

---

# Les processus du type « ARIMA » pour la prévision et la simulation en hydrométéorologie

## « ARIMA » processes for hydrometeorological forecasting and simulation

J.W. Delleur

Professeur de Génie Hydraulique, Purdue University (USA)  
Détaché au Laboratoire National d'Hydraulique, Chatou

---

### Introduction

#### Introduction

Les modèles mathématiques en hydrométéorologie se classifient en modèles déterministes et en modèles stochastiques. Dans cette étude on considère une classe de processus stochastiques en temps discrets susceptibles de modéliser des séries chronologiques telles que celles des pluies, des débits ou des températures des rivières et d'autres phénomènes de l'environnement. Ces séries chronologiques se caractérisent par une forte dépendance entre observations successives. On supposera de plus qu'elles peuvent être représentées par des modèles linéaires, c'est-à-dire que la série  $Z_t$  peut être considérée comme étant la sortie d'un système linéaire dont l'entrée est un bruit blanc discret  $a_t$ , c'est-à-dire :

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (1)$$

En introduisant un opérateur de translation en arrière  $B$ , tel que  $BZ_t = Z_{t-1}$  et  $B^m Z_t = Z_{t-m}$  l'équation (1) se réécrit :

$$Z_t = \mu + \psi(B) a_t \quad (2)$$

$$\text{où : } \psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (3)$$

est la fonction de transfert du filtre.

Parmi ces modèles linéaires, le modèle autorégressif présente un intérêt particulier en hydrologie, car chaque valeur du processus est une somme pondérée et finie de valeurs antérieures de ce processus plus une composante aléatoire. Ce genre de processus s'adapte particulièrement bien aux étiages des rivières dont les débits sont alors dus à la vidange du stockage dans le bassin. Pour  $\mu = 0$ , ce modèle s'exprime :

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (4)$$

$$\text{ou encore : } \phi(B) Z_t = a_t$$

où  $\phi(B)$  est un polynôme en  $B$  de degré  $p$ , et le processus autorégressif d'ordre  $p$  est désigné par le symbole  $AR(p)$

Le processus dit de moyenne mobile représente la valeur actuelle de  $Z_t$  comme une somme pondérée et finie de  $q$  valeurs antérieures du bruit blanc :

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (5)$$

Il est désigné par le symbole  $MA(q)$ , de l'anglais *moving average*, et s'écrit :

$$Z_t = \theta(B) a_t \quad (6)$$

où  $\theta(B)$  est un polynôme en  $B$  de degré  $q$ .

La combinaison des deux processus donne lieu au processus mixte autorégressif d'ordre  $p$  et de moyenne mobile d'ordre  $q$  désigné par  $ARMA(p, q)$  qui s'écrit :

$$\left[ \sum_{j=0}^p \phi_j B^j \right] Z_t = \left[ \sum_{j=0}^q \theta_j B^j \right] a_t \quad (7)$$

$$\text{ou encore } \phi(B) Z_t = \theta(B) a_t \quad (8)$$

D'une manière heuristique on peut imaginer, par exemple, que les débits d'une rivière dépendent d'une part du tarissement (partie autorégressive) et d'autre part des apports aléatoires des pluies. On peut donc espérer que ce genre de modèles se prête à la représentation de débits des rivières.

Les débits comme les pluies annuelles sont, en général stationnaires, et ces modèles, en particulier les modèles autorégressifs, se prêtent à leur schématisation. Si on s'intéresse aux pluies ou débits mensuels, hebdomadaires ou journaliers on a affaire à des phénomènes non stationnaires. On résumera donc les conditions de stationnarité et d'inversibilité des processus ARMA qui sont stationnaires, et ensuite leur extension aux chroniques non stationnaires. On considérera principalement

ensuite les applications des modèles ARIMA aux pluies et débits mensuels et aux débits hebdomadaires et journaliers qui ont été étudiés à Purdue University.

### Stationnarité et inversibilité des processus ARMA

#### *Stationarity and invertibility of ARMA processes*

Un processus est strictement stationnaire si sa densité de probabilité est invariante à une translation dans le temps, et il est stationnaire du second ordre lorsque les deux premiers moments de la densité de probabilité du processus sont invariants aux translations dans le temps. En hydrologie on se contente souvent de la stationnarité du second ordre. Cette dernière implique la stationnarité stricte si on a une loi normale.

Considérons le processus autorégressif ARMA (1,0) donné par  $(1 - C_j B) Y_t = a_t$ . Alors :

$$\begin{aligned} Y_t &= C_j Y_{t-1} + a_t \\ Y_{t+1} &= C_j^2 Y_{t-1} + C_j a_t + a_{t+1} \\ Y_{t+2} &= C_j^3 Y_{t-1} + C_j^2 a_t + C_j a_{t+1} + a_{t+2}, \text{ etc } \quad (9) \end{aligned}$$

lorsque  $|C_j| > 1$  l'effet du passé sur la valeur présente croît au fur et à mesure que la série se déplace vers le futur. Lorsque  $|C_j| = 1$  l'effet du passé sur la valeur présente reste constant. Ces deux cas concrétisent l'expérience hydrologique selon laquelle l'effet du passé sur le présent décroît au fur et à mesure que la série se déplace dans le futur. Pour la stationnarité on doit avoir  $|C_j| < 1$ , ceci se généralise pour les processus AR.

Considérons le processus à moyenne mobile ARMA (0,1) donné par :

$$Y_t = (1 - C_i B) a_t$$

Si  $\theta(B) = (1 - C_i B)$  est inversible, on peut écrire :

$$a_t = \frac{1}{1 - C_i B} Y_t = (1 - C_i B + C_i^2 B^2 + \dots) Y_t \quad (10)$$

qui est une série divergente pour  $|C_i| > 1$ . Mais pour  $|C_i| < 1$  on peut écrire la série convergente :

$$\begin{aligned} \frac{1}{1 - C_i B} Y_t &= \left[ -\frac{1}{C_i B} - \frac{1}{C_i^2 B^2} - \dots \right] Y_t \quad (11) \\ &= -\frac{1}{C_i} Y_{t+1} - \frac{1}{C_i^2} Y_{t+2} - \frac{1}{C_i^3} Y_{t+3} - \dots \end{aligned}$$

selon laquelle la valeur actuelle de  $Y_t$  dépendrait des valeurs futures  $Y_{t+1}$ ,  $Y_{t+2}$ , ce qui n'est pas physiquement réalisable. Donc pour obtenir une génération physiquement réalisable il faut imposer la condition d'inversibilité  $|C_i| < 1$  qui se généralise pour les processus MA.

### Modèles stationnaires et non stationnaires

#### *Stationary and non-stationary models*

Comme nous l'avons déjà mentionné les chroniques hydrométéorologiques à des pas de temps plus petits que l'année ne sont pas stationnaires. Il faut donc leur faire subir une transformation pour les rendre stationnaires avant de pouvoir caler un modèle stationnaire du type ARMA. Ces non-stationnarités se traduisent par des pics dans le spectre de variance et par des oscillations périodiques dans la fonction d'autocorrélation. La figure 1 illustre le cas de la série des racines carrées des pluies mensuelles, dont les valeurs ont été divisées par la variance du mois correspondant de manière à obtenir une série à variance constante.

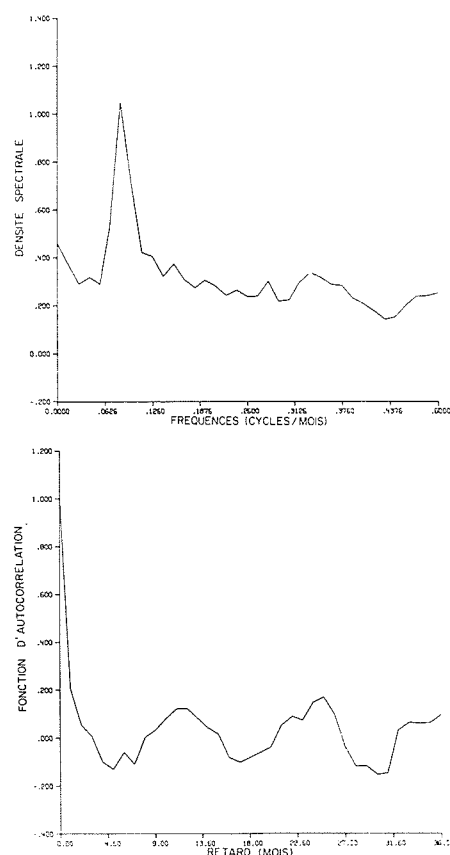


Figure 1 - Densité spectrale et fonction d'autocorrélation des racines carrées des pluies mensuelles (divisées par leurs variances mensuelles)

*Spectral density and autocorrelation function of squareroots of monthly rainfall (divided by monthly variance).*

Kavvas et Delleur [1975.b] ont étudié trois méthodes de filtrage des périodicités dans les séries mensuelles à variance constante : par différence non-saisonnnière (pas unité), par différence saisonnière (pas de 12) et par soustraction de la moyenne mensuelle, c'est-à-dire en formant la variable réduite. En décomposant la série originale en une composante circulairement stationnaire et une composante aléatoire stationnaire, ils ont évalué l'effet des différentes méthodes de filtrage sur les spectres. L'emploi de variables réduites élimine

essentiellement la périodicité dans la fonction d'autocovariance et donne une série qui satisfait aux conditions de stationnarité faible. La première différence  $\nabla Z_t = (1 - B) Z_t = w_t$  réduit d'une manière appréciable la composante périodique mais produit des distorsions dans la partie continue du spectre et dans la fonction de covariance. La différence saisonnière,  $\nabla_{12} Z_t = (1 - B^{12}) Z_t = v_t$ , élimine complètement la contribution périodique mais produit des oscillations sinusoïdales dans la partie stationnaire du spectre et introduit des distorsions dans la fonction de covariance. Les modèles ARMA ( $p, q$ ) appliqués aux séries filtrées par des différences sont appelés processus ARIMA ( $p, d, q$ ) de l'anglais *autoregressive integrated moving average*. Il faut effectuer des sommes infinies pour retrouver la série  $Z_t$ ; pour le cas d'une seule opération de différence on a :

$$\begin{aligned} Z_t &= S w_t = \nabla^{-1} w_t = (1 - B)^{-1} w_t \\ &= (1 + B + B^2 + \dots) w_t \end{aligned} \quad (12)$$

Les modèles ainsi obtenus ont une variance infinie et les modèles non-stationnaires du type ARIMA ne peuvent donc être employés dans la génération de séries hydrologiques synthétiques, mais peuvent néanmoins être utilisés avec succès pour la prévision. (Delleur, Tao et Kavvas, 1976).

La théorie des processus ARMA et ARIMA est développée en détail dans l'ouvrage de Box et Jenkins (1976) ainsi que les méthodes d'identification des modèles, d'estimation des paramètres et les tests de validité de l'ajustement. Dans ce qui suit on concentrera donc l'attention sur les applications aux séries hydro-météorologiques.

### Application aux pluies mensuelles

#### Application to monthly rainfall

Les modèles des types ARMA et ARIMA ont été appliqués par Kavvas et Delleur (1975) et par Delleur et Kavvas (1978) aux séries des pluies moyennes mensuelles sur 15 bassins situés au centre des Etats-Unis d'Amérique (Indiana, Illinois et Kentucky). Les bassins ont une superficie variant de 630 km<sup>2</sup> à 10 240 km<sup>2</sup>. Les pluies moyennes ont été obtenues par la méthode du polygone de Thiessen. La longueur des chroniques varie de 468 à 484 mois. On a pris la racine carrée de ces précipitations moyennes afin d'obtenir des séries approximativement gaussiennes.

#### Modèles non saisonniers

Sous cette rubrique on considère les modèles du type ARIMA appliqués directement aux séries des racines carrées des pluies mensuelles, et les modèles du type ARMA appliqués aux variables réduites :

$$Y_t = \frac{Z_t - \bar{Z}_i}{S_i} \quad (13)$$

où  $Z_t$  est la racine carrée des pluies au mois  $t$ , et  $\bar{Z}_i$  et  $S_i$  sont les moyennes et écarts types des échantillons du mois  $i$  où  $i = 1, \dots, 12$ .

L'identification des modèles se fait en examinant le comportement des fonctions d'autocorrélation et d'autocorrélation partielle. La fonction d'autocorrélation d'un processus AR( $p$ ) ressemble à une courbe de tarissement et la fonction d'autocorrélation partielle se termine au pas  $p$ , tandis que pour le processus MA( $q$ ) la fonction d'autocorrélation se termine au pas  $q$  et la fonction d'autocorrélation partielle décroît. Pour les modèles ARMA( $p, q$ ) la fonction d'autocorrélation présente une décroissance exponentielle accompagnée d'oscillations harmoniques après les premiers  $q - p + 1$  pas tandis que la fonction d'autocorrélation partielle décroît d'une manière exponentielle accompagnée d'oscillations.

Après examen des fonctions d'autocorrélation et d'autocorrélation partielle on a retenu les modèles ARIMA (1.1.1.) et ARMA (1.1.), ce dernier appliqué aux variables réduites. Le modèle ARIMA (1.1.1.) s'écrit :

$$(1 - \phi_1 B) \Delta Z_t = (1 - \theta_1 B) a_t \quad (14)$$

et le modèle ARMA (1.1.) ou ARIMA (1.0.1.) s'écrit :

$$(1 - \phi_1 B) Y_t = (1 - \theta_1 B) a_t \quad (15)$$

L'estimation préliminaire des paramètres se fait au moyen des formules de Yule-Walker. On peut ensuite obtenir une estimation plus raffinée des paramètres qui maximisent la fonction de vraisemblance logarithmique en minimisant la somme des carrés des résidus.

$$S(\hat{\phi}, \hat{\theta}) = \sum a_t^2(\hat{\phi}, \hat{\theta}) \quad (16)$$

Pour les valeurs calculées de  $\phi$  et  $\theta$  on obtient les courbes d'égalité de  $S$ . Dans le cas des pluies mensuelles les courbes  $S(\phi, \theta)$  n'ont pas de minimum bien défini. On emploie une méthode itérative non linéaire d'estimation des paramètres qui cherche les valeurs des paramètres pour lesquels la somme des carrés des résidus est minimum.

Pour les 15 bassins on a trouvé pour le modèle ARIMA (1.1.1.) que la valeur de  $\theta$  était très proche de l'unité ( $0,991 \leq \theta_1 \leq 0,999$ ). Or lorsque  $\theta_1 = 1$  le modèle 14 s'écrit :

$$(1 - \phi_1 B) (1 - B) Z_t = (1 - B) a_t \quad (17)$$

ou :

$$(1 - \phi_1 B) Z_t = a_t \quad (18)$$

et le modèle ARIMA (1.1.1.) se réduit au modèle AR(1).

Les modèles sont ensuite évalués au moyen du test de l'ajustement de Box et Pierce (1970) dit *test du porte-manteau* (écrit portmanteau dans la littérature américaine !) et le test de Kolmogorov-Smirnov appliqué au périodogramme accumulé. Les résultats pour les pluies mensuelles sont les suivants :

	ARIMA (1.1.1.)		ARMA (1.1.)	
Test	nombre de fois accepté rejeté		nombre de fois accepté rejeté	
Porte-manteau	7	8	15	0
Périodogramme	13	2	14	1

On observe donc la supériorité du modèle ARMA (1.1.) appliqué aux variables réduites. Ce modèle est applicable tant à la génération qu'à la prévision. Il requiert 27 paramètres, les 12 moyennes mensuelles  $\bar{Z}_j$ , les 12 écarts types mensuels  $S_j$ , les paramètres  $\phi_1$  et  $\theta_1$  et la variance de la série résiduelle  $\sigma_a^2$ . Par contre le modèle ARIMA (1.1.1.) n'est utilisable que pour la prévision et ne requiert que 3 paramètres  $\phi_1$ ,  $\theta_1$  et  $\sigma_a^2$ .

Modèles saisonniers

Le filtrage par différence de pas 12 élimine le cycle annuel dans les séries chronologiques mensuelles. On a calé le modèle ARIMA (1.1.1.)<sub>12</sub> qui s'écrit :

$$(1 - \Phi B^{12}) \nabla_{12} Z_t = (1 - \Theta B^{12}) \alpha_t \quad (19)$$

En posant  $\nabla_{12} Z_t = v_t$  le modèle se réécrit :

$$v_t = \Phi v_{t-12} + \alpha_t - \Theta \alpha_{t-12} \quad (20)$$

Ce modèle attribue une dépendance significative pour le même mois pendant plusieurs années successives, et suppose que la dépendance est la même pour les 12 mois de l'année.

On peut introduire une structure de corrélation entre les mois d'une même année en corrélant les résidus  $\alpha_t$  du modèle ARIMA (1.1.1.)<sub>12</sub> par un modèle non saisonnier du type ARIMA (1.0.0.). Le modèle multiplicatif est de la forme :

$$(1 - \phi B) (1 - \Phi B^{12}) \nabla_{12} Z_t = (1 - \Theta B^{12}) a_t \quad (21)$$

ou :

$$v_t = \phi v_{t-1} + \Phi v_{t-12} - \phi \Phi v_{t-13} + a_t - \Theta a_{t-12} \quad (22)$$

Le résultat de l'application des tests est le suivant :

	ARIMA(1.1.1.) <sub>12</sub>		ARIMA(1.0.0.) x (1.1.1.) <sub>12</sub>	
Test	nombre de fois accepté	rejeté	nombre de fois accepté	rejeté
Porte-manteau	12	3	15	0

qui montre bien la supériorité du modèle multiplicatif. Les modèles saisonniers s'appliquent à la prévision mais pas à la génération.

Prévision des pluies mensuelles

La série (1) peut se réécrire pour  $\mu = 0$  :

$$Z_{t+r} = \sum_{j=0}^{r-1} \psi_j a_{t+r-j} + \sum_{j=r}^{\infty} \psi_j a_{t+r-j} \quad (23)$$

Le deuxième terme du membre de droite est la prévision au temps  $t$  pour une avance  $r$  que l'on dénote par

$\hat{Z}_t(r)$ , et le premier terme à droite est l'erreur  $e_t(r)$ . On montre que  $\hat{Z}_t(r)$  est la prévision correspondant à une erreur minimum en moyenne quadratique, que cette prévision n'est pas biaisée et que la variance de l'erreur de prévision est donnée par :

$$\text{var } e_t(r) = \sum_{j=0}^{r-1} \psi_j^2 \sigma_a^2 \quad (24)$$

Il est donc important de connaître les poids  $\psi_j$  pour différents modèles en fonction des paramètres  $\phi$  et  $\theta$ , Box et Jenkins (1976) donnent cette information pour les modèles les plus courants. Kavvas et Delleur (1975) ont calculé les poids  $\psi_j$  pour les modèles saisonniers de la section précédente et Meckerchav et Delleur (1972) les donnent pour le modèle ARIMA (2.0.0.) x (0.1.1.)<sub>12</sub> qu'ils ont employé pour les débits mensuels.

La série de variables réduites  $Y_t$  a une moyenne nulle et une variance unitaire (approximativement). Pour retrouver la racine carrée des pluies mensuelles avec leurs variations saisonnières il faut faire la transformation inverse :

$$\hat{Z}_t(r) = \hat{Y}_t(r) S_j + \bar{Z}_j \quad (25)$$

où  $\hat{Z}_t(r)$  est la prévision, au temps  $t$  pour une avance  $r$ , de la racine carrée de la pluie mensuelle et  $S_j$  et  $\bar{Z}_j$  sont l'écart-type et la moyenne de la racine carrée des pluies au mois  $j$ . L'erreur type de la prévision  $\hat{Z}_t(r)$  est :

$$\hat{\sigma}_{Z_t}(r) = \hat{\sigma}_{Y_t}(r) S_j \quad (26)$$

où  $\hat{\sigma}_{Y_t}(r)$  est l'erreur type de  $\hat{Y}_t(r)$  qui peut être obtenue de (24).

Le bruit blanc ou processus ARMA (0.0.0.), qui avait été suggéré par Roesner et Yevjevich (1966), et le processus ARIMA (1.0.1.) appliqués aux variables réduites ont été utilisés pour la prévision des racines carrées des pluies mensuelles. Le bruit blanc a la fonction de prévision la plus simple :

$$\hat{Y}_t(r) = 0 \quad ; \quad r \geq 1 \quad (27)$$

et :  $\hat{Z}_t(r) = \bar{Z}_j$  avec  $\hat{\sigma}_{Z_t}(r) = S_j \quad (28)$

Les modèles ARIMA (1.1.1.)<sub>12</sub> et (1.0.0.) x (1.1.1.)<sub>12</sub> ont été également appliqués à la prévision des racines carrées des pluies mensuelles. La figure 2 donne un exemple de l'application de ces différents modèles correspondant aux pluies tombant sur le bassin de la East Fork White River en amont de Seymour (Indiana) de 6 040 km<sup>2</sup>. On observe que les prévisions sont essentiellement les mêmes pour les différents modèles, et

ARIMA	$\phi$	$\theta$	$\sigma_a^2$	$\sigma_{Z_t}^2$ (24)	
(1.0.1.)	0.248	0.20	0.963	5.049	
(1.1.1.)	0.171	0.998	0.976	-	
(0.0.0.)	0	0	-	4.988	
	$\phi$	$\Phi$	$\Theta$	$\sigma_a^2$	$\sigma_{Z_t}^2$ (24)
(1.1.1.) <sub>12</sub>	0	0.008	0.950	0.255	4.906
(1.0.0.) x (1.1.1.) <sub>12</sub>	0.129	-0.004	0.952	0.252	4.959

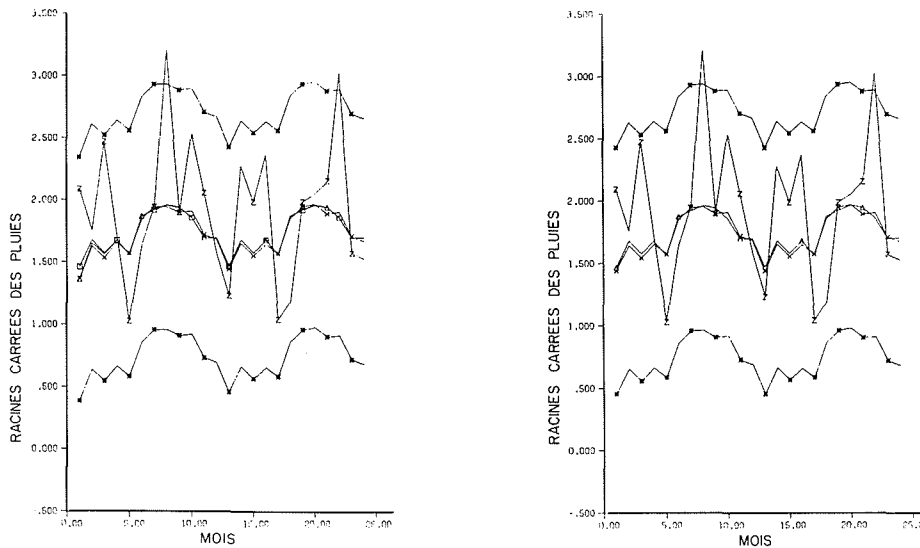


Figure 2 – Prédiction des racines carrées des pluies mensuelles en pouces (inch) par différents modèles

- Racines moyennes mensuelles
  - Δ Prévisions ARIMA (1.0.1.)
  - X Prévisions ARIMA saisonnier (1.0.0.) x (1.1.1.)<sub>12</sub>
  - \* Intervalle de confiance 95 % pour ARIMA (1.0.0.) x (1.1.1.)<sub>12</sub>
  - Z Racines carrées des observations
- Square roots of monthly rainfall (inches) forecast with different models.*
- Mean monthly roots
  - Δ ARIMA forecasts (1.0.1.)
  - X ARIMA seasonal forecast (1.0.0.) x (1.1.1.)<sub>12</sub>
  - \* 95 % confidence interval for ARIMA (1.0.0.) x (1.1.1.)<sub>12</sub>
  - Z Square roots of observed data

après 3 pas de temps les prévisions sont très proches de la moyenne mensuelle. Enfin en Table 1 on trouvera les paramètres des modèles étudiés et l'erreur quadratique moyenne de la prévision pour le même bassin qu'à la figure 2. On remarque que l'erreur de prévision à 24 mois est presque la même pour les différents modèles considérés.

porte-manteau a donné les résultats suivants (au niveau de 90 % de signification) :

Modèle	Nombre de cas	
	acceptés	rejetés
ARMA (1.0.)	10	6
ARMA (2.0.)	16	0

**Application aux débits mensuels**

*Application to monthly discharge*

Les débits mensuels ont été étudiés par Mc Kerchar et Delleur (1974, 1972) pour 16 bassins dont 15 sont ceux qui ont été étudiés pour les pluies mensuelles moyennes. La longueur de la chronique des débits varie entre 38 et 56 ans. Les logarithmes des débits ayant une distribution essentiellement gaussienne, on a formé la variable réduite :

$$Y_t = \frac{Z_t - \bar{Z}_i}{S_i} \tag{29}$$

où  $Z_t$  est le logarithme du débit mensuel du mois  $t$ ,  $\bar{Z}_i$ ,  $i = 1, \dots, 12$ , est la moyenne des logarithmes des débits du  $i$ -ème mois et  $S_i$  est l'écart type correspondant.

Des processus autorégressifs du premier et du deuxième ordre ont été calés aux séries des  $Y_t$ . Le test du

Nous avons donc retenu le modèle ARMA (2.0) qui s'écrit :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t \tag{30}$$

La série des logarithmes des débits mensuels a été filtrée en prenant la première différence au pas de 12. La forte dépendance sérielle décelée dans le cas précédent a suggéré l'application du modèle multiplicatif (2.0.0) x (0.1.1)<sub>12</sub> qui s'écrit :

$$(1 - \phi_1 B - \phi_2 B^2) \nabla_{12} Z_t = (1 - \Theta_1 B^{12}) a_t \tag{31}$$

Le modèle a été accepté dans 15 des 16 cas (au niveau de 90 % de signification). Le modèle ARMA (2.0) requiert 27 paramètres (12 valeurs de  $\bar{Z}_i$ , 12 valeurs de  $S_i$ ,  $\phi_1$ ,  $\phi_2$  et  $\sigma_a^2$ ). Il peut être appliqué à la génération de séries synthétiques ainsi qu'à la prévision. Le modèle multiplicatif ARIMA (2.0.0) x (0.1.1)<sub>12</sub> requiert seulement 4 paramètres ( $\phi_1$ ,  $\phi_2$ ,  $\Theta_1$  et  $\sigma_a^2$ ) mais s'applique seulement à la prévision.

L'équation de prévision pour le modèle ARMA (2.0) s'écrit :

$$\begin{aligned} \hat{Y}_t(1) &= \phi_1 Y_t + \phi_2 Y_{t-1} \\ \hat{Y}_t(2) &= \phi_1 \hat{Y}_t(1) + \phi_2 Y_t \\ \hat{Y}_t(r) &= \phi_1 \hat{Y}_t(r-1) + \phi_2 \hat{Y}_t(r-2), \quad r \geq 3 \end{aligned} \quad (32)$$

La prévision à un mois ne dépend que de valeurs observées, tandis que la prévision à deux mois dépend de la prévision à un mois et une valeur observée et la prévision à 3 mois dépend des prévisions à deux mois et à un mois, et ainsi de suite. Il apparaît donc que les prévisions à court terme, c'est-à-dire à un ou deux pas de temps sont plus sûres que celles à long terme. La prévision du logarithme des débits mensuels est donnée par la transformation inverse :

$$\hat{Z}_t(r) = \hat{Y}_t(r) S_i + \bar{Z}_i \quad (33)$$

et l'erreur type de  $\hat{Z}_t(r)$  est :

$$S_Z(t, r) = S_Y(r) S_i \quad (34)$$

où  $S_Y(r)$  est l'écart type de l'erreur de prévision des  $\hat{Y}_t(r)$  calculés par (24) et  $S_i$  est l'écart type du logarithme des débits du  $i$ -ème mois, où la valeur de  $i$  est déterminée à partir de l'origine  $t$  plus l'avance  $r$ . On voit donc que la prévision des logarithmes des débits et son erreur type prennent en compte la variation saisonnière de l'écart type des logarithmes des débits. Ce n'est pas le cas du modèle multiplicatif qui ne requiert pas de transformation inverse et pour lequel l'erreur type  $S_Z(r)$  est calculée directement à partir de (24).

Pour obtenir la prévision des débits  $\hat{Q}_t(r)$  à partir de la prévision de leurs logarithmes  $\hat{Z}_t(r)$  on emploie la méthode des moments qui donne :

$$\hat{Q}_t(r) = \exp(m + S^2/2) \quad (35)$$

et l'erreur type de la prévision est donnée par :

$$S_t(r) = \hat{Q}_t(r) [\exp(S^2) - 1]^{1/2} \quad (36)$$

où  $m$  et  $S$  représentent  $\hat{Z}_t(r)$  et  $S_Z(t, r)$  obtenus par (33) et (34) pour ARMA (2.0) ; et  $\hat{Z}_t(r)$  et  $S_Z(r)$  obtenus par application du modèle (31) dans le cas du modèle ARIMA (2.0.0) x (0.1.1)<sub>12</sub>. Les prévisions obtenues par ces deux modèles A et B respectivement, sont illustrées sur la figure 3, pour la rivière Wabash à Logansport (Indiana). La différence entre les prévisions par les modèles A et B est due au fait que le modèle B ne tient pas compte de la variation saisonnière de l'écart type.

Une application intéressante est celle de la prévision en temps réel, c'est-à-dire avec une avance d'une unité de temps. Cette application a été traitée par Rao et Kashyap, (1971), dans laquelle les paramètres du modèle sont réestimés à chaque intervalle de temps au fur et à mesure que les nouvelles informations deviennent disponibles. Pour des séries relativement longues comme celles qui ont été utilisées dans cette étude les paramètres sont stables, et il n'est pas nécessaire de les réajuster à chaque pas de temps. La figure 4 donne la prévision en temps réel par les modèles A et B. En général ces prévisions sont plus proches des débits observés que les débits moyens qui peuvent

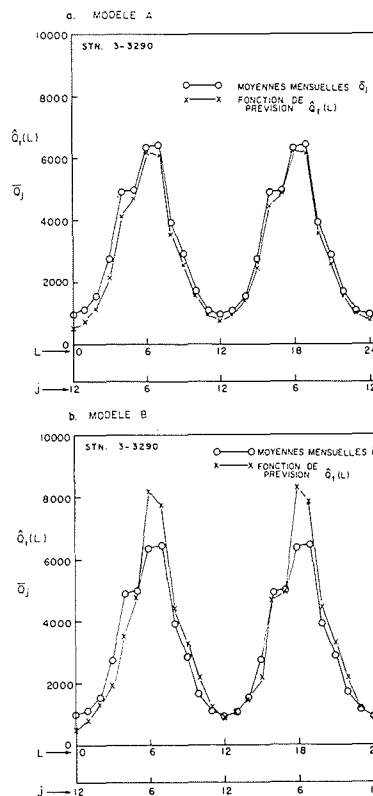


Figure 3 – Fonction de prévision pour les débits  
Forecasting function for discharge

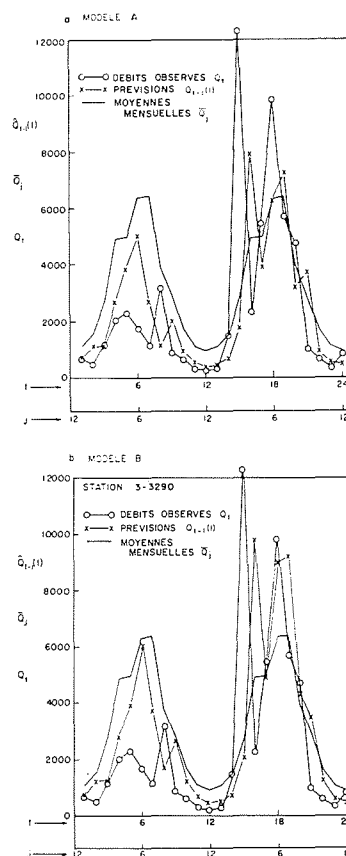


Figure 4 – Prévisions un pas de temps en avant et débits observés  
One time-step ahead forecasts and observed discharge.

être considérés comme prévisions inconditionnelles ou prévision à longue échéance.

La Table II donne les valeurs des paramètres pour la Rivière Wabash à Logansport pour trois modèles.

Modèle	Paramètres			
	$\phi_1$	$\phi_2$	$\theta_1$	$\sigma_a^2$
ARMA (1.0) *	0.510	—	—	0.742
ARMA (2.0)	0.468	0.082	—	0.738
ARIMA (2.0.0) x (0.1.1) <sub>12</sub>	0.489	0.72	0.952	0.546

(\*) Ce modèle passe le test du porte-manteau pour cette station.

### Application aux débits journaliers

#### Application to daily discharge

Les débits journaliers ont été étudiés par Tao et Delleur (1975, 1976) pour 32 bassins dont 16 ont été étudiés pour les débits mensuels et 15 pour les pluies mensuelles. Ces bassins sont situés au centre des Etats Unis (Indiana, Illinois, Ohio et Kentucky), ont une superficie qui varie entre 74 074 km<sup>2</sup> et 606 km<sup>2</sup> et la longueur des chroniques varie entre 29 et 63 ans.

Dénotons par  $Q_{p,t}$  la variable hydrologique au jour  $t$  de la  $p$ -ième année. Soient  $\bar{Q}_t$  et  $S_t$  la moyenne et l'écart type à la position  $t$ , estimés par :

$$\bar{Q}_t = \frac{1}{n} \sum_{p=1}^n Q_{p,t} \quad (37)$$

$$S_t = \frac{1}{n-1} \left[ \sum_{p=1}^n (Q_{p,t} - \bar{Q}_t)^2 \right]^{1/2} \quad (38)$$

On ajuste une série de Fourier à  $\bar{Q}_t$  et  $S_t$  :

$$\bar{Q}_t = \bar{Q}_t + \sum_{j=1}^m \left( A_j \cos \frac{2\pi jt}{w} + B_j \sin \frac{2\pi jt}{w} \right) \quad (39)$$

$$\bar{S}_t = \bar{S}_t + \sum_{j=1}^m \left( A_i \cos \frac{2\pi it}{w} + B_i \sin \frac{2\pi it}{w} \right) \quad (40)$$

où  $\bar{Q}_t$  et  $\bar{S}_t$  sont les valeurs moyennes de  $\bar{Q}_t$  et  $\bar{S}_t$ , et où  $w = 365$  pour des séries journalières. On forme la variable réduite :

$$Y_{p,t} = \frac{Q_{p,t} - \mu_t}{\sigma_t} \quad (41)$$

qui est une standardisation cyclique paramétrique. Le coefficient de corrélation saisonnier de la variable  $Y_{p,t}$  de moyenne  $\bar{Y}$  et variance  $S^2$  au pas  $k$  et au temps  $t$  est défini par :

$$\rho_{k,t} = \frac{\text{COV}(Y_{p,t}, Y_{p,t-k})}{\text{Var}(Y_{p,t})} \quad (42)$$

$$= \frac{E[(Y_{p,t} - \bar{Y})(Y_{p,t-k} - \bar{Y})]}{S^2}$$

Comme la fonction de corrélation varie dans le temps on est amené à généraliser les modèles ARMA pour inclure le cas où les paramètres  $\phi_{k,t}$  et  $\theta_{k,t}$  varient dans le temps :

$$Y_{p,t} = \sum_{k=1}^p \phi_{k,t} Y_{p,t-k} - \sum_{k=1}^q \theta_{k,t} a_{p,t-k} + a_{p,t} \quad (43)$$

Les équations de Yule-Walker pour l'estimation préliminaire des paramètres peuvent se généraliser aussi. Par exemple pour les modèles ARMA (1.0) et (1.1) on a :

$$\phi_{1,t} = r_{1,t} \quad (44)$$

et :  $r_{2,t} = r_{1,t} \cdot \phi_{1,t} \quad (45a)$

$$r_{1,t} = \frac{(1 - \phi_{1,t} \theta_{1,t})(\phi_{1,t} - \theta_{1,t})}{1 + \theta_{1,t}^2 - 2\phi_{1,t} \theta_{1,t}} \quad (45b)$$

où  $r_{k,t}$  est l'estimateur de  $\rho_{k,t}$ . En pratique on estime les paramètres par la méthode des moindres carrés auxquels on cale ensuite une série de Fourier. Six harmoniques sont suffisantes pour caler les séries de  $\mu_t$ ,  $\sigma_t$ ,  $\phi_{k,t}$  et  $\theta_{k,t}$ .

Cette méthodologie a été appliquée aux logarithmes des débits exprimés en variables réduites par (41) où  $Q_{p,t}$  sont les logarithmes des débits journaliers,  $\mu_t$  et  $\sigma_t$  leurs moyenne et écart type cyclique. La figure 5 montre les valeurs de  $\mu_t$  et  $\sigma_t$  pour la rivière Salamonie à Dora (Indiana). La superficie du bassin versant est de 1443 km<sup>2</sup> et la longueur de la chronique est de 45 ans. La figure 6 montre le coefficient de corrélation  $\rho_{k,t}$  pour  $k = 1, 2, 3$  pour la même station, et la figure 7 donne les valeurs des paramètres  $\phi_1$  et  $\theta_1$  du modèle ARMA (1.1) ajusté à cette station, ainsi que les séries de Fourier obtenues avec 6 harmoniques.

Pour comparer la performance des modèles ARMA à coefficients constants et à coefficients variables on a généralisé la statistique  $Q$  du test dit du porte-manteau. Pour le cas de coefficients constants cette statistique est donnée pour un nombre maximum de pas  $K$  par

$$Q = n w \sum_{k=1}^K (r_k)^2 \quad (46)$$

où  $n w$  est la taille de l'échantillon,  $r_k$  est le coefficient de corrélation du résidu au pas  $k$ .  $Q$  a approximativement une distribution de  $\chi_v^2$  avec un nombre de degrés de liberté  $v = K - p - q$ . Pour le cas de coefficients variables dans le temps on a introduit :

$$Q_1 = n \sum_{k=1}^K \sum_{t=1}^w (r_{k,t})^2 \quad (47)$$

où  $n$  est le nombre d'années d'observation,  $w$  le nombre d'observations pendant l'année, et  $r_{k,t}$  est le coefficient de corrélation saisonnier des résidus au temps  $t$  et au pas  $k$ . Les propriétés statistiques de  $Q_1$  ne sont pas connues mais les statistiques  $Q_1$  pour les modèles à coefficients variables sont systématiquement inférieures aux statistiques  $Q$  des modèles à coefficients constants. Ceci indique donc le meilleur ajustement des modèles à coefficients variables.



Figure 5 – Moyenne et écart type des logarithmes des débits journaliers (station 33245)  
Average and standard deviation of the logarithms of daily discharge (station 33245)

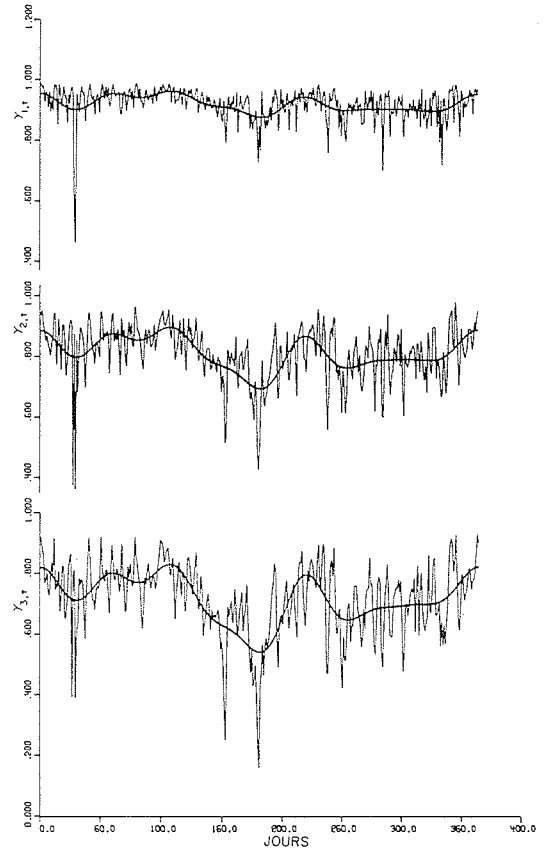


Figure 6 – Coefficient d'autocorrélation saisonnier de pas 1, 2 et 3 des logarithmes des débits journaliers (station 33245)  
Seasonal autocorrelation coefficient for steps 1, 2 and 3 of the logarithms of daily discharge (station 33245)

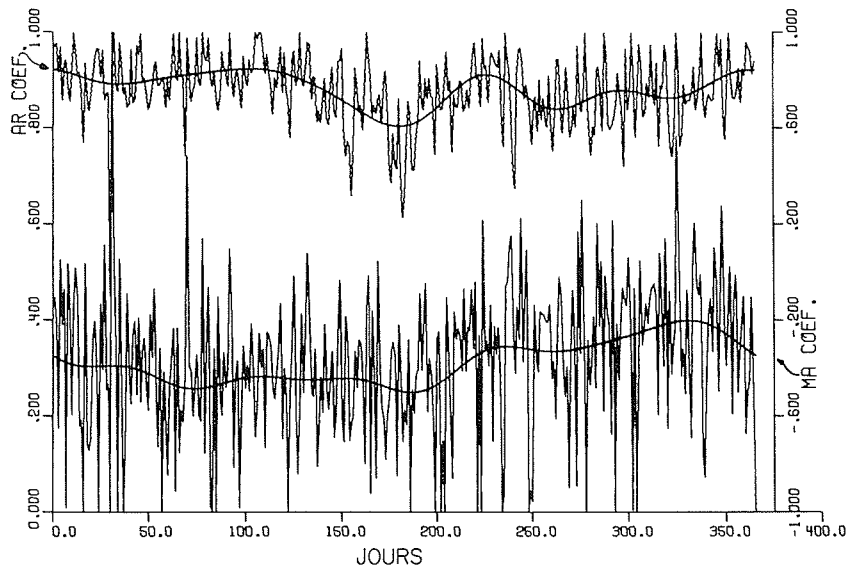


Figure 7 – Coefficients autorégressifs (AR) et de moyenne mobile (MA) pour un modèle ARMA non stationnaire appliqué aux logarithmes des débits journaliers standardisés (station 33245)  
Autoregressive (AR) and moving average (MA) coefficients for a non-stationary ARMA model, applied to the logarithms of standardized daily discharge (station 33245)



A titre d'exemple, le modèle ARMA (1.1) appliqué aux débits hebdomadaires de la rivière Salamonie à Dora donne une statistique  $Q = 133$  pour le cas de coefficients constants et une statistique  $Q_1 = 83$  pour le cas de coefficients variables.

Il est aussi intéressant de noter que les résidus du modèle ARMA (1.1) appliqués aux débits journaliers de la Salamonie ont une distribution essentielle-ment bi-gamma.

**Prévision des débits en tenant compte de la prévision des pluies.**

*Forecasting of discharge with allowance for rainfall forecast*

L'équation (29) peut se réécrire :

$$Q_i = Q_t + S_t \cdot Y_i \quad (48)$$

où  $Q_i$  est le débit,  $Q_t$  et  $S_t$  sont les débits moyens et l'écart type au temps  $t$  qui peuvent être considérés comme une composante déterministe, et  $Y_i$  est la composante aléatoire. Etant donné que l'espérance mathématique de  $Y_i$  est nulle, la prévision inconditionnelle au temps  $i + 1$  est :

$$\hat{Q}_i(1) = Q_{t+1}$$

Si la composante aléatoire est représentée par un processus ARMA (2.0) qui représente bien les débits mensuels comme on l'a vu plus haut, la fonction de prévision s'écrit :

$$\hat{Q}_i(1) = Q_{t+1} + S_{t+1} \sum_{k=1}^2 \phi_k Y_{i-k+1} \quad (49)$$

Par contre si on emploie le concept du filtre de Kalman, la sortie  $Y_i$  du filtre récursif du premier ordre soumis à des entrées aléatoires  $X_i$  est donnée par :

$$Y_i = C_1 Y_{i-1} + C_2 X_i, \quad |C_1| < 1 \quad (50)$$

Si ce système représente un bassin versant on pourrait supposer que  $X_i$  soit la précipitation et  $Y_i$  le débit. En combinant (48) et (50) on obtient :

$$\hat{Q}_i(1) = Q_{t+1} + S_{t+1} (C_1 Y_i + C_2 P_{i+1}) \quad (51)$$

où  $P$  est la précipitation.

Tao et Delleur (1975) ont appliqué les modèles (49) et (51) à la prévision des débits en employant un processus ARMA (14.0) pour la prévision des pluies. Ils ont trouvé que l'erreur quadratique moyenne de la prévision des débits mensuels est du même ordre de grandeur par les modèles (48, 49 et 51).

Tao, Rao et Rukvichai (1975) ont étudié des modèles de prévision de débits journaliers dans un bassin de 2797 km<sup>2</sup>, situé dans le Nord de l'Etat de Kentucky (USA). Pour de tels bassins il s'est avéré que la prédiction des débits peut être améliorée si

1) on tient compte de la prévision des précipitations que l'on peut obtenir des services météorologiques,

2) si on incorpore dans le modèle l'effet de stockage du bassin

3) si on considère la structure de l'erreur résiduelle après l'application et 1 et 2.

L'équation de continuité entre le changement de stockage dans le bassin  $S_t - S_{t-1}$  entre les jours  $t$  et  $t - 1$ , l'apport des pluies  $P_t$ , le débit  $Q_t$  et l'évapotranspiration  $ET_t$  (calculée par une formule empirique à partir des observations de température) s'écrit :

$$S_t = S_{t-1} + P_t - Q_t - ET_t \quad (52)$$

en prenant une valeur arbitraire du stockage initial on ajuste ensuite une série de Fourier aux stockages :

$$\hat{S}_t = \bar{S} + \sum_{j=1}^6 \left( A_j \cos \frac{2\pi jt}{w} + B_j \sin \frac{2\pi jt}{w} \right), \quad t = 1, \dots, 365 \quad (53)$$

et on calcule l'erreur :

$$Se_t = S_t - \hat{S}_t \quad (54)$$

à laquelle on a ajusté un processus ARMA (3.0). On peut alors obtenir la prévision du stockage du jour avec un jour d'avance :

$$\hat{S}_t(1) = \hat{S}_{t+1} + \sum_{k=1}^3 \phi_k Se_{t+1-k} \quad (55)$$

En s'inspirant de la continuité hydrologique on peut écrire une équation de régression entre la prévision des débits  $\hat{q}_t(1)$ , celle des pluies  $\hat{P}_t(1)$  estimée par les services météorologiques, celle de l'évapotranspiration  $\hat{ET}_t(1)$  estimée à partir de la prévision de température des services météorologiques et la prévision du stockage donnée par l'équation précédente. L'équation s'écrit :

$$\begin{aligned} \hat{q}_t(1) = & a_1 + a_2 [\hat{P}_t(1) - \hat{ET}(1)] + a_3 (P_t - ET_t) \\ & + a_4 [\hat{S}_t(1) - S_t] + a_5 (S_t - S_{t-1}) + a_6 Q_t \end{aligned} \quad (56)$$

où  $Q_t$  est le débit observé au jour  $t$ . L'erreur de prévision est :

$$e_{t+1} = Q_{t+1} - \hat{q}_t(1) \quad (57)$$

à laquelle on a ajusté un processus ARMA (14.0.0.). La prévision du débit est finalement donnée par :

$$\hat{Q}_t(1) = \hat{q}_t(1) + \sum_{i=0}^{13} \phi_i [q_{t-i} - \hat{q}_{t-i-1}(1)] \quad (58)$$

On arrive donc à la conclusion que pour la prévision à court terme des débits provenant de petits bassins, il est désirable de prendre en considération la prévision des pluies, de l'évapotranspiration et de l'état du bassin.

**Le modèle ARMA (1.1.).**

*The ARMA model (1.1.).*

Le processus ARMA (1.1.) a mérité une attention spéciale dans la littérature en ce qui concerne la transformation logarithmique des débits et des propriétés de persistance.

La transformation logarithmique a été employée dans la simulation et prévision des débits mensuels et journaliers. Mejia et Rodriguez-Iturbe (1974) ont étudié la relation entre le coefficient de corrélation  $\rho_x(\tau)$  et  $\rho_y(\tau)$  où les variables  $x$  et  $y$  sont reliées par l'expression  $x = \exp y + a$ . Cette relation est :

$$\rho_x(\tau) = \frac{\exp(\sigma_y^2 \rho_y^{|\tau|}(1)) - 1}{\exp(\sigma_y^2) - 1} \quad (59)$$

Si on désire préserver la structure de corrélation, le processus  $y$  devrait être généré avec la fonction de corrélation :

$$\rho_y(\tau) = \log \{1 + \rho_x(\tau) [\exp(\sigma_y^2) - 1]\} / \sigma_y^2 \quad (60)$$

qui se comporte asymptotiquement comme :

$$\rho_y(\tau) = \rho_x^{|\tau|}(1) [\exp(\sigma_y^2) - 1] / \sigma_y^2 \quad (61)$$

qui à son tour est de la forme de la fonction de corrélation du processus ARMA (1.1.)

$$\rho(\tau) = \frac{(1 - \phi_1 \theta_1)(\phi - \theta_1)}{(1 + \theta^2 - 2\phi_1 \theta_1)\phi_1} \phi_1^{|\tau|} \quad (62)$$

En comparant (61) et (62), Mejia et Rodriguez-Iturbe évaluent les paramètres  $\theta_1$  et  $\phi_1$  du processus ARMA (1.1.) appliqué aux logarithmes des débits en conservant la structure de corrélation des débits.

O'Connell (1971-1974) étudie le processus ARMA (1.1.) lorsque  $\phi$  approche l'unité et approche donc la non stationnarité. Alors ce processus peut générer des séquences dont le comportement le rapproche de celui du bruit gaussien fractionnaire proposé par Mandelbrot et Wallis (1969) pour simuler les effets de persistance.

## Remerciements

### Acknowledgments

Cet article est basé sur les résultats de recherches subventionnées en partie par l'"Office of Water Research and Technology, Département of the Interior, USA" sous le projet OWRT-B-036-IND, et en partie par la Purdue University à West Lafayette, Indiana, USA.

## Bibliographie

### References

[1] BOX (G.) et JENKINS (G.), 1976. – *Time Series Analysis, Forecasting and control*, Holden-Day, 553 pp. Revised ed.

- [2] BOX (G.), PIERCE (D.), 1970. – Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, *Jour. Amer. Stat. Assoc.* 64.
- [3] DELLEUR (J.), KAVVAS (M.), 1978. – Stochastic Models for Monthly Rainfall Forecasting and Synthetic Generation, *Jour. of Applied Meteorology*.
- [4] DELLEUR (J.), TAO (P.), KAVVAS (M.), 1976. – An Evaluation of the Practicality and Complexity of some Rainfall and Runoff Time Series Models, *Water Res. Res.* 12, n° 5, pp. 953-969.
- [5] KAVVAS (M.), DELLEUR (J.), 1975 a. – The Stochastic and Chronologic Structure of Rainfall Sequences – Application to Indiana, *Purdue Univ., Water Resources Research Center, Tech. Rept. n° 57*, 199 pp.
- [6] KAVVAS (M.), DELLEUR (J.), 1975 b. – Removal of Periodicities by Differencing and Monthly Mean Subtraction, *Jour. of Hydrology*, 26, 335-353.
- [7] Mc. KERCHAR (A.), DELLEUR (J.), 1974. – Application of seasonal Parametric Linear Stochastic Models to Monthly Flow Data, *Water Resour. Res.* 10, 246-255.
- [8] Mc. KERCHAR (A.), DELLEUR (J.), 1972. – Stochastic Analysis of Monthly Flow Data – Application Lower Ohio River Tributaries, *Purdue Univ. Water Resour. Research Center, Tech. Rept. n° 26*, 104 pp.
- [8 bis] MANDELBROT (B.), WALLIS (J.), 1969. – Computer experiments with fractional gaussian noise, *Water Resour. Research*, 5, 1, 228-267.
- [9] MEJIA (J.), RODRIGUEZ-ITURBE (I.), 1974. – Correlation Links Between Normal and Log Normal Processus, *Water Resour. Research*. 10, n° 4, pp. 689-690.
- [10] O'CONNEL (P.), 1971. – *A Simple Stochastic Modelling of Hurst's Law*. Math. Models in Hydrology. Symp. Varsovie (AINS Pub. n° 100) Vol. 1, pp. 169-187.
- [11] O'CONNEL (P.), 1974. – *Stochastic Modelling of Long-Term Persistence in Streamflow Sequences*, Thèse, Ph.D. Civil Eng. Imperial College, Londres.
- [12] RAO (A.), KASHYAP, 1971. – *Stochastic Difference Equation Modeling of Hydrologic Processes*. Math. Models in Hydrology, Symp. Varsovie (AIHS Pub. n° 100).
- [13] ROESNER (L.), YEVJEVICH (Y.), 1966. – Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff, *Colorado State Univ. Hydrology Paper 15*
- [14] TAO (P.), DELLEUR (J.), 1976. – Seasonal and non seasonal ARMA Models in Hydrology, *Jour. Hydr. Div. ASCE*. 102, HY10, 1541-1559.
- [15] TAO (P.), DELLEUR (J.), 1975. – Models of the stochastic and chronologic structure, prediction and simulation of runoff sequences. Application to Lower Ohio Basin *Purdue Univ., Water Resources Research Center. Tech. Rept. n° 60*, 124 pp.
- [16] TAO (P.), RAO (R.), RUKVICHAI (C.). – Stochastic Forecasting Models of Reservoir Inflows for Daily Operation, I. Model Development and Application to Rough River Reservoir Inflow. *Purdue Univ. School of Civil eng. Rept. N. CE.HYD.* 75-8.