SIMULATION DES DEBITS MENSUELS ET DECADAIRES PAR LES MODELES ARMA

N. DECHEMI
Maître de Conférence - ENP d'Alger
D. SOUAG
Maître Assistante - USTHB

Résumé

L'objectif du présent article est de trouver le modèle adéquat pour la simulation des débits mensuels et décadaires de certains oueds algériens. Nous avons choisi pour cela de caler des modèles stationnaires de type ARMA à coefficients constants p et q. Les techniques d'identification et d'estimation des paramètres ont été développées, et les modèles validés par plusieurs tests appliqués pour la génération de séries synthétiques de débits mensuels et de débits décadaires préalablement normalisés par différentes méthodes.

Mots clés : simulation • ARMA • débit • pas de temps • identification • estimation•

1 INTRODUCTION

Pour la résolution de différents problèmes en hydrologie en général, et pour l'optimisation du dimensionnement et de la gestion des ressources hydriques en particulier, les projeteurs ont besoin de séries de débit assez longues qui ne sont malheureusement pas toujours disponibles ce qui a imposé l'introduction de la notion de simulation dans le domaine hydrologique au début des années soixante.

Depuis, de grands efforts ont été consacrés en vue d'améliorer, de développer les premiers modèles, et de leur donner des justifications physiques.

Plusieurs modèles ont été proposés parmi lesquels - Les modèles autoregressifs AR (Thomas et Fiering, 1962), modèles du bruit Gaussien fractionnel FGN (Mandelbort et Wallis, 1968, Matalas et Wallis, 1971), les modèles de désagrégation (Valencia et Shaake, 1973), et les modèles ARMA-Markov (Lettenmaier et Burges, 1977).

2 MODELES AUTOREGRESSIFS ET DE MOYENNE MOBILE ARMA

2.1 Formulation mathématique du modèle

Les modèles mathématiques en hydrologie se subdivisent en modèles déterministes et stochastiques. Dans ce dernier cas, on considère une classe de processus stochastique (en temps discret) susceptibles de modéliser les séries chronologiques des débits qui se caractérisent par une certaine dépendance entre observations successives pouvant être représentées par des modèles linéaires.

Le modèle linéaire autoregressif (AR) présente un intérêt particulier dans la modélisation des séries chronologiques en hydrologie. C'est l'un des modèles qui s'adaptent particulièrement bien aux étiages des rivières.

Chaque valeur du processus est une somme pondérée et définie de p valeurs de celui-ci, à laquelle on ajoute une variable aléatoire.

Le modèle autoregressif d'ordre p est représenté par:

$$Z_t = \sum_{j=1}^{p} f_j Z_{t-j} + \varepsilon_t \qquad (1)$$

ϕ_j : Paramètres du modèle (AR,).

Un autre processus linéaire dit de moyenne mobile (MA) représente la valeur au temps t par la somme pondérée et finie de q valeurs antérieures d'une variable aléatoire, il est donné par :

$$Z_t = \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad (2)$$

θ_j : Paramètres du modèle (MA)

La combinaison des deux modèles donne le modèle mixte autoregressif et de moyenne mobile d'ordre (p,q) défini par :

$$Z_t = \sum_{j=1}^{p} \phi_j Z_{t-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \qquad (3)$$

Les paramètres du modèle ARMA sont : μ, θ_1, θ_2, θ_q, $\phi_1, \phi_2, \ldots \phi_p$, soit au total p+q+2 paramètres à estimer à partir des données.

2.2 Justification physique de l'utilisation des modèles ARMA

L'examen du système hydrologique permet de constater qu'un débit Z_t est dû à la contribution des eaux de stockage c.St-1 et des eaux de ruissellement $d.\varepsilon_t$.

$$Z_t = cS_{t-1} + d\varepsilon_t \qquad (4)$$

Avec :

$C \geq 0$

$d \leq 0$

En considérant l'équation de continuité du stockage, on a :

$$S_t = S_{t-1} + cS_{t-1} + a\varepsilon_t \qquad (5)$$

En combinant les équations (4) et (5), Salas et Smith (1981) ont montré que la modélisation des débits peut se faire de la manière suivante

$$Z_t = (1-c)Z_{t-1} + d\varepsilon_t - [(1-c)d - ac]\varepsilon_{t-1} \qquad (6)$$

Par analogie, on montre que ce modèle correspond au cas d'un modèle ARMA (1, 1)

$$Z_t = \phi Z_{t-1} + \theta\varepsilon_{t-1} - \varepsilon_t \qquad (7)$$

2.3 Procédure de la modélisation

La procédure de modélisation suivie dans ce cas est celle établie par Box et Jenkins (1976) et utilisée pour les paramètres hydrologiques par Salas et Smith (1981) :

- Identification de la composition du modèle ;
- Identification du type de modèle ;
- Identification de la forme du modèle ;
- Estimation des paramètres du modèle;
- Tests d'adéquation du modèle ;
- Tests de validation.

2.3.1 Identification de la forme du modèle

L'identification de la forme du modèle, c'est à dire la détermination des ordres p et q est faite en utilisant les caractéristiques des fonctions d'autocorrélation et d'autocorrélation partielle du modèle ARMA.

Chaque modèle APMA(p,q) a des fonctions d'autocorrélations et d'autocorrélations partielles de formes et de caractéristiques différentes. En général, si les autocorrélations tendent exponentiellement vers 0, on se trouve devant un modèle ARMA(p,0) dont l'ordre p dépend du nombre d'autocorrélations partielles significativement différentes de 0. Si les autocorrélations partielles tendent exponentiellement vers 0, c'est un modèle ARMA (0,q), son ordre dépend du nombre d'autocorrélations significatives statistiquement. Lorsque les deux fonctions tendent exponentiellement vers 0, c'est un modèle mixte ARMA(p,q). p>0 et q>0 le problème d'identification est plus complexe, on teste alors les différentes combinaisons possibles avec $p \leq 2$ et $q \geq 2$.

Dans le cas d'un modèle autoregressif d'ordre m (avec m élevé) et des coefficients intermédiaires nuls, on fait appel aux modèles autorégressifs simplifiés SAR(1,1,m) respectant la règle de parcimonie. Les coefficients non nuls de ce modèle se trouvent aux rangs 1,*l* et m.

L'équation de ce modèle est donnée par:

$$Z_t = \Phi_1 Z_{t-1} + \Phi_l Z_{t-l} + \Phi_m Z_{t-m} + \varepsilon_t \qquad (8)$$

2.3.2 Estimation des paramètres du modèle

L'estimation des paramètres du modèle identifié est faite généralement en trois phases de précision croissante.

- Estimation préliminaire qui se fait par la résolution des équations de Yules-Walker pour les paramètres autoregressifs ϕ_i.

$$C_k = \phi_i C_{k-1} \qquad k \geq p+1,\ i = 1, \ldots, p \qquad (9)$$

C_k sont les autocovariances des séries Z_t. Les paramètres de moyenne mobile θ_i sont estimés par un processus itératif utilisant les autocovariances C_j' de la série résultant de la différence entre la série originale et la série formée par le modèle autoregressif

- Une estimation plus raffinée des paramètres peut être obtenue par la méthode du maximum de vraisemblance, qui consiste à maximiser la fonction de vraisemblance logarithmique en minimisant la somme des carrés des résidus.

$$S(\phi,\theta) = \sum \varepsilon_t^2(\phi,\theta) \qquad (10)$$

- Enfin, si des estimateurs plus précis sont requis, on peut utiliser une méthode d'estimation non linéaire qui cherche les paramètres pour lesquels la somme des carrés des résidus est minimale.

2.3.3 Tests du modèle

Après identification et estimation des paramètres du modèle choisi, on effectue des tests pour vérifier la capacité de ce dernier à reproduire les caractéristiques des séries historiques.

Pour ce faire on utilise les tests suivants :

- Tests d'adéquation qui comportent:
  - Test d'indépendance des résidus ;
  - Test de normalité des résidus ;
  - Test d'indépendance des résidus vis à vis des variables à modéliser ;
  - Test de parcimonie.

- Test de validation comportant les tests de ressemblance des caractéristiques statistiques.

3 APPLICATION

3.1 Données utilisées

Ce travail a été réalisé sur six séries de débits mensuels et deux séries de débits décadaires représentant différentes régions de l'Algérie,

Les modèles ARMA utilisés pour la génération de séries synthétiques de débits supposent que la variable étudiée suit une distribution de probabilité normale, un test est donc appliqué pour vérifier ce critère.

Pour les séries s'avérant non symétriques, on a proposé différentes méthodes de normalisation à savoir :

- LT : Transformation du log-translaté;
- LTM : Transformation du log-translaté périodique;
- BD : Transformation de Beard.

Les deux premières méthodes peuvent être facilement appliquées pour les séries utilisées, ce qui n'est pas le cas pour la troisième qui peut poser des problèmes d'indétermination pour les séries présentant des valeurs nulles, car dans ce cas, la variable doit passer par une transformation logarithmique.

Pour pouvoir caler un modèle de type ARMA, la série à modéliser doit aussi vérifier le critère de stationnarité. Le dégagement de la périodicité est fait en utilisant la transformation suivante (passage à la variable centrée réduite)

$$Z_t = (y_{v,\tau} - \mu_\tau) / \sigma_\tau \qquad (11)$$

avec $t = (v-1)\omega + \tau$

μ_τ : Moyenne saisonnière de la variable normalisée $y_{v,\tau}$

σ_τ : Ecart-type de $y_{v,\tau}$

$Z_{v,\tau}$: Variable transformée (v désignant l'année et τ la période),

ω : Nombre de périodes (pour le cas mensuel ω=12, pour le cas décadaire ω=36).

3.2 Simulation des débits mensuels

A partir des fonctions d'autocorrélation partielle, on détermine les ordres des valeurs des coefficients qui sont significatifs à un intervalle de confiance de 95% (tableau 1)et on propose d'utiliser selon les cas:

- Les modèles autoregressifs ARMA (p,O),
- Les modèles autoregressifs et de moyenne mobile ARMA (p,q),
- Les modèles autoregressifs simplifiés SAR(1,1,m).

SERIE	Ordre des coefficients d'autocorrélations significatifs à 95%	Modèles proposés	Transformations utilisées
OUED-BOUSSELAM	1,25	ARMA (1,0), ARMA (1,1) ARMA (1,2), ARMA (2,1)	LT LTM BD
OUED DOUS	1,6,14 1,6,24 1,6,14	ARMA (1,0), ARMA (6,0) ARMA (1,1), ARMA (1,2) ARMA (2,1)	LT LTM BD
OUED FODDA	1,11,14,25 2,17,14,25 1,11,14,25	ARMA (1,0), ARMA (2,0) ARMA (1,1), ARMA (1,2) ARMA (2,1)	LT LTM BD
BENI-BAHDEL	1 1	ARMA (1,0), ARMA (2,1) ARMA (1,2), ARMA (2,1)	LT LTM
CHEFFIA	1,5,6,24	ARMA (1,0), ARMA (5,1) ARMA (6,0), ARMA (1,1) ARMA (1,2), ARMA (2,1)	LT BD
CHEURFA	1,3,5,12,24 1,3,5,12,24	ARMA (1,0), ARMA (3,0) ARMA (5,0), ARMA (1,1) ARMA (1,2), ARMA (2,1) SAR(1,3,12), SAR (1,5,15)	LT LTM

Tableau 1 : Résultats de l'identification des ordres p et q

L'estimation des paramètres des modèles identifiés, a été faite par la méthode des moments et la méthode de maximum de vraisemblance, les résultats obtenus pour la série des débits d'oued Dous sont représentés par le tableau 2 :

	Méthodes des moments			Méthode de M.L.E		
Transfomation	ϕ_1	θ_1	σ_ε^2	ϕ_1	θ_1	σ_ε^2
LT	0.115	0.0119	0.7050	0.514	0.010	0.7193
LTM	0.490	-0.0061	0.7150	0.485	-0.035	0.7185
BD	0.500	-0.0021	0.7180	0.501	-0.035	0.7155

Tableau 2 : Résultats comparatifs des méthodes d'estimation des paramètres.

On a remarqué que dans le cas de séries de grandes tailles (N = 360 mois), les deux méthodes d'estimation des paramètres conduisent à des résultats assez proches.

En ce qui concerne les tests d'adéquation utilisés, le test d'indépendance des résidus a été rejeté par certains modèles, on a remarqué que pour certaines séries (Oued Dous, Oued Fodda et Beni-Bahdel) la transformation utilisée pour normaliser ces séries a un effet sur la satisfaction de ce test, tandis que le test d'indépendance des ε_t vis-à-vis des Z_t a été vérifié par tous les modèles.

Le test d'Akalke a pour but de choisir les modèles sur la base du critère de parcimonie (modèle avec le moins de paramètres à estimer) en comparant les valeurs des AIC calculés pour les différents modèles (Tableau 3).

$$AIC(p,q) = N.Ln(\sigma_\varepsilon^2) + 2.(p.q) \qquad (12)$$

N : Taille de l'échantillon.

σ_ε^2 : Variance des résidus estimée par la méthode du maximum de vraisemblance, .

$$\overline{VER} = \frac{1}{NC} \cdot \sum_{n=1}^{NC} VER_n \qquad (13)$$

NC : Nombre de caractéristiques.

Les modèles les plus satisfaisants du point de vue test d'Akaike sont les SAR et ARMA (p,q) (p et q≠0).

Modèle / Série	1er	2ème	3ème	4ème	5ème	6ème
OUED BOU -SSELAM	SAR (1,3,12) LT (-149.1)	ARMA (1,2) BD (-144.8)	ARMA (1,2) BD (-142.9)	ARMA (1,2) BD (-124.3)	ARMA (1,2) LT (-141.7)	ARMA (1,2) LT (-141.4)
OUED DOUS	ARMA BD (-117.5)	ARMA (1,2) BD (-117.3)	ARMA (1,2) BD (-115.7)	ARMA (1,2) BD (-115.1)	ARMA (1,2) BD (-115.0)	ARMA (1,2) LT (-114.4)
OUED FODDA	SAR (1,2,11) LT (-113.9)	SAR (1,2,11) LTM(-109.3)	SAR (1,2,11) LTM(-107.8)	ARMA (1,2) LT (-103.7)	ARMA (1,2) LT (-100.1)	ARMA (1,2) LT (-98.6)
BENI BAHDEL	ARMA (1,2) LTM	ARMA (1,2) LTM	ARMA (1,2) LTM	SAR (1,3,12) LT	ARMA (1,2) LT	ARMA (1,2) LT
CHEFFIA	ARMA (1,2) BD (-77.3)	SAR (1,5,12) BD (-76.3)	ARMA (1,2) BD (-76.1)	ARMA (1,2) BD (-75.3)	ARMA (1,2) BD (-73.7)	ARMA (1,2) LT (-70.8)
CHEURFA	SAR (1,3,12) LTM(-118.6)	SAR (1,3,12) LTM(-112.6)	ARMA (1,2) LTM(-105.9)	ARMA (1,2) LTM(-102.3)	ARMA (1,2) LT(-100.8)	ARMA (1,2) LT(-99.6)

Tableau 3 : Classement des six premiers modèles suivant le critère de parcimonie

Après le choix des modèles, on génère (à partir de ces données) des séries grâce à des variables indépendantes simulées à partir de variables aléatoires uniformes.

Pour la transformation de la variable uniforme (0,1), on a utilisé les équations de Box et Muller:

$$\varepsilon_1 = (Ln(1/u_1)^{1/2} - Cos(2\pi.u_2) \qquad (14)$$

$$\varepsilon_2 = (Ln(1/u_1)^{1/2} - Sin(2\pi.u_2) \qquad (15)$$

où: ε_1 et ε_2 sont des variables aléatoires suivant la loi normale,

u_1 et u_2 variables aléatoires suivant la loi uniforme (0,1),

Pour la comparaison des différentes caractéristiques statistiques des séries historiques et simulées, le classement des modèles diffère d'une caractéristique à une autre.

Pour toutes les transformations utilisées, et pour tous les modèles, on constate que la caractéristique qui est la mieux reproduite est la moyenne, la transformation qui conduit aux erreurs minimales est la méthode de Beard (quant son utilisation est possible), suivie de la méthode du Log translaté mensuel, puis par la méthode du Log-translaté.

En complément des tests utilisés, on a vérifié la capacité des modèles ARMA à reproduire les caractéristiques statistiques des périodes obtenues, il s'est avéré que les modèles ARMA pris en considération reproduisent mieux les caractéristiques des périodes humides que celles des périodes sèches. Cependant, ces tests n'affectent pas le classement des modèles faits sur la base des tests précédents, les meilleurs résultats ont été obtenus grâce aux modèles ARMA (2, 1), ARMA (1, 1) et ARMA (1,2).

3.3 Simulation des débits décadaires

Deux séries de débits ont été prises en considération (Cheffia et Oued Fodda), ces dernières s'étant avérées non symétriques (après application du test de normalité), ont été transformées par le Log-translaté et Beard, et stationnarisées par la méthode non paramétrique.

L'examen des fonctions d'autocorrélations partielles a conduit à l'étude des modèles présentés dans le tableau 4 :

Série	Modèles	Transformations
CHEFFIA	ARMA (2,0), ARMA (1,1), ARMA (2,1), ARMA (1,2), SAR(1,2,8)	Log translaté
OUED FODDA	ARMA (1,0), ARMA (3,0), ARMA (4,0), ARMA (2,2), ARMA (1,2), ARMA (2,1), SAR (1,4,13) SAR (1,4,11), SAR (1,3,11)	Log translaté et Beard

Tableau 4 : Résultats de l'identifications des ordre p et q pour les séries décadaires

Les paramètres identifiés ont été obtenus par la méthode du maximum de vraisemblance, et les tests d'indépendance des résidus ont été vérifiés pour tous les modèles étudiés, les modèles les plus satisfaisants de point de vue test d'AKAIKE sont les SAR, ARMA(1,2) et ARMA (1, 1).

Trasformation	Modèle	φ_1	φ_2	θ_1	θ_2	σ_ε^2
Log Translaté	ARMA (1,0)	0.351	0.127	-	-	0.7386
	ARMA (1,1)	0.645	-	0.053	-	0.7321
	ARMA (1,2)	0.600	-	0.247	-0,022	0.7313
	ARMA (2,1)	0.383	0.022	0.378	-	0.7314

Tableau 5 : Paramètres des modèles ARMA(p,q) (CHEFFIA)

La comparaison des erreurs relatives des caractéristiques statistiques prises séparément a permis de constater que le classement des modèles diffère d'une caractéristique à une autre.

Série	Modèle	Moyenne	Variance	Coefficient de Variation
CHEFFIA	ARMA (1,1)	0.1409	0.3957	0.2807
	ARMA (1,2)	0.530	0.4307	0.4052
	ARMA (2,0)	0.1273	0.4225	0.5172
OUED FODDA	SAR (1,4,11) LT	-0.0334	-0.0386	-0,4258
	SAR (1,4,13) BD	0.0551	0.0015	-0.2846
	ARMA (3,0) BD	0.0510	0.1892	0.1034
	ARMA (1,2) BD	0.0059	-0.1196	-0.2374

Tableau 6 : Erreurs relatives sur les caractéristiques statistiques

Les séries générées des débits d'oued Fodda sont meilleures que celles de la Cheffia (pour le pas de temps décadaire), du fait de la non stationnarité (même après transformation) de la fonction d'autocorrélation de cette dernière (Tableau 6).

4 CONCLUSION

L'utilisation des modèles stochastiques du type ARMA pour la génération de séries synthétiques de débits mensuels et décadaires d'oueds algériens a conduit aux conclusions suivantes

- Concernant les transformations utilisées pour normaliser les séries dissymétriques, on recommande les transformations du Log-translaté et du Log-translaté périodique pour éviter tous les problèmes d'indétermination rencontrés par les autres méthodes. Ces transformations ont une influence importante sur les tests d'adéquation et de ressemblance des caractéristiques statistiques des séries simulées.
- Les modèles ARMA à coefficients constants reproduisent plus facilement les caractéristiques des périodes humides que celles des périodes sèches, ces dernières sont, pour la plupart des séries sous-estimées, et les écarts correspondants dépassent les 20%.
- Pour les séries dont la fonction d'autocorrélation est non stationnaire, les modèles ARMA à coefficients constants ne sont pas recommandés.
- Les modèles ARMA recommandés pour les séries étudiées sont ARMA(1,2) et ARMA(2,1) pour les séries de débits mensuels avec les transformations du Log-translaté et du Log-translaté mensuel. Pour les séries de débits décadaires, on recommande l'utilisation des modèles ARMA (1,2) et les modèles autoregressifs simplifiés SAR accompagnés des transformations du Log-translaté et de Beard.

BIBLIOGRAPHIE

[1] G. E. P. BOX, and G. M. JENYINS, "*Ti series analysis forecasting and conti Holden-Day*", San Francisco. 1976.

[2] D. P. Lettenmaier, and S. J.Burges, "*Operational assessment of hydrölogie mod of long-term persistance*". Jour. Wa ressources research , 13(l), pp -. 113-1: 1977.

[3] B. B. MANDELBORT, and J. R. WALLI "*Computer experiments with fraction Gaussian noise, 2, Rescaled ranges a spectre*", Jour. Water ressources research, 5(pp : 242-259. 1969.

[4] N. C. MATALAS, and J. R. WALLIS, "*Statistical properties of multivaria Fractionna noise processes*". Jour. Wa ressources research, 7(6), pp-1460-146 1971.

[5] J. D. SALAS, , R. A. SMITH, , "*Physical bas of annualflows*". Jour. Water ressourc research, 17(2), pp - 428-430, 1981.

[6] H. A. THOMAS, and M. B.FIERIN "*Mathematical synthesis of streamflo séquences for the analysis of river basins l simulation.In design of water resssource.* research, 23(l), pp : 32-36. 1962.

[7] D. VALANCIA, and J. C. SHAAKE, "*Disagregation process in stochast hydrology*". Jour. Water ressources researc 9(3), pp : 580-585. 1973.