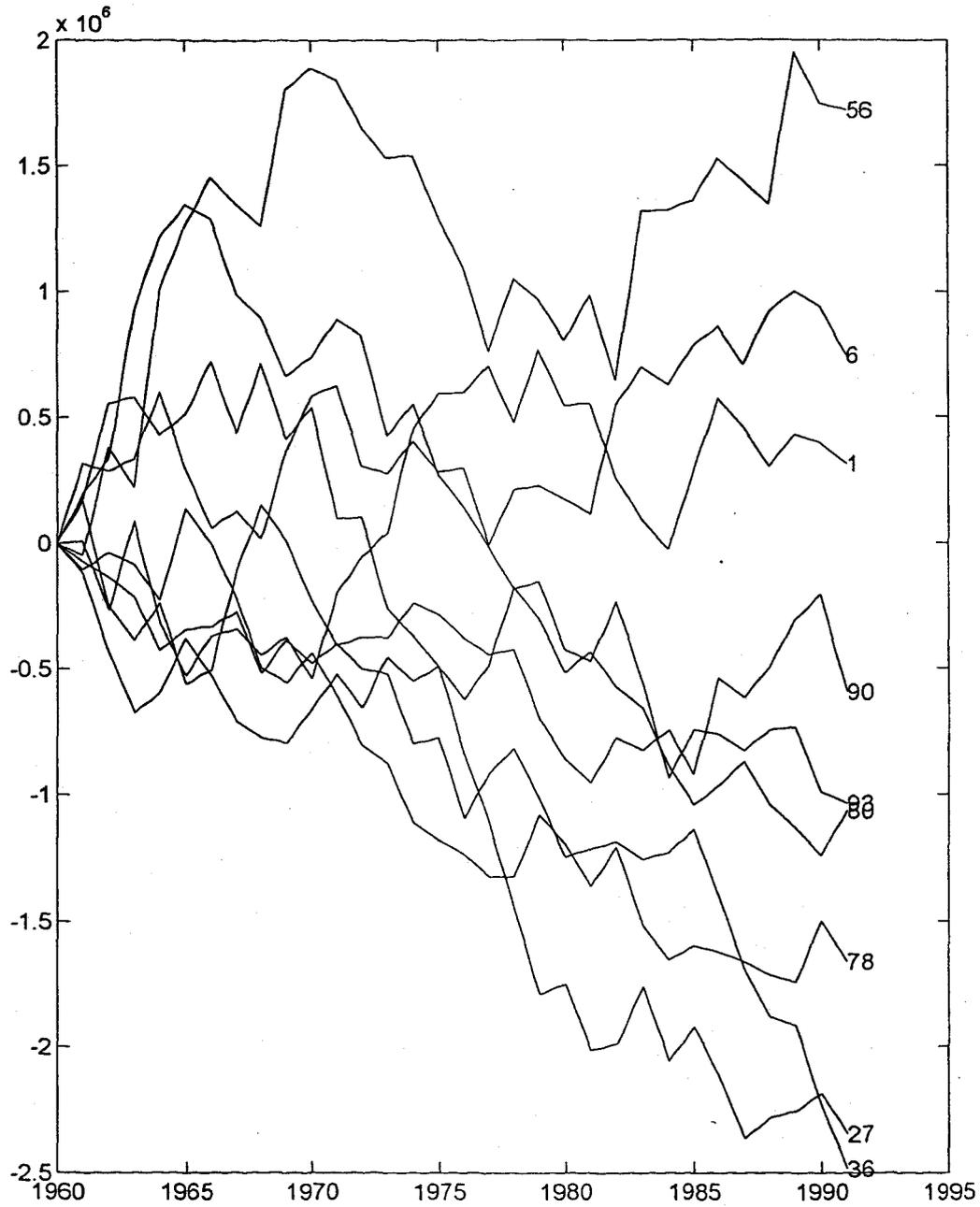


Figure 5.6 Sommes partielles ajustées avec la moyenne à long terme pour neuf scénarios types

## 6 Conclusions

---

Le présent projet s'inscrit dans le cadre du développement du modèle MINERVE, une version améliorée du modèle ATHENA. Ces modèles permettent de déterminer l'aménagement optimal d'un nouveau bassin à des fins de production d'hydroélectricité. Dans ATHENA, seuls les apports historiques sont considérés pour déterminer l'aménagement optimal. L'objectif spécifique du sous-projet présenté dans ce rapport a été d'élaborer une méthodologie pour la détermination de séries mensuelles types couvrant une gamme réaliste de scénarios futurs possibles. Ces séries seront ensuite utilisées dans MINERVE dans l'optique d'introduire l'aspect aléatoire des apports naturels dans la détermination de l'aménagement optimal.

Idéalement, le choix de scénarios types devrait se faire de façon automatique. Cependant, ce travail a mis en évidence que, pour faire un choix adéquat de scénarios types, l'utilisateur doit prendre un nombre de décisions de caractère subjectif qui peuvent difficilement être automatisées. Plus particulièrement, ces décisions concernent:

- *La modélisation stochastique des données historiques*

L'estimation d'un modèle stochastique pourrait, en principe, se faire de manière automatique. Un bon exemple est le logiciel SPIGOT qui donne de nombreuses options pour une modélisation automatisée. Toutefois, cette procédure est rarement avantageuse, chaque système hydrique étant unique et représentant des particularités dont on doit tenir compte lors de la modélisation. Il est généralement préférable de comparer plusieurs modèles afin de choisir le meilleur.

- *Le choix de critères*

Les séries synthétiques doivent être discriminées et classées selon des critères. Il est évident que le choix de critères est extrêmement important pour le classement des séries. Préférentiellement, les critères seront choisis en fonction du problème étudié. Certaines caractéristiques statistiques d'étiage, comme le domaine de variation par exemple, sont traditionnellement définies pour les

réservoirs infinis, mais il pourrait s'avérer plus approprié de considérer des réservoirs semi-infinis ou finis.

• *L'analyse en composantes principales*

L'interprétation des résultats d'une analyse en composantes principales ne peut être automatisée. Il est important que l'analyse conduise à des facteurs auxquels on peut associer un sens physique. Il peut s'avérer avantageux d'exclure certaines variables de l'analyse pour obtenir une meilleure représentation des critères dans l'espace factoriel.

• *Le choix de classes*

Le choix de classes est probablement l'élément le plus crucial dans la méthodologie proposée. Dans notre exemple, nous avons considéré des classes équiprobables dans l'espace factoriel. Toutefois, il n'est pas nécessaire que les classes soient équiprobables. Il pourrait s'avérer avantageux de retenir des scénarios plus extrêmes. Bien entendu, plus un scénario est extrême, plus la probabilité (poids) qui lui est associée est faible. La meilleure façon d'évaluer l'impact du choix de classes est d'analyser la sensibilité des résultats produits par MINERVE en fonction de différents choix de classes/probabilités. Ce dernier aspect sera étudié dans la deuxième phase du projet.

## 7 Bibliographie

---

- Acres International Limited (1990) Évaluation des modèles de débit mensuel pour l'exploitation et la planification de la production hydro-électrique, Rapport no. 720 G 682, Acres International Limited, Niagara Falls, Ontario, 133 pp.
- Bartolini, P. et Salas, J.D. (1993) Modeling of streamflow processes at different time scales. *Water Resour. Res.*, 29(8): 2573-2587
- Bartolini, P., Salas, J.D., et Obeysekera, J.T.B. (1988) Multivariate periodic ARMA(1,1) processes. *Water Resour. Res.*, 24(8): 1237-1246
- Bras, R.L. et Rodriguez-Iturbe, I. (1985) *Random Functions and Hydrology*. Addison-Wesley, Reading Mass.
- Box, G.E.P. et Cox, D.R. (1964) An analysis of transformations. *J. Roy. Statist. Soc.*, B26: 211-243
- Camacho, F., McLeod, A.I., et Hipel, K.W. (1985) Contemporaneous autoregressive-moving average (CARMA) modeling in water resources. *Water Resour. Bull.*, 21(4): 709-720
- Camacho, F., McLeod, A.I. et Hipel, K.W. (1987) Multivariate contemporaneous ARMA models with hydrological applications. *Stochastic Hydrol. Hydraul.*, 1: 141-154.
- Caron, P. et Cuerrere, A. (1993) Complexe Grande Baleine, Harmonisation des séries d'apports mensuels avant et après aménagement pour la période de 1942 à 1991. Rapport technique, Direction aménagements de centrales, Service Hydraulique, Hydro-Québec, Montréal
- Dracup, J.A., Lee, K.S. et Paulson, E.G. (1980a) Statistical characteristics of drought events, *Water Resour. Res.*, 16(2): 289-296
- Dracup, J.A., Lee, K.S. et Paulson, E.G. (1980b) On the definition of droughts, *Water Resour. Res.*, 16(2): 297-302
- Duquette, R. et Lafond, L. (1992) Une nouvelle approche d'aménagement optimal d'un bassin hydraulique pour fin de production d'électricité, Rapport IREQ-92-171.
- Feller, W. (1951) The asymptotic distribution of the range of sums of independent random variables, *Ann. Math. Stat.*, 22: 427-432
- Fernandez, B. et Salas, J.D. (1986) Periodic gamma autoregressive processes for operational hydrology. *Water Resour. Res.*, 22(10): 1385-1396

- Fernandez, B. et Salas, J.D. (1990) Gamma-autoregressive models for stream-flow simulation. *J. Hydraul. Eng. (ASCE)*, 116(11): 1403-1414
- Filliben, J.J. (1975) The probability plot correlation coefficient test for normality, *Technometrics*, 17(1): 111-117
- Frick, D.M., Bode, D. et Salas, J.D. (1990) Effect of drought on urban water supplies. I: Drought analysis, *J. Hydraul. Eng., ASCE*, 116(6): 733-753
- Grygier, J.C. et Stedinger, J.R. (1988) Condensed disaggregation procedures and conservation corrections for stochastic hydrology. *Water Resour. Res.*, 24(10): 1574-1584
- Grygier, J.C. et Stedinger, J.R. (1990) SPIGOT. A Synthetic Streamflow Generation Software Package, Technical Description, Version 2.6, Cornell University, Ithaca, New York
- Grygier, J.C. et Stedinger, J.R. (1991) SPIGOT. A Synthetic Streamflow Generation Software Package, User's Manual, Version 2.6, Cornell University, Ithaca, New York
- Haltiner, J.P. et Salas, J.D. (1988) Development and testing of a multivariate, seasonal ARMA(1,1) model. *J. Hydrol.*, 104: 247-272
- Hurst, H.E. (1951) Long-term storage capacity of reservoirs, *Trans, ASCE*, 116: 770-808
- Koutsoyiannis, D. (1992) A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic models. *Water Resour. Res.*, 28(12): 3175-3191
- Lane, W.L. (1979) Applied Stochastic Techniques: User manual. Div of Plann, Tech. Serv., Water and Power Resour. Serv., Denver, Colo.
- Lin, G.F. (1990) Parameter estimation for seasonal to subseasonal disaggregation. *J. Hydrol.*, 120: 65-77
- Lin, G.-F. et Lee, F.-C. (1992) An aggregation - disaggregation approach for hydrologic time series modelling. *J. Hydrol.*, 543-557
- McLeod, A.I. et Hipel, K.W. (1978) Preservation of the rescaled adjusted range, 1, A reassessment of the Hurst phenomenon, *Water Resour. Res.*, 14(3): 491-508
- Mejia, J.M. et Rousselle, J. (1976) Disaggregation models in hydrology revisited. *Water Resour. Res.*, 12(2): 185-186
- Moran, P.A.R. (1959) *The Theory of Storage*, Methuen, London
- Pegram, G.G.S. et James, W. (1972) Multilag multivariate autoregressive model for the generation of operational hydrology. *Water Resour. Res.*, 8(4): 1074-1076
- Pegram, G.G.S., Salas, J.D., Boes, D.C. et Yevjevich, V. (1980) *Stochastic Properties of Water Storage*, Hydrology Papers 100, Colorado State University, Fort Collins, CO
- Rasmussen, P.F. (1995a) Preservation of variance and autocorrelation in synthetic streamflow series, article présenté au congrès "Statistical and Bayesian methods in Hydrological Sciences", UNESCO, Paris, septembre 1995

- Rasmussen, P.F., Mathier, L., Salas, J.D., Fagherazzi, L. et Rassam, J.-C. (1995b) Calibration of a multivariate PARMA model for the Ottawa River system, Rapport de recherche R-432, INRS-Eau, Université du Québec, Ste-Foy, 77p
- Rasmussen, P.F., Salas, J.D., Fagherazzi, L., Rassam, J.-C. et Bobée, B. (1996) Calibration and testing of contemporaneous PARMA models for streamflow simulation, *Water Resour. Res.*, sous presse
- Santos, E.G. et Salas, J.D. (1992) Stepwis disaggregation scheme for synthetic hydrology. *J. Hydraul. Eng. (ASCE)*, 118(5): 765-784
- Salas, J.D., J.W. Delleur, V. Yevjevich, et Lane, W.L. (1980) *Applied Modeling of Hydrologic Time Series*. Water Resources Publications, Littleton, Colo.
- Salas, J.D., Tabios, G.Q., et Bartolini, P. (1985) Approaches to multivariate modeling of water resources time series. *Water Resour. Bull.*, 21(4): 683-708
- Sim, C.H. (1987) A mixed gamma ARMA(1,1) model for river flow time series. *Water Resour. Res.*, 23(1): 32-36
- Stedinger, J.R., Lettenmaier, D.P., et Vogel, R.M. (1985a) Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. *Water Resour. Res.*, 21(4): 497-509
- Stedinger, J.R., Pei, D., et Cohn, T.A. (1985b) A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resour. Res.*, 21(5): 665-675
- Stedinger, J.R. et Taylor, M.R. (1982) Synthetic streamflow generation, 1, Model verification and validation. *Water Resour. Res.*, 18(4): 909-918
- Stedinger, J.R. et Vogel, R.M. (1984) Disaggregation procedures for generating serially correlated flow vectors. *Water Resour. Res.*, 20(1): 47-56
- Thomas, H.A. et Fiering, M.B. (1962) Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. *Design of Water Resources Systems* (Mass et al.), Harvard Univ. Press, Cambridge, Mass., pp. 459-493
- Todini, E. (1980) The preservation of skewness in linear disaggregation schemes. *J. Hydrol.*, 42: 199-214
- Valencia, R.D. et Schaake, J.C. (1973) Disaggregation processes in stochastic hydrology. *Water Resour. Res.*, 9(3): 580-585
- Vecchia, A.V. (1985) Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resour. Bull.*, 21(5): 721-730
- Wilson, E.B. et Hilferty, M.M. (1931) The distribution of chi-square. *Proceedings, Nat. Acad. Sci.*, New York, 17(12): 684-688