

## ETUDE DU CARACTERE STOCHASTIQUE DES PRECIPITATIONS A PARTIR DES DONNEES RADAR METEOROLOGIQUE

N. LEHTIHET<sup>(1)</sup>, F. YUCEF ETTOUMI<sup>(1)</sup>, A. ADANE<sup>(1)</sup>, H. SAUVAGEOT<sup>(2)</sup>

<sup>^</sup> Université des Sciences et de la Technologie Houari Boumédiène, Faculté d'Électronique et d'Informatique,

Département Télécommunications, Laboratoire Traitement d'Images et Rayonnement,

BP N° 32 El Alia, Bab Ezzouar, Alger, Algérie. Télé/Fax: +213 21 24 71 87

e-mail : [lehtihet.nouria@caramail.com](mailto:lehtihet.nouria@caramail.com), [yucefettoumi@yahoo.fr](mailto:yucefettoumi@yahoo.fr)

<sup>(2)</sup> Université Paul Sabatier Toulouse, Observatoire Midi Pyrénées, Laboratoire d'Aérodynamique,

Centre de Recherches Atmosphériques, 8, route de Lannemezan, 65300 Campistrous, France.

### Résumé

Des images collectées toutes les cinq minutes par le radar de Bordeaux (France) en 1996, ont fait l'objet d'une analyse statistique basée sur des processus autorégressifs. Cette analyse appliquée aux échos de précipitations, a montré que le processus considéré présente un fort coefficient d'autocorrélation et que les variations quotidiennes des surfaces occupées par ces échos sont convenablement décrites par un processus autorégressif du premier ordre. Pour généraliser l'approche autorégressive, de telles variations ont aussi été modélisées par un processus ARIMA. On trouve que les paramètres du modèle ARIMA changent d'un mois à l'autre, reflétant ainsi le comportement de ce processus à long terme. A l'échelle horaire, les meilleurs résultats sont obtenus à l'aide d'un modèle A R I M A ( 1,1,0).

### Abstract

Images collected every five minutes by the radar of Bordeaux (France) in 1996, have been processed using an autoregressive approach. This kind of statistical analysis applied to precipitation echoes, has shown that the process under consideration, is characterised by a strong autocorrelation coefficient and the daily surface variations of precipitation echoes can satisfactorily be described by a first-order autoregressive process. To generalise the autoregressive approach, such variations have also been modelled using an ARIMA process. The A R I M A parameters are found to change from a month to another on a daily basis. Then, they express the long-term behaviour of such a process. At hourly scale, the best results are obtained using an ARIMA (1, 1, 0) model.

**Mots-clés** : radar ; météorologie; précipitations ; statistique ; AR ; ARIMA.

**Keywords** : radar; meteorology; precipitation; statistics; AR; ARIMA.

### Introduction

Les inondations, les phénomènes causés par les déboisements excessifs de certaines régions ou tout simplement par les feux de forêts, la désertification de certaines zones, la réduction de la diversité des espèces et l'explosion démographique nous conduisent à réfléchir quant à la gestion rationnelle et l'utilisation efficace des ressources en eau dans la planète. En effet, les possibilités de développement dans certaines régions sont freinées par le manque d'information relative à une quantification des potentialités hydriques. Notamment, la gestion raisonnable des aménagements hydrauliques permettrait d'alléger les nombreuses contraintes telles que celles citées précédemment. En fait, les ressources en eau reçues au sol sont difficilement évaluées au moyen de pluviomètres car l'indication enregistrée par ces appareils est spécifique de l'endroit où la mesure est effectuée. Or cette évaluation est rendue complexe par le fait qu'on a affaire à une variation spatio-temporelle selon le déplacement de la perturbation, le lieu de l'averse, la topographie et les obstacles géographiques locaux. Dans ce cas, l'utilisation de radars météorologiques serait judicieuse pour déterminer les quantités de précipitations reçues au sol [Devuyst, 1972]. En effet, les données radar permettent de localiser et de suivre des

champs de précipitations. De telles données sont maintenant largement utilisées pour les besoins hydrologiques de la prévision. Dans ce contexte, les données prises par le radar de Bordeaux en 1996 sont analysées ci-après et modélisées par des processus autorégressifs.

## 1. Observation des précipitations à l'aide d'un radar météorologique

En émettant des ondes électromagnétiques, le radar localise des zones de précipitations à des distances pouvant atteindre 200 à 300 km. La mesure de l'intensité des précipitations se fait généralement dans une zone annulaire de rayon compris entre 80 et 120 km. Cette mesure est réalisée par détection des ondes rétrodiffusées par les champs de précipitations. Cependant, le signal utile est le plus souvent perturbé par des échos parasites dûs aux obstacles de la surface terrestre comme les montagnes ou les constructions. De tels échos sont éliminés par masquage.

## 2. Relation entre le facteur de réflectivité et l'intensité des précipitations

Rappelons que les précipitations sont généralement exprimées en hauteur ou lame d'eau précipitée par unité de surface horizontale (en mm) et que l'intensité de pluie (R) est la hauteur d'eau rapportée à l'unité de temps. Le radar donne des indications sur les champs de précipitations par la mesure d'un facteur de réflectivité (Z) qui est par définition [Sauvageot, 1982] :

$$Z = \frac{1}{V} \sum_i D_i^6 \quad (1)$$

V est le volume de résolution contenant diverses gouttes de pluie de diamètre  $D_i$ . En unités logarithmiques, le facteur de réflectivité Z est exprimé en dBZ par :  $Z(\text{dBZ}) = 10 \log [Z(\text{mm}^6/\text{m}^3)]$ . Pour des précipitations stratiformes, Marshall-Palmer [Sauvageot, 1982] ont établi une relation Z-R :

$$Z = 200.R^{1.6} \quad (2)$$

Dans cette expression, R est en mm/h et Z en  $\text{mm}^6/\text{m}^3$ . Pour les images radar de Bordeaux, les variations du facteur de réflectivité sont exprimées en dixièmes de dBZ et la palette de couleurs de chaque image est formée de 28 niveaux de réflectivité.

## 3. Données météorologiques

Nous disposons des données météorologiques acquises par un radar puisé à 2,8 Ghz, installé sur une tour de 18 m de hauteur, à l'aéroport de Bordeaux-Mérignac (dans le sud-ouest de la France). A partir de ce site, des images radar de format 512x512 pixel et d'une résolution de 1 km par pixel ont été prises régulièrement toutes les 5 min, jour par jour, pour l'ensemble des mois de l'année 1996.

## 4. Traitement des images

Les images ont été classées en deux catégories selon le type d'échos, à savoir les images contenant des échos de précipitations et celles ne comportant que des échos du sol [Haddad, 2000]. Ces derniers sont quasiment permanents et peuvent être filtrés ou masqués [Adane et al., 2000]. Lorsque les surfaces des échos de précipitations sont calculées et cumulées jour par jour, nous obtenons une série chronologique, notée série<sub>1</sub>, reflétant la variation de ces surfaces au cours du temps. La figure 1 représente les variations journalières des surfaces d'échos de précipitations observées en 1996. Cette série traduit l'état du climat de la côte atlantique durant cette année. Elle indique que de telles surfaces persistent sur plusieurs jours, sans aucun mouvement saisonnier particulier et que le pic de pluie le plus élevé, s'est produit le 21 Juin 1996. Cette journée a été prise comme référence pour l'étude des variations plus courtes que la journée. Nous obtenons une autre série, notée série<sub>2</sub>, formée des surfaces horaires de précipitations apparaissant au cours de cette journée type. Dans cette étude, nous avons utilisé des variables aléatoires qui sont centrées et réduites.

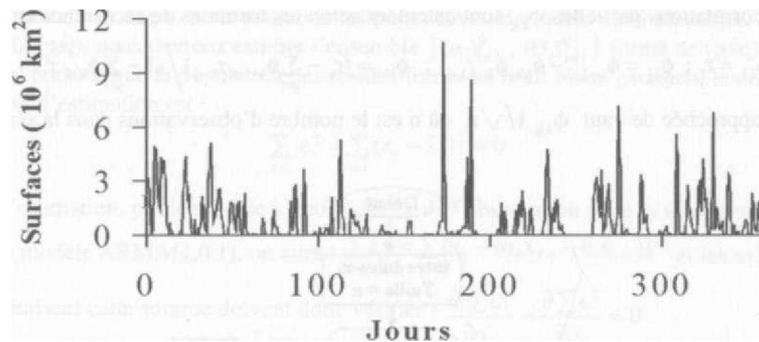


Figure 1 : Variation annuelle des surfaces journalières de précipitations en l'an 1996.

### 5. Construction d'un processus ARIMA ( p,d,q )

Le modèle ARIMA, Auto-regressive Integrated Moving Average, qui est une généralisation des modèles AR, comporte trois types de paramètres. Ce sont les paramètres autorégressifs (p), les paramètres de moyennes mobiles (q), et les différenciations explicites (d) [Box et Jenkins, 1976]. Le modèle ARIMA(p,d,q) est exprimé sous la forme abrégée par [Bosq et Lecoutre, 1992] :

$$(1 - \sum_{i=1}^p \alpha_i \cdot L^i) \cdot \Delta^d \cdot x_t = (1 - \sum_{i=1}^q \theta_i \cdot L^i) \cdot \varepsilon_t \tag{3}$$

Dans cette équation, L est l'opérateur de retard ou 'lag', tel que :  $L^1 \cdot x_t = x_{t-1}$ ; et  $\Delta^d$  est un opérateur de différence de degré d. Le terme  $\Delta^d \cdot x_t$  est une différenciation fractionnaire d'ordre d, définie par un développement en série, tel que :

$$\Delta^d x_t = x_t - d \cdot x_{t-1} + \frac{d(d-1)}{2} \cdot x_{t-2} + \dots + (-1)^n \frac{d(d-1)\dots(d-n+1)}{n!} \cdot x_{t-n} + \dots \tag{4}$$

où d est généralement égale à un, la série différenciée devient :  $\Delta \cdot x_t = x_t - x_{t-1}$ .

Le principal problème est celui de la stratégie de modélisation, c'est à dire le choix des paramètres p, d et q. L'organigramme de la figure 2 décrit la technique optée pour la construction d'un modèle ARIMA. Elle est composée de trois étapes essentielles : identification, estimation et vérification.

La première étape consiste d'abord, à analyser la tendance pour s'assurer de la stationnarité de la série  $x_t$ , et ensuite, à identifier le type et l'ordre du modèle AR MA. Ceci, en analysant les séquences de la FAC (fonction d'autocorrélation) et FACP (fonction d'auto corrélation partielle) et en recherchant les niveaux de coupures de ces dernières avec les intervalles de confiance, déterminés par  $a(r_k)$  et  $cr(<I>_{k,k})$  respectivement [Bosq et Lecoutre, 1992]. Pour les processus AR(p) et MA(q), ces intervalles valent  $1 / \sqrt{n}$  [Box & Jenkins, 76]. Pour un décalage k d'une série  $x_t$ , qui est centrée et réduite ( $x_t = 0, cr^{\wedge} = 1$ ), l'auto corrélation  $r_R$  ainsi que l'erreur-type associée sont définies par :

$$r_k = \frac{\sum_{t=1}^{n-k} x_t \cdot x_{t+k}}{\sum_{t=1}^n x_t^2}, \quad \text{Err-Type } (r_k) = \sqrt{\frac{n-k}{n(n+2)}} \tag{5}$$

Les auto corrélations partielles  $\phi_{kk}$  sont calculées selon les formules de récurrence suivantes :

$$\phi_{11} = r_1 ; \phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} ; \phi_{kk} = (r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}) / (1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j) \quad (6)$$

L'erreur approchée de vaut  $1/\sqrt{n}$  où n est le nombre d'observations dans la série.

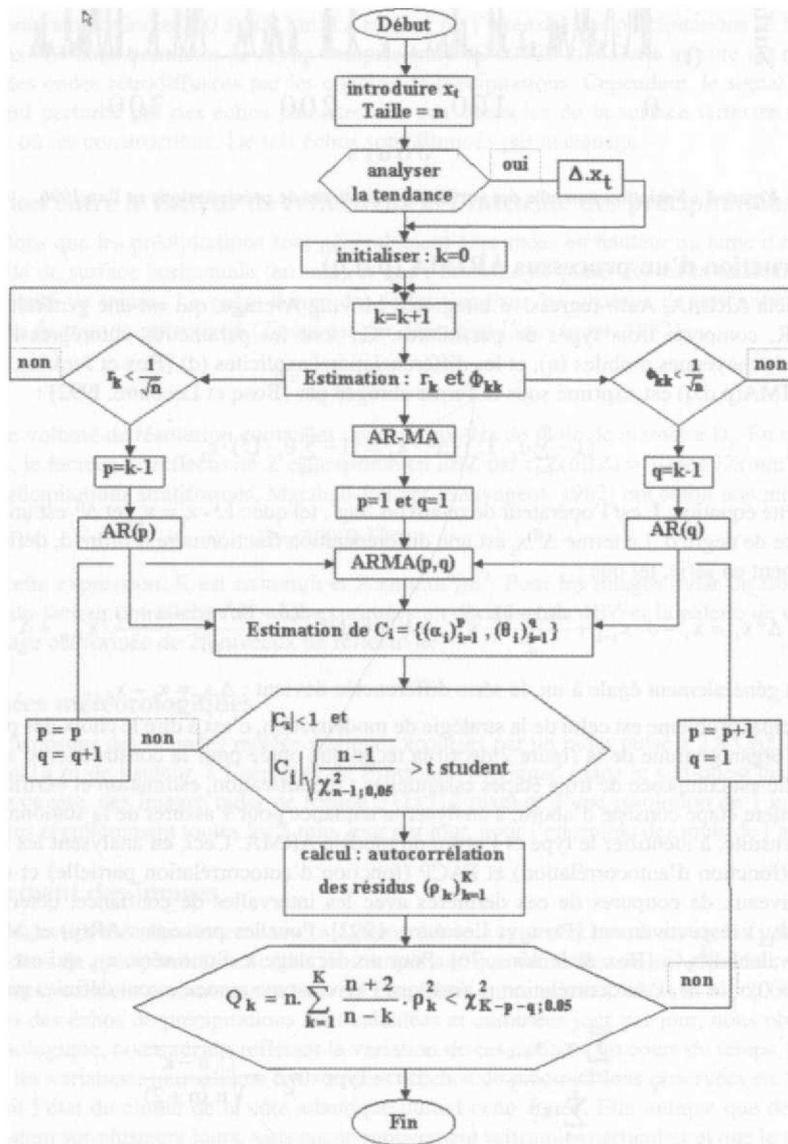


Figure 2 : Organigramme de la procédure de modélisation par des processus ARIMA.

Dans la deuxième étape, nous estimons les paramètres du modèle ainsi identifiés. Si ce modèle est un ARMA(p,q), nous devons estimer l'ensemble  $\{ \hat{a}_i, \hat{\theta}_j \}_{i=1}^p, j=1}^q$  formé de (p+q) paramètres.

Sous l'hypothèse que la population des résidus forme un bruit blanc gaussien, le critère employé dans la phase d'estimation est :

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (x_i - \hat{x}_i)^2 = 0 \tag{7}$$

où  $\hat{x}_i$  est l'estimation, par le modèle identifié, de la  $i^{\text{ème}}$  observation ( $x_i$ ) de la série considérée.

Pour un modèle ARMA(1,0,1), on aura :  $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (x_i - (\alpha_1 x_{i-1} - \theta_1 \varepsilon_{i-1}))^2$  et les valeurs de  $\alpha_1$  et  $\theta_1$  qui minimisent cette somme doivent donc vérifier :  $\frac{\partial \sum \varepsilon_i^2}{\partial \alpha_1} = \frac{\partial \sum \varepsilon_i^2}{\partial \theta_1} = 0$  (8)

Ces conditions s'écrivent aussi :

$$\begin{cases} \sum x_i \cdot x_{i-1} - \alpha_1 \sum x_{i-1}^2 + \theta_1 \sum x_{i-1} \cdot \varepsilon_{i-1} = 0 \\ \sum x_i \cdot \varepsilon_{i-1} - \alpha_1 \sum x_{i-1} \cdot \varepsilon_{i-1} + \theta_1 \sum \varepsilon_{i-1}^2 = 0 \end{cases} \tag{9}$$

Pour un AR(1), par exemple, la solution du système défini par la relation (9) est  $\alpha_1 = \frac{\sum x_i \cdot x_{i-1}}{\sum x_{i-1}^2}$ . L'incertitude commise sur l'estimation de chaque paramètre est donnée par :

$$r^2 = 1 - \frac{\sum (\hat{x}_i - x_i)^2}{\sum x_i^2} = \frac{\sum \hat{x}_i^2}{\sum x_i^2} \tag{10}$$

Les valeurs des paramètres  $C_j$ , issues de cette estimation, doivent être comprises entre -1 et 1 (bornes de stationnarité). Ils doivent aussi être statistiquement significatifs sous le critère que  $|C_j| \geq t_{\alpha/2, n-k} \sqrt{\frac{1}{n-k}}$  supérieure au t de Student, déterminé par la table, pour degrés de liberté et un seuil de confiance de 95 % [Bosq et Lecoutre, 1992]. S'ils ne répondent pas à ces conditions, l'ordre de la moyenne mobile est augmenté pour entamer une nouvelle estimation.

La dernière étape est relative à l'adéquation statistique du modèle retenu avec les observations. Elle s'appuie sur l'analyse des résidus qui doivent suivre un bruit blanc. On calcule alors les K premières autocorrélations des résidus où  $K = 13$ , ce nombre est assez grand devant  $n^{0.9}$  et ne dépasse pas 20 d'après la littérature [Bosq et Lecoutre, 1992]. Ensuite, on utilise la statistique de Box-Ljung qui donne une meilleure approximation pour les petites tailles d'échantillon. Cette statistique, définie par  $Q_k = n \sum_{k=1}^K \frac{\hat{\rho}_k^2}{1-k}$  suit approximativement une loi du khi-deux à K-p-q degrés de liberté,  $\alpha \in [0,1]$  étant le risque de première espèce associé à ce test. Si l'hypothèse d'indépendance est rejetée, nous identifions un nouveau modèle dont le degré de mémoire sera augmenté et le cycle d'estimation recommence.

### 5. Application du modèle AR

La série, formée de 365 observations a été modélisée à l'aide d'un modèle AR(p) d'équation :

$$x_t = \varepsilon_t + \sum_{i=1}^p \alpha_i \cdot x_{t-i} \tag{H}$$

Dans cette équation,  $x_t, x_{t-1}, \dots, x_{t-p}$ , sont des variables aléatoires prises aux instants  $t, \dots, t-p$ , ap...,  $\alpha_p$  sont les paramètres du processus autorégressif et  $\varepsilon_t$  est le bruit blanc. La condition de stationnarité de ce processus est :  $-1 < \alpha_i < +1$ . Le programme que nous avons élaboré pour générer le modèle AR(p) et modéliser la série est décrit par l'organigramme de la figure 2. Les paramètres du modèle et leur

nombre ont été déterminés en calculant les fonctions d'autocorrélation  $r_k$  (FAC) et d'autocorrélation partielle  $\phi_{kk}$  (FACP). La figure 3 montre que la FAC décroît selon une loi exponentielle et que la contribution la plus importante est apportée par le terme du premier ordre. Pour les ordres supérieurs, les coefficients d'autocorrélation partielle sont pratiquement nuls dans l'intervalle de confiance marqué par des traits discontinus (figure 3). Le modèle AR(1) étant ainsi identifié [Box et Jenkins, 1976], ses paramètres sont calculés en utilisant la technique des moindres carrés, décrite ci-dessus, par les équations (7), (8), (9) et (10). On trouve que  $a_1$  vaut 0,41 avec une erreur d'estimation de 0,05. Ce paramètre est significatif puisque  $t_{calculé}$  valant 9,93 est supérieure au  $t$  de Student qui est de 1,96, pour 364 degrés de liberté avec un risque de se tromper inférieur ou égale à 0,05. Un test d'adéquation basé sur la statistique de Box-Ljung a aussi été effectué. Au seuil de signification de 5%, on admet la validité de l'ajustement car  $Q_k$  qui vaut 13,6 est inférieure au quantile de 0,95 de la loi du khi-deux à 12 degrés de liberté. Il montre que le modèle AR(1) produit des résidus statistiquement indépendants.

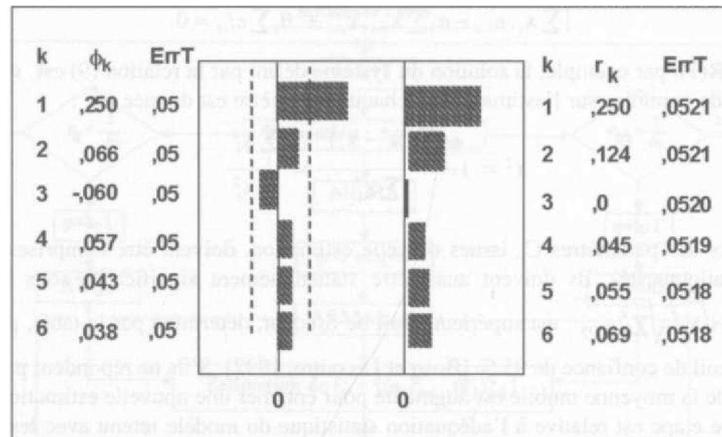


Figure 3 : Valeurs des coefficients d'autocorrélation ( $r_k$ ) et d'autocorrélation partielle ( $\phi_k$ ) pour le modèle AR.

## 6. Application du modèle ARIMA

Le modèle AR(p) ne peut pas être appliqué aux données formant la série<sub>2</sub> car celles-ci sont de durée beaucoup plus courte que la journée et forment un processus aléatoire non stationnaire. La figure 4 montre que cette série présente une tendance linéaire. Une première différence appliquée sur la série<sub>2</sub> éliminera cette dernière.

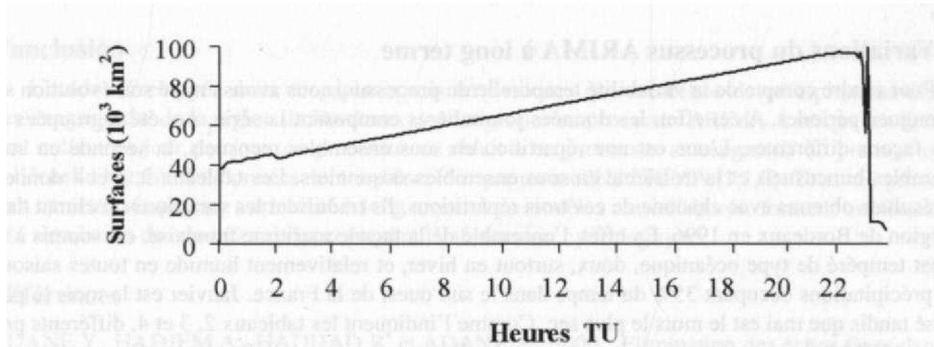


Figure 4 : Variations horaires des surfaces de précipitations pour la journée type du 21 juin 1996.

Dans ce cas, le modèle A R I M A semble se prêter mieux à la description d'un tel processus. Le programme décrit par la figure 2, permet aussi de calculer les paramètres du modèle ARIMA. Comme pour le modèle AR, les paramètres de ce modèle et leur nombre ont été déterminés en calculant les fonctions d'autocorrélation (FAC) et d'auto corrélation partielle (FACP). Les tracés de ces fonctions sont illustrés par la figure 5.

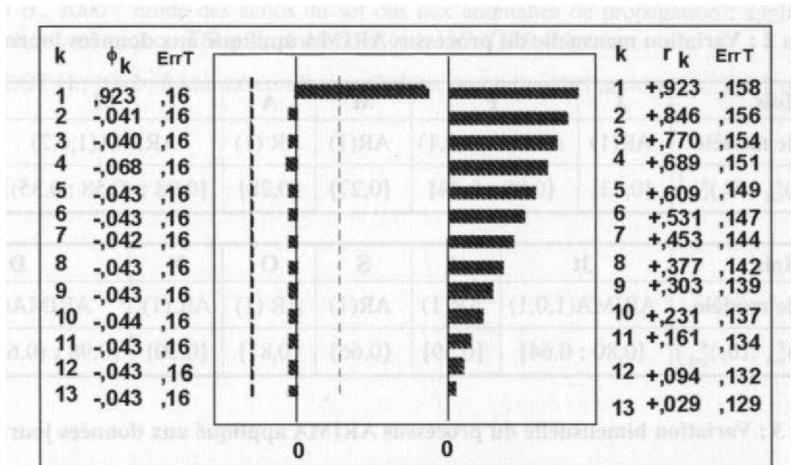


Figure 5 : Valeurs des coefficients d'autocorrélation ( $r_k$ ) et d'autocorrélation partielle ( $\phi_k$ ) pour le modèle ARIMA.

On trouve que la série, composée de 288 observations est représentée convenablement par un modèle ARIMA( 1,1,0) où le paramètre ( $a_1$ ) a été obtenu avec une erreur-type de 0,05 et vaut 0,97. Les résultats de ce type de modélisation satisfont au test de Student et au test de Box-Ljung du point de vue des hypothèses et de l'adéquation (tableau 1).

Tableau 1 : Résultats des tests de significativité et d'adéquation pour le modèle ARIMA( 1,1,0).

Type série	modèle	v = n-1	$\hat{a}_{calculé}$	tv; 0,05	$v^v = K-p$	QK	$5C^2v-0,05$
Série2	ARIMA( 1,1,0)	287	18,03	1,96	12	5,00	21,00

## 7. Variations du processus ARIMA à long terme

Pour rendre compte de la variabilité temporelle du processus, nous avons étudié son évolution sur de longues périodes. A cet effet, les données journalières composant la série, ont été regroupées en trois façons différentes. L'une est une répartition en sous ensembles mensuels, la seconde en sous ensembles bimensuels et la troisième en sous ensembles saisonniers. Les tableaux 2, 3 et 4 donnent les résultats obtenus avec chacune de ces trois répartitions. Ils traduisent les variations du climat dans la région de Bordeaux en 1996. En effet, l'ensemble de la façade maritime française, est soumis à un climat tempéré de type océanique, doux, surtout en hiver, et relativement humide en toutes saisons. Les précipitations occupent 35% du temps dans le sud ouest de la France. Janvier est le mois le plus arrosé tandis que mai est le mois le plus sec. Comme l'indiquent les tableaux 2, 3 et 4, différents processus ARIMA représentent toutes les nuances de climat de la région de Bordeaux. Ces modèles ont été déterminés grâce au programme décrit par l'organigramme de la figure 2. D'après le tableau 2, les variations journalières des surfaces de précipitations observées en janvier, août, septembre et octobre, sont très bien décrites par un modèle AR(1). Pour le mois de juin, aucun processus n'a pu être identifié car les données sont en nombre réduit et se sont concentrées sur le 21 juin 1996, ce jour ayant été pris comme référence pour l'étude à l'échelle horaire. De même, aucun processus n'a été identifié pour la période mai - juin étant donné le manque d'échos de pluie (tableau 3). Le tableau 4 montre que la saison d'hiver est modélisée par un processus autorégressif du premier ordre et les autres saisons sont caractérisées par un processus ARIMA(1,0,1) à paramètres très élevés.

**Tableau 2 : Variation mensuelle du processus ARIMA appliqué aux données journalières.**

Mois	J	F	M	A	M	J
Type de modèle	AR(1)	ARIMA(1,0,1)	AR(1)	AR(1)	ARIMA(1,0,2)	
$C = \{(oc,)_E, ;(e,)^{\wedge}\}$	{0,73}	{0,98 ; 0,84}	{0,27}	{0,28}	{0,98 ; (0,58 ; 0,35)}	
Mois	Jt	A	S	O	N	D
Type de modèle	ARIMA(1,0,1)	AR(1)	AR(1)	AR(1)	AR(1)	ARIMA(1,0,2)
$c=k)f=i-.(\gg i)?=.$	{0,80 ; 0,64}	{0,59}	{0,66}	{0,87}	{0,30}	{0,98 ; (0,65 ; 0,26)}

**Tableau 3 : Variation bimensuelle du processus ARIMA appliqué aux données journalières.**

Deux mois	J-F	M-A	M-J	Jt-A	S-O	N-D
Type de modèle	AR(1)	ARIMA(2,0,2)	*	ARIMA(1,0,1)	AR(1)	ARIMA(1,0,1)
$C - fa > w ;(e,)^{\wedge}\}$	{0,71}	{(0,42 ; 0,58) ; (0,24 ; 0,58)}	*	{0,82 ; 0,58}	{0,54}	{0,98 ; 0,92}

**Tableau 4 : Variation saisonnière du processus ARIMA appliqué aux données journalières.**

Saison	printemps	été	automne	hiver
Type de modèle	ARIMA(1,0,1)	ARIMA(1,0,1)	ARIMA(1,0,1)	AR(1)
$C = \}a_i!L. XQ^{\wedge}j$	{0,97 ; 0,86}	{0,79 ; 0,48}	{0,98 ; 0,91}	{0,64}

## Conclusion

L'étude développée dans ce papier a mis en évidence que les échos de précipitations extraits des images radar de la région de Bordeaux, sont régis par un processus ARIMA. Lorsque les données sont collectées jour par jour, ce processus se réduit à un processus autorégressif du premier ordre.

A l'échelle horaire, les données sont représentées par un modèle ARIMA (1,1,0). Tous ces processus sont caractérisés par une variabilité à long terme. Les modèles ainsi élaborés, forment une base utile à la simulation des données radar contenant l'information hydrologique.

## Références

- AD ANE Y., HADJEM A., HADDAD B. et AD ANE A., 2000 : Elimination des échos fixes dans les images radar utilisées pour l'observation météorologique. Proc. DAT 2000, Réghaia, Alger.
- B O S Q D. et L E C O U T R E J.P., 1992 : Analyse et prévision des séries chronologiques. Ed. Masson, Paris.
- BOX G.E.P. and JENKINS G.M., 1976 : Time séries analysis, forecasting and control. Ed. Holden Dey, San Francisco.
- D E V U Y S T P., 1972 : La météorologie : comprendre et interpréter. Ed. Eyrolles, Paris.
- H A D D A D B., 2000 : Etude des échos du sol dus aux anomalies de propagation : application aux radars météorologiques. Thèse de doctorat d'état, U.S.T.H.B-Algérie.
- SAUVAGEOT H., 1982 : Radarmétéorologie, télédétection active de l'atmosphère. Ed. Dunod, Paris.