

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Université Abderrahmane Mira de Bejaia**

**Faculté de la Technologie**

**Département d'Electronique**

MEMOIRE

Présenté en vue de l'obtention du diplôme de

**MAGISTER**

Filière: Automatique et Traitement de Signal

Option: Système

par :

**M<sup>me</sup> BELAID SABRINA**

Ingénieur en Electronique option contrôle

Thème

**MODELISATION DES DONNEES DE L'IRRAIATION SOLAIRE**

**CAS DU SITE GHARDAIA**

<b>Jury</b>	<b>Grade</b>	<b>Affiliation</b>	<b>Qualité</b>
<b>Mr. Mr Y.ZEBOUDJ</b>	<b>Professeur</b>	<b>Univ-Bejaia</b>	<b>Président</b>
<b>Mr. K. MOKRANI</b>	<b>M.C.A</b>	<b>Univ-Bejaia</b>	<b>Rapporteur</b>
<b>Mr. B.MENDIL</b>	<b>Professeur</b>	<b>Univ-Bejaia</b>	<b>Examineur</b>
<b>Mr. S.BERRAH</b>	<b>M.C.A</b>	<b>Univ-Bejaia</b>	<b>Examineur</b>
<b>Mr. A. Alliche</b>	<b>M.A</b>	<b>Univ-Bejaia</b>	<b>Invité</b>

2010

## **REMERCIEMENTS**

*Avant tout, je remercie DIEU le Tout-puissant de m'avoir donné le courage, la volonté, la patience et la santé durant toutes ces années d'étude et que grâce à lui ce travail a pu être réalisé.*

*A la mémoire de mon chère PERE sans qui je ne serai pas là aujourd'hui, il m'a transmis la volonté et le courage d'aller toujours en avant ! Que dieux l'accueille dans son vaste paradis.*

*Mes remerciements bien sur à mon adorable MAMA. Je sais que je ne pourrais jamais te rendre ce que tu m'as donné, mais comme vous le dites papa et toi « notre réussite dans la vie c'est votre bonheur ». « Que Dieu te protège et te garde en bonne santé ».*

*Je remercie mon époux qui a toujours cru en moi, et être à mes côtés pour le meilleur et pour le pire. A nos deux adorables enfants.*

*Je tiens à exprimer mon remerciement et ma gratitude à mes encadreurs Mr. MOKRANI KARIM, Maître de conférence à l'université de Bejaia et Mr ALLICHE ABDENOUR de m'avoir proposé ce sujet de thèse et d'être patients avec moi. Je leurs suis très reconnaissante.*

*Mes remerciements vont de mêmes aux membres de jury qui m'ont fait l'honneur de participer au jury de mon mémoire, il s'agit en l'occurrence de :*

*Mr. Y.ZEBOUDJ : Professeur de l'université de Bejaia.*

*Mr. B.MENDIL : Professeur de l'université de Bejaia.*

*Mr. S. BERRAH : Maître de conférences de l'université de Bejaia.*

*Je remercie tous ceux qui m'ont soutenu de pré ou de loin a la réalisation de mon mémoire et je ne peux clore mes remerciements, sans citer mes frères et sœurs.*

*« Que DIEU guide nos pas dans la voie du savoir, de la vérité et de la justice. »*

# SOMMAIRE

## Liste des symboles

## Introduction générale

1

## Chapitre1. Irradiation solaire et états de l'art

I. Les différentes approches utilisées pour l'estimation du rayonnement solaire	3
I.1 Les modèles empiriques	3
I.1.1 Le modèle d'Angström	3
I.1.2 Le modèle de Coppolino	4
I.1.3 Le modèle de Collarès-Pareira	4
I.1.4 Le modèle de Liu Jordan	4
I.1.5 Le modèle de Perrin Brichambaut	5
I.2. Les modèles stochastiques	7
II. Bibliographie	8
III. Notions fondamentales sur l'énergie solaire	12
III.1 Généralités	12
III.2 Influence de l'atmosphère terrestre sur le rayonnement solaire	14
III.3 La constante solaire	15
III.4 L'irradiation solaire hors atmosphère	16
III.4.1 Calcul de l'irradiation solaire globale hors atmosphère	16
III.4.2 Application pour le site de Ghardaïa	17

## Chapitre2. Analyse des séries temporelles

I. Introduction	19
I.1 Histoire des séries temporelles	19
I.2 Développement de l'analyse des séries temporelles	20
II. Rappels de Probabilité et de Statistiques	20
II.1 Introduction :	20
II.2 La population des moments théoriques	21
II.2.1 l'espérance mathématique	22
II.2.2 La variance	23
II.2.3 Le Kurtosis	23
II.2.4 Le Skewness	23
III. Processus aléatoire stationnaire	24

III.1	La stationnarité au sens stricte	24
III.2	La stationnarité au sens large	24
III.3	Le Théorème de Wold	25
III.3.1	Enoncée du théorème de la décomposition de Wold	25
III.4	Définition de l'opérateur retard	25
III.5	Théorème de Wold et polynômes retard	26
III.6	Les processus ARMA	27
III.6.1	Introduction	27
III.6.2	Le processus AR(p)	27
III.6.2.1	La moyenne du processus AR(p)	28
III.6.2.2	La variance du processus AR(p)	29
III.6.3	Le processus MA	31
III.6.3.1	La moyenne du processus MA(q)	31
III.6.3.2	La variance du processus MA(q)	31
III.6.4	Le processus ARMA (p,q)	32
III.6.4.1	La moyenne du processus ARMA	33
III.6.4.2	La variance du processus ARMA	33
IV.	Processus aléatoire non stationnaire	34
IV.1	Introduction	34
IV.2	Le processus TS	34
IV.3	Le processus DS	36
IV.4	Détection de la non stationnarité et sa nature	37
IV.5	Test de Dickey-Fuller augmenté	38
IV.5.1	La stratégie du test de racine unitaire	39
IV.6	Les processus ARIMA (p,d,q)	40
IV.7	Les processus SARIMA (p,d,q)(P,D,Q)	40

### **Chapitre3. La méthodologie de Box & Jenkins**

I.	Introduction	42
II.	L'identification	42
III.	L'estimation	45
III.1	Système d'équations de Yule-Walker :	45
III.2	Méthode du maximum de vraisemblance	46

IV Tests de validation du modèle	47
IV.1 La parcimonie :	47
IV.2 Test sur les résidus	48
IV.2.1 Test d'autocorrélations des résidus	48
IV.2.2 Test de normalité des résidus	49
V La prévision	49
V.1 La prévision dans le cas d'un modèle AR(p)	49
V.2 La prévision dans le cas d'un modèle MA(q)	50
V.2.1 Prévision à l'aide d'un processus $AR(\infty)$ pour un MA(q)	51
V.3 La prévision dans le cas d'un modèle ARMA (p,q)	51
V.4 La prévision dans le cas d'un processus ARIMA (p,d,q)	52
V.4.1 Prévision d'un ARIMA (p,d,q) en utilisant l'approximation AR	53
V.4.2 Prévision d'un ARIMA (p,d,q) en utilisant l'approximation MA	54
V.5 Prévision dans le cas des modèles SARIMA	55

## **Chapitre4. Analyse et résultats**

I. Introduction	58
II. Etape d'identification	58
III. Estimation des paramètres	62
III.1 Principe de calcul	62
IV. La prévision	65
IV.1 Résultats de la prévision des moyennes journalières de l'irradiation globale	66
IV.2 Test de validation du modèle	67
V. Résultats et discussion	68
Conclusion 1	73
VI. Autres techniques de modélisations	73
Conclusion 2	88
<b>Conclusion Générale</b>	89
<b>Perspectives</b>	90
<b>Références Bibliographiques</b>	91
<b>Annexes</b>	

## La nomenclature

- ✓  $G$  : L'irradiation globale exprimée en Wh/m<sup>2</sup>.
- ✓  $G_0$  : L'irradiation globale hors atmosphère exprimée en Wh/m<sup>2</sup>.
- ✓  $S$  : La durée d'insolation effective exprimée en heure et dixième d'heure.
- ✓  $S_0$  : La durée maximale d'insolation exprimée en heure et dixième d'heure.
- ✓  $\omega$  : L'angle horaire.
- ✓  $\omega_s$  : Le temps solaire vrai.
- ✓  $\theta_z$  : La distance zénithale (°)
- ✓  $\Theta$  : L'angle d'incidence (°)
- ✓  $\beta$  : L'angle optimal annuel d'inclinaison (°)
- ✓  $G_G$  : L'éclairement global reçu sur un plan horizontal (W/m<sup>2</sup>)
- ✓  $D_G$  : L'éclairement direct mesuré sur un plan horizontal (W/m<sup>2</sup>)
- ✓  $DF_G$  : L'éclairement diffus mesuré sur un plan horizontal (W/m<sup>2</sup>)
- ✓  $E_G$  : L'éclairement global reçu sur un plan incliné (W/m<sup>2</sup>)
- ✓  $K, K', K'', M, M''$  : sont des constantes qui dépendent de l'état de l'atmosphère propres au modèle de Perrin Brichambaut
- ✓  $\lambda$  : La longueur d'onde du photon (micro mètre)
- ✓  $h$  : La constante de Planck.
- ✓  $c$  : La célérité de la lumière (m/s)
- ✓  $m$  : Le nombre d'Air Masse
- ✓  $p$  : La pression (N/m<sup>2</sup>)
- ✓  $A(^{\circ})$  : L'élévation du soleil sur l'horizon (90° au zénith)
- ✓  $Z$  : L'altitude (Km)
- ✓  $I_0$  : La constante solaire,  $I_0 = 1367(W/m^2)$
- ✓  $N$  : Le numéro du jour à partir du 1<sup>er</sup> Janvier (on considère 365 jours dans l'année)

- ✓  $\lambda$  : La latitude du lieu
- ✓  $\delta$  : La déclinaison du soleil
- ✓  $E_t$  : La série temporelle observée qui est la variable irradiation globale journalière moyenne.
- ✓  $G_t$  : La transformée logarithmique de  $E_t$
- ✓  $X_t$  : La série différenciée d'ordre 365 de  $G_t$  ;  $X_t = G_t(1 - L^{365})$
- ✓  $U_t$  : La composante périodique décomposée en coefficients de Fourier
- ✓  $Y_t$  : Le résidu entre la série observée et la composante périodique.
- ✓  $L(\theta)$  : La fonction log vraisemblance.
- ✓  $f_X(x)$  : La fonction de densité
- ✓  $E(X^k)$  : La population des moments associée à tout ordre  $k \in \mathbb{N}$
- ✓  $\mu_k$  : La population des moments centrés
- ✓  $E[X_t] = \mu$  : L'espérance mathématique de  $x$
- ✓  $V[X_t] = \sigma_x^2$  : La variance de la série  $x$
- ✓  $cov[X_t, X_{t-k}] = \gamma_k$  : La covariance de  $x$  au retard  $k$
- ✓  $\psi_j$  : Les coefficients de pondérations du bruit blanc
- ✓  $L$  : Opérateur retard
- ✓  $(\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p)$  : Les coefficients du polynôme autorégressif AR
- ✓  $(\theta_0, \theta_1, \theta_2, \dots, \theta_q) \in \mathbb{R}^{q+1}$  : coefficients du polynôme moyenne mobile MA
- ✓  $\Phi(L^S)$  et  $\vartheta(L^S)$  : Des polynômes qui caractérisent respectivement les parties AR et MA saisonnières.
- ✓  $\varphi_0$  : constante
- ✓ ACF : La fonction d'autocorrélation
- ✓ PACF : La fonction d'autocorrélation partielle
- ✓ RACF : La fonction d'autocorrélation des résidus
- ✓  $\rho_k$  : La fonction d'autocorrélation ACF

- ✓  $\gamma_0$  : La variance de  $X$
- ✓  $\rho_{ii}$  : Le coefficient d'ACP autocorrélation partiel
- ✓  $\varepsilon_t$  : Le bruit blanc
- ✓  $\beta, a, \alpha_i$  : Des coefficients propre au test de Dickey-Fuller
- ✓  $t$  : La statistique de Student
- ✓  $\nabla$  : L'opérateur retard
- ✓ AIC : Le critère d'information d'Akaike
- ✓ BIC : Le Critère d'information Bayésien
- ✓  $Q(r)$  : La statistique de Ljung-Box
- ✓  $r_k(\varepsilon)$  : La fonction d'autocorrélation ACF du Bruit  $\varepsilon_k$  au retard  $k$
- ✓  $T$  : La taille de l'échantillon
- ✓  $X_{T+h}^*$  : La prévision de la variable  $X$  à l'horizon  $h$
- ✓  $\widehat{X}_{T+h}^*$  : L'estimée de la prévision de la variable  $X$  à l'horizon  $h$



# Introduction générale

---

## **Introduction générale :**

Les besoins énergétiques de l'humanité n'ont pas cessé de croître avec son évolution : on a utilisé d'abord des ressources naturelles telle que le charbon, le bois et par la suite avec le progrès scientifique, le pétrole, le gaz, l'énergie hydraulique, et l'énergie nucléaire.

Actuellement les énergies fossiles (pétrole et gaz) sont consommées bien plus rapidement qu'elles ne se forment dans la nature, et selon les estimations les réserves mondiales seront épuisées vers 2030 si la consommation n'est pas radicalement modifiée, et au plus tard vers 2100 si des efforts sont fournis sur la production et la consommation nécessaire [1]. Cette forme d'énergie étant limitée dans le temps, il s'avère nécessaire de trouver une autre solution pour prendre le relais. La contrainte imposée est d'utiliser d'autres sources d'énergie, tout en tenant compte des facteurs économiques et environnementaux.

L'énergie nucléaire occupe une place grandissante dans la production mondiale d'énergie car elle permet de produire énormément d'énergie suffisante pour couvrir les besoins des mégapoles, mais elle demeure toujours une solution à risques (exemple Tchernobyl ?). De plus, les déchets nucléaires posent de véritables problèmes pour l'environnement. Enfin ces centrales nucléaires ne permettent pas de couvrir les besoins de petits sites isolés.

L'énergie solaire qui est l'objectif de notre étude se caractérise au contraire par une absence de pollution et une disponibilité exceptionnelle. D'autre part, les systèmes d'exploitations qui utilisent cette forme d'énergie demandent une maintenance légère et présente une bonne fiabilité de fonctionnement, une autonomie de plus en plus accrue, une résistance extrême aux conditions naturelles (température, humidité, vent, corrosion,...etc.) et donc une grande longévité. Il apparaît dès lors que l'énergie solaire peut apporter de réelles solutions.

Au niveau mondial, le développement des énergies renouvelables, notamment l'énergie solaire, est d'une très grande importance et qui ne cesse d'accroître, d'ailleurs on l'appelle l'énergie du futur. En Algérie, cette énergie propre a bénéficiée ces dernières années des préoccupations des pouvoirs publics vis son important potentiel en gisement solaire et sa disponibilité, aussi en prévision de la future pénurie des ressources fossiles et surtout des problèmes écologiques qui résultent de leur utilisation.

L'exploitation du gisement solaire intéresse de plus en plus la communauté scientifique. Cependant le dimensionnement et l'utilisation des équipements liés à cette ressource, et plus

particulièrement les dispositifs photovoltaïques ou thermiques, nécessitent la connaissance aussi bien quantitative que qualitative du gisement solaire disponible en un lieu donné au sol.

Ainsi, l'un des enjeux scientifiques de notre projet consiste en la modélisation suivie d'une prévision des moyennes journalières de l'irradiation globale sur le site de Ghardaïa où le gisement solaire est bien présent tout au long de l'année et le rayonnement atteint même des niveaux considérables.

Le modèle que nous avons adopté est le modèle stochastique SARMA (Seasonal Autoregressive Moving Average) de Box Jenkins. Le choix de ce modèle par rapport aux modèles empiriques, généralement utilisés, réside dans son aptitude à tenir compte de la majorité des phénomènes aléatoires qui peuvent influencer cette variable météorologique, tel que les nuages, la pluie,.....et surtout le vent de sable qui est un phénomène fréquent sur ce site de Ghardaïa (Sud de l'Algérie).

La validation de cette prévision sera réalisée par une étude comparative entre les valeurs calculées et les mesures directes recueillies sur le site de Ghardaïa, sur la période allant du 1 Juin 2009 au 31 Mai 2010.

A cet effet, notre travail est structuré en 4 chapitres. Un premier chapitre qui comprend une partie bibliographique consacré aux différents travaux menés dans l'axe de la modélisation des données de l'irradiation globale par les différentes méthodes (empiriques et stochastique) et une partie sur les notions fondamentales de l'énergie solaire. Un second chapitre qui détaille la théorie de l'analyse stochastique des séries temporelles. Dans un troisième chapitre, nous avons introduit la méthodologie de Box-Jenkins. Finalement et dans un dernier chapitre nous avons exposé et discuté les résultats issus de l'application de la dite méthode sur l'analyse des moyennes journalières de l'irradiation globale sur le site de Ghardaïa.

# Chapitre 1. Irradiation solaire et états de l'art

---

## I. Les différentes approches utilisées pour l'estimation du rayonnement solaire :

Afin d'évaluer le gisement solaire dans un site donné, il ya lieu d'évoquer les principaux modèles rencontrés dans la littérature et qui sont en fonction du type d'approche utilisée qu'ont peut classer en deux grandes classes essentielles ; modèles empiriques et modèles stochastiques

### I.1 Les modèles empiriques :

Ces modèles sont nombreux et permettent d'évaluer, Grace à des formules empiriques, le rayonnement solaire au sol à partir d'un nombre plus au moins réduit de paramètres astronomiques et/ou météorologiques. Parmi ces paramètres, on nommera ; les heures d'ensoleillement, l'humidité relative, la déclinaison solaire, la durée astronomique du jour, la constante solaire, la latitude, l'altitude, la pluviométrie, la température, etc....

Citons en exemples : le modèle de A.A.M. Sayigh [2], le modèle de Reddy [3], le modèle de Garg *et al.* [4], le modèle de Rietveld *et al.* [5], le modèle de A.S Sambo [6], Modèle de Liu & Jordan [7]...ces modèles sont tellement nombreux qu'on reprendra seulement quelques uns les plus rencontrés :

#### I.1.1 Le modèle d'Angström :

C'est un modèle prévisionnel de base, établi en 1924.

$$G = G_0 \left( A + B \frac{S}{S_0} \right) \quad (1)$$

Avec :

- ✓  $G$  : l'irradiation globale exprimée en Wh/m<sup>2</sup>.
- ✓  $G_0$  : l'irradiation globale hors atmosphère exprimée en Wh/m<sup>2</sup>.
- ✓  $S$  : durée d'insolation effective exprimée en heure et dixième d'heure.
- ✓  $S_0$  : durée maximale d'insolation exprimée en heure et dixième d'heure.
- ✓  $A$  et  $B$  sont les coefficients d'Angstrom à estimer, ils sont caractéristiques au site donnée et généralement validés avec une bonne précision.

### I.1.2 Le modèle de Coppolino :

Ce modèle permet de prédire la moyenne mensuelle de l'irradiation journalière globale  $G$  de n'importe quel site. Cette relation est liée à la durée théorique de l'ensoleillement  $S_t$  et à la hauteur du soleil à midi au 15<sup>ième</sup> jour du mois  $h_0$  exprimée en degré.

$$G = A S_t^{0.5} (\sin(h_0))^{1.15} \quad (2)$$

$A$  : est un coefficient à calculer et dépend du site.

### I.1.3 Le modèle de Collarès-Pareira :[34]

C'est un modèle qui permet d'évaluer le rapport entre l'irradiation horaire et la moyenne mensuelle de l'irradiation journalière

Pour l'irradiation globale, la formule est la suivante :

$$G = \frac{\pi}{24} (A + B \cdot \cos(\omega_s)) \left( \frac{\cos(\omega_s) - \cos(\omega)}{\sin(\omega) - \omega \sin(\omega)} \right) \quad (3)$$

- ✓  $\omega$  est l'angle horaire (en radians).
- ✓  $\omega_s$  est en fonction du temps solaire vrai  $t$  (en heure).  $\omega_s = 15(t - 12)$  (en degré)
- ✓  $A$  et  $B$  sont les coefficients à estimer et différent d'un site à un autre.

### I.1.4 Le modèle de Liu Jordan :[7]

Ce modèle nous permet de calculer l'éclairement global sur un plan incliné avec un angle  $\alpha$ , en tenant compte de l'éclairement direct  $D_G$  et de l'éclairement diffus  $DF_G$  mesurés sur un plan horizontal.

$$E_G = D_G \frac{\cos \theta}{\cos \theta_z} + DF_G \left( \frac{1 + \cos \beta}{2} \right) + G_G \cdot a \cdot \left( \frac{1 - \cos \beta}{2} \right) \quad (4)$$

Avec :

$a$  Albédo régional

$\theta_z$  Distance zénithale (°)

$\theta$  Angle d'incidence (°)

$\beta$  : Angle optimal annuel d'inclinaison (°)

$G_G$  : Eclairement global reçu sur un plan horizontal (W/m<sup>2</sup>)

$D_G$  : Eclairement direct mesuré sur un plan horizontal (W/m<sup>2</sup>)

$DF_G$  : Eclairement diffus mesuré sur un plan horizontal (W/m<sup>2</sup>)

$E_G$  : Eclairement global reçu sur un plan incliné (W/m<sup>2</sup>)

Il existe aussi d'autres modèles plus ou moins complexes qui nécessitent la connaissance exacte de certains phénomènes physiques tel que la constitution ou la description précise de l'atmosphère. Certains de nos collègues les chercheurs ont pensé à la décomposition de l'atmosphère en couches horizontales et étudient successivement les phénomènes d'absorption et de diffusion auxquels est soumis le rayonnement solaire par ces différentes couches atmosphériques, ainsi ils élaborent des modèles de calcul des différentes composantes du rayonnement solaire.

A titre d'exemple, nous allons voir le modèle de Perrin Brichambaut qui donne de très bons résultats comparés aux autres modèles :[9]

### **I.1.5 Le modèle de Perrin Brichambaut :**

La particularité de ce modèle c'est qu'il permet d'estimer l'énergie solaire reçue par un capteur pour différentes orientations  $\alpha$  ou angle d'inclinaison.

Les puissances des rayonnements ; direct  $E_d$ , diffus  $E_{df}$  et le globale  $E_G$ , exprimées en W/m<sup>2</sup>, sont données par les relations suivantes :

$$\left\{ \begin{array}{l} E_d = K \cos(i) e^{-\frac{1}{M \sin(h+2)}} \\ E_{df} = \left( \frac{1 + \cos(\alpha)}{2} \right) D_G + \left( \frac{1 + \cos(\alpha)}{2} \right) a \cdot G_G \\ E_G = E_d + E_{df} \end{array} \right. \quad (5)$$

$$\left\{ \begin{array}{l} D_G = K'(\sin(h))^{0.4} \\ G_G = K''(\sin(h))^{M''} \end{array} \right. \quad (6)$$

Tel que :

$D_G$ : Éclairement direct reçus par une surface horizontale.

$G_G$ : Éclairement global reçus par une surface horizontale.

$a$ : Est le coefficient de réflexion du sol, appelé albédo du sol et vari selon l'état du sol.

$$a = \left\{ \begin{array}{l} 0.9 \text{ à } 0.8 \text{ neige} \\ 0.8 \text{ à } 0.4 \text{ sol clair} \\ 0.4 \text{ à } 0.2 \text{ verdures} \end{array} \right. \quad (7)$$

$K, K', K'', M, M''$  sont des constantes qui dépendent de l'état de l'atmosphère :

Etat de l'atmosphère	$K$	$M$	$K'$	$K''$	$M''$
Ciel bleu foncé	1300	6	87	1150	1.15
Ciel bleu clair	1230	4	125	1080	1.22
Ciel bleu laiteux	1200	2.5	187	990	1.25



## I.2 Les modèles stochastiques :

Les modèles stochastiques sont de modèles mathématiques décrivant des phénomènes physiques aléatoires car les données d'observation de ces derniers font apparaître des trajectoires ayant des caractéristiques d'imprévisibilité et de régularité en même temps. Tel que la caractéristique d'imprévisibilité renvoie à la notion de chance, de hasard ou d'aléa. Elle signifie qu'une des réalisations du phénomène en question est entachée d'incertitude. Alors que la notion de régularité renvoie en revanche à la celle de prévisibilité du phénomène considéré dans son ensemble. Comme le souligne Spanos [1998], ces deux notions semblent a priori être antinomiques puisque l'imprévisibilité signifie le désordre contrairement à la régularité qui elle signifie l'ordre. Mais bien sur, nous constaterons en observant les différentes séries temporelles que c'est l'ordre qui domine.

De façon général, les modèles stochastique appartiennent à la classe des modèles ARMA ou ses dérivées (ARIMA, SARIMA, AR, MA,...) dont le principe est base sur la méthodologie de Box-Jenkins. L'idée est de générer séquentiellement la série temporelle à partir d'un bruit blanc et d'un filtre linéaire. Mais nous prendrons le soin de détailler cette méthode dans les prochains chapitres puisqu'elle fait objet de notre travail.

## II. Bibliographie :

Les modèles empiriques développés pour l'estimation des irradiances solaires globale que ça soit journalières, mensuelles ou horaires sont nombreux et comme nous l'avons déjà évoqué, ils exigent la connaissance de certains paramètres géographiques ou/et atmosphériques avec des formules plus complexes, Ces modèles restent applicables pour certaines régions et pas pour d'autres aussi leur application peut se résumer seulement à un ciel clair, par exemple pour l'Espagne ; Palomo (1989), pour l'Italy ; Mustacchi et al (1979), pour Athènes ; Balouktsis et Tsalides (1986), pour Singapour ; Goh et Tan (1977), pour le Canada ; Graham et al (1987) et pour les USA ; Knight (1988).[10]

Et pour ce qui est des travaux menés en Algérie pour l'estimation de l'irradiation solaire par les modèles empiriques, il ya lieu de citer quelque uns les plus importants :

M.Koussa et al [11] ont fait une analyse comparative des différents modèles proposés dans la littérature par différents scientifiques, et relatifs à plusieurs sites dans le monde pour l'estimations des différentes composantes (le global et le diffus) de l'irradiation solaire que se soit journalières ou mensuelles. Ces modèles ont été appliqué sur des valeurs réellement mesurées sur plusieurs sites du territoire national; Bouzaréah, comme site côtier, Batna sur les Hauts Plateaux, Ghardaïa et Adrar au Nord et au Sud du Sahara. Et comme résultats, ils ont constaté que seuls certains de ces modèles qui tendent à être en bonne concordance avec les valeurs mesurées. Ainsi les modèles retenus sont : les modèles de Hussain et al et de Garg et al pour l'estimation de l'irradiation globale et les modèles de Erbs, Klein et Duffie et de Liu & Jordan pour la reconstitution de la composante diffuse.

Il ya aussi les travaux de S. *Benkacali* et al [12] ou leurs choix s'est porté sur cinq modèles empiriques basés sur l'estimation du rayonnement global reçu sur un plan inclinés à la latitude du lieu de 32°, ainsi ils ont testé chacun de ces modèles et confronté les résultats calculés aux données expérimentales obtenues par des mesures prises à l'aide d'une station radio métrique installée a l'unité de recherches de Ghardaïa.

M.Cheggar et al [13] ont songé d'estimer l'irradiation solaire globale moyenne mensuelle sur une surface horizontale et sur le site de Mascara en utilisant plusieurs modèles. Il s'agit de la régression linéaire du type Angstrom-PreScott :  $G/G_0 = A + B(S/S_0)$ , de la forme quadratique ;  $G/G_0 = A + B(S/S_0) + C(S/S_0)^2$  de la forme logarithmique :  $G/G_0 = A + B \ln(S/S_0)$ , et d'une forme exponentielle :  $G/G_0 = A + B \exp(S/S_0)$ . Tous ces modèles sont

basés sur la fraction d'ensoleillement mesurée  $S/S_0$ . Une analyse statistique a été faite en utilisant les écarts relatifs moyens algébrique, absolu et quadratique. Ils sont arrivés à la conclusion que ces différents modèles peuvent être utilisés pour des sites dont le climat est similaire à celui de Mascara.

Récemment, les études penchent beaucoup plus sur les modèles stochastiques bien que leur emploi a commencé bien avant avec Klein(1976), Brinksmorth (1977), Pearson et al (1981), Exell et al (1981), Vergam-Dominguez (1985), Amato et al (1986), Graham et Hollands (1988),...etc. Tous ces derniers ont développés des modèles stochastiques pour des séquences d'irradiation à l'échelle journalière. Sur ceux on citera quelques exemples de travaux les plus récents ;

En 1994, A.Zeroual et al [14] ont développé deux modèles stochastiques présentant l'irradiation solaire globale quotidienne obtenue à partir de trois ans de données mesurées sur une surface horizontale à Marrakech, Maroc (latitude  $31^{\circ}37'N$ , longitude  $08^{\circ}02'W$ , altitude 463 m). Le développement de ces modèles est basé sur la périodicité annuelle et la variation saisonnière du rayonnement solaire en utilisant deux types de normalisation. Le premier modèle est développé en utilisant une décomposition classique de l'irradiation journalière comme somme de deux composantes: une composante tendancielle dont la modélisation se fait par la décomposition en série de Fourier à sept harmoniques et une composante stochastique qui est décrite par trois modèles autorégressifs AR(2), AR(1) et AR(3) représentant respectivement les périodes saisonnières : Janvier-Avril, Mai-août et Septembre-Décembre . Le deuxième modèle est développé en utilisant une variable adimensionnelle ; l'index de clarté, qui est modélisé comme processus stochastique après avoir transformé la variable en une variable réduite dont la moyenne nulle et la variance unité afin d'obtenir une série chronologique stationnaire qui sera ensuite modélisée avec le modèle autorégressif AR du premier ordre. Ces deux modèles mènent à des résultats satisfaisants et sont validés pour la prévision et la simulation des données de l'irradiation solaires globales journalières.

En 2006, Kadri Yûrekli [15] modélisent via la technique de Box et Jenkins les moyennes journalières de trois variables météorologiques en Turquie à savoir : la température, l'humidité et l'irradiation solaire en aboutissant à une modélisation ARIMA(2,1,1) pour la température , ARIMA(1,1,1) pour l'humidité et ARIMA(0,1,2) pour l'irradiation solaire journalière.

En 1996, M.Yusof Sulaiman et Al [16] ont appliqué le modèle ARIMA pour modéliser les données de l'irradiation journalières de 4 différentes stations météorologiques en Malaisie,

en appuyant leurs résultats par des tests statistiques des résidus, tel que, le test de Box-Pierce et de Ljung-Box. Pour vérifier la normalité des résidus, les coefficients de Skewness et de Kurtosis sont testés. Santos et al [17], se s'ont inspiré du modèle ARIMA pour analyser les séquences de données de l'indice de clarté journalière.

En 2001 P.K. Jain et al [18] proposent une analyse harmonique de la durée de l'ensoleillement et de l'irradiation solaire mesurées à Sebele. Les données utilisées comprennent les moyennes mensuelles ainsi que journalières. Leur étude repose sur la décomposition de la série chronologique en deux composants : déterministes et stochastiques. La composante stochastique est analysée par la technique de Box et Jenkins et on trouve que le modèle autorégressif de second ordre est le mieux adapté pour représenter les données de l'irradiation solaire et de la durée de l'ensoleillement.

Quand à la modélisation de l'irradiation solaire à l'échelle horaire, on parlera des plus importants et qui ont été menée par :

Boileau et al (1981 et 1985) [19], qui ont modélisé à l'aide de modèle ARMA la variable irradiation solaire globale au pas de temps de l'heure après avoir classé les données selon le beau ou le mauvais temps et transformé afin de rendre la variable gaussienne dans l'objectif d'améliorer les modèles de prédiction de l'irradiation solaire globale horaire.

Ensuite, en 1998 LL.Mora-Lopez et al [20] [21] proposent une méthodologie pour générer une série de données de l'irradiation globale horaire basée sur l'utilisation d'un modèle multiplicative ARMA(1,0)x(0,1) avec des composantes régulières et saisonnières qui changent d'un mois à l'autre dans dix différentes régions espagnoles.

On trouve aussi par la suite une succession d'études réalisées toutes aussi importantes dans ce même axe de recherche comme celles de Zeroual (1988), Graham et Hollands (1990), Aguiar et Collares-Pareira (1992),.....

Pour ce qui est des travaux effectués en Algérie dans la modélisation de l'irradiation solaire par les méthodes stochastiques sont rares. Mefti et Bouroubi [10] sont parmi les premiers chercheurs à contribuer à la modélisation stochastique via l'approche Box Jenkins des données de la fraction d'insolation journalière mesurées sur quelques sites algériens ainsi ils permettent d'identifier un modèle ARMA(1,0) pour Alger et un ARMA(2,0) pour Ghardaïa.

Ensuite Il ya eu Mellit et al, qui eux, ont développé des modèles hybrides en combinant les modèles stochastique aux modèles d'intelligence artificiel comme les réseaux de neurones afin d'estimer l'irradiation journalière globale dans différentes localités algériennes. Et A.Khireddine et al [22] qui ont élaboré un modèle permettant de décrire ou de reproduire les variations horaires et quotidiennes de l'irradiation solaire, ce modèle est basé sur l'emploi de processus AR (autoregressif).

Concernant notre travail, nous essayerons de développer au mieux ces modèles stochastiques par la méthodologie de Box-Jenkins en vu de modéliser et de prédire au mieux l'irradiation solaire globale journalière reçu sur une plan horizontale à Ghardaïa.

### III. Notions fondamentales sur l'énergie solaire

#### III.1 Généralités

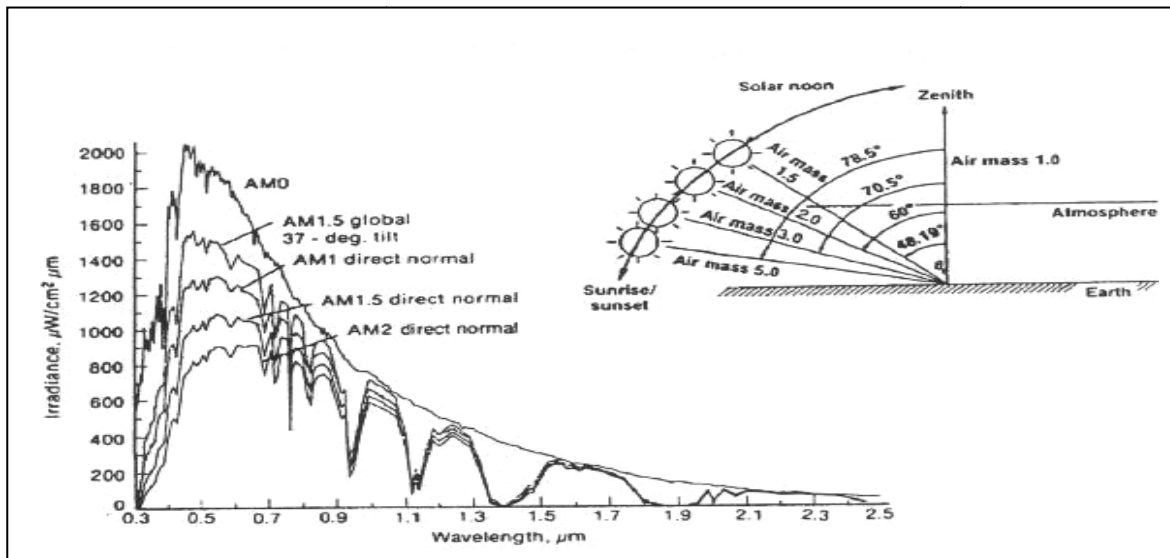
L'émission du rayonnement solaire dans chaque coin de l'espace apparaît sous forme d'ondes électromagnétiques qui portent l'énergie à la vitesse de la lumière. Nous rappelons qu'un rayonnement électromagnétique a un double aspect ; ondulatoire et corpusculaire. Le corpusculaire est le photon de masse nulle ; son énergie  $E$  est reliée à la fréquence  $\nu$  ou : [23]

$$E = h * \nu = \frac{h * c}{\lambda} \tag{8}$$

$\lambda$  est la longueur d'onde du photon et  $h$  est la constante de Planck.

Le rayonnement solaire est soit absorbé, reflété, ou répandu par les particules pleines dans n'importe quel endroit de l'espace et plus particulièrement par la terre. Selon la géométrie de la terre, sa distance du soleil, l'endroit géographique de tout point sur terre, coordonnées astronomiques, et la composition de l'atmosphère, l'irradiation solaire diffère d'un endroit à un autre.

Sa répartition spectrale (en longueur d'onde) est assez simple puisqu'elle est à peu près celle d'un corps noir à 5700 K.



**Figure 1** : L'allure du spectre ; flux de puissance en fonction de la longueur d'onde en microns. [24]

\* 8% de la puissance se situe dans l'ultra violet (0.2-0.4 μm).

\* 48% de la puissance dans le spectre visible (0.4-0.8  $\mu\text{m}$ ).

\* 37% de la puissance dans le proche infra rouge (0.8-2  $\mu\text{m}$ ).

Pour mieux comprendre la courbe de la Figure 1, nous définissons l'AM ou l'Air Masse qui est l'épaisseur de l'atmosphère que la lumière doit pénétrer (figure.5).

Cette épaisseur change d'heure en heure selon la position du soleil dans le ciel. Si le soleil est au zénith, la lumière traverserait une épaisseur de l'atmosphère ou un AM1 pour atteindre la surface de la terre. L'après-midi ou le matin, le soleil est à un angle inférieur, et la distance que doit traverser la lumière pourrait être deux fois ou plus que celle au zénith.

L'équation mathématique définissant le nombre d'Air Masse est comme suit : [26]

$$m = \frac{(p * e^{(-z/7.8)})}{(1013 * \sin(A))} \quad (9)$$

avec :

p : est la pression (N/m<sup>2</sup>)

A(°) : élévation du soleil sur l'horizon (90° au zénith)

Z : altitude (Km)

Dans les conditions normales et au niveau de la mer, l'expression simplifiée suivante est utilisée :  $m = 1/\sin(A)$

Ainsi, lorsque le soleil est au zénith, on dit qu'on a les conditions AM1 car les rayons lumineux traversent une atmosphère unité de 7.8 km et nombre Air Masse 1

Car :  $m = 1 / \sin (90^\circ) = 1$ .

Avec un soleil à 30° sur l'horizon, on obtient les conditions AM2

Car :  $m = 1 / \sin (30^\circ) = 2$ .

Hors atmosphère, à haute altitude, on définit les conditions AM0.

Ce nombre d'Air Masse va servir à définir un spectre de référence pour calibrer dans le domaine pratique les cellules étalonnées destinées à qualifier les performances des dispositifs photovoltaïques.

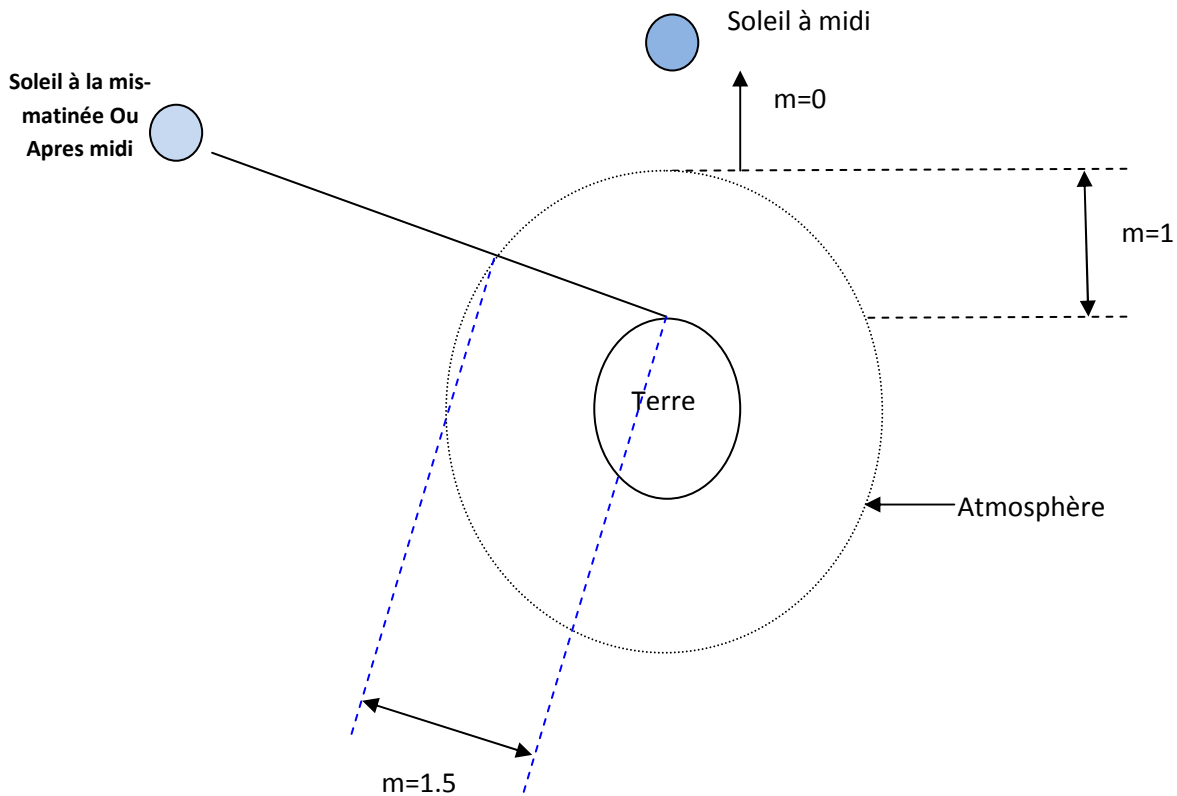


Fig.2 : L'angle et la distance du soleil par rapport à l'atmosphère

### III.2 Influence de l'atmosphère terrestre sur le rayonnement solaire

Hors atmosphère, le flux énergétique moyen reçu par la terre est d'environ  $1400 \text{ W/m}^2$  et par jour. En traversant l'atmosphère, le rayonnement solaire subit une atténuation et une modification de son spectre, à la suite de phénomènes complexes d'absorption et de diffusion par les poussières et les aérosols (Fig. 3) Ainsi, la couche d'ozone absorbe la majeure partie du rayonnement ultraviolet, tandis que la vapeur d'eau absorbe le rayonnement infrarouge. Le flux énergétique reçu au sol dépend ainsi de nombreux facteurs, et son maximum au sol ne dépasse guère  $1000 \text{ W/m}^2$ . [4]

En ramenant toutes les couches de l'atmosphère dans des conditions normales ( $p = 1013 \text{ mbar}$  et  $T = 0^\circ\text{C}$ ), on a défini une atmosphère standard d'épaisseur verticale moyenne de  $7.8 \text{ km}$  prise pour référence unité et formée de couches planes et stratifiées par les divers gaz comme l'azote (couche de  $1650 \text{ m}$ ), l'argon ( $74 \text{ m}$ ), le gaz carbonique ( $24 \text{ m}$ )... L'eau est



représentée par une couche d'épaisseur variable de quelques dizaines de mètres pour la vapeur et de quelques centimètres pour le liquide. [25]

Le rayonnement global au sol est donc fonction de la composition et de l'épaisseur de l'atmosphère traversée par les rayons lumineux au cours de la journée. Il se décompose en rayonnement direct mesuré à l'aide d'un appareil de mesure appelé pyrhéliomètre, et rayonnement diffus mesuré à l'aide d'un autre appareil appelé pyranomètre qui peut aussi mesurer le rayonnement global. Ce sont ces deux rayonnements qui sont exploités par les dispositifs solaires.

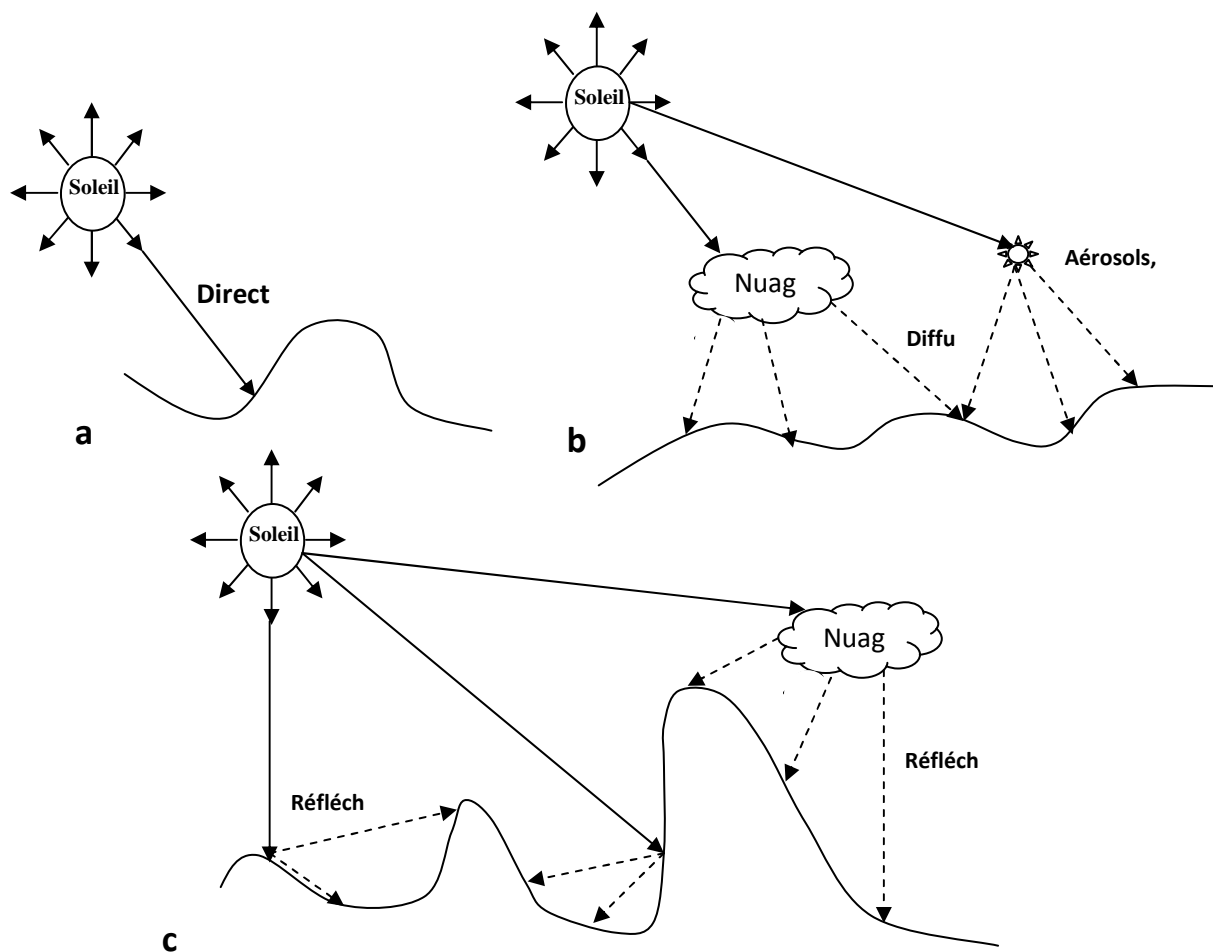


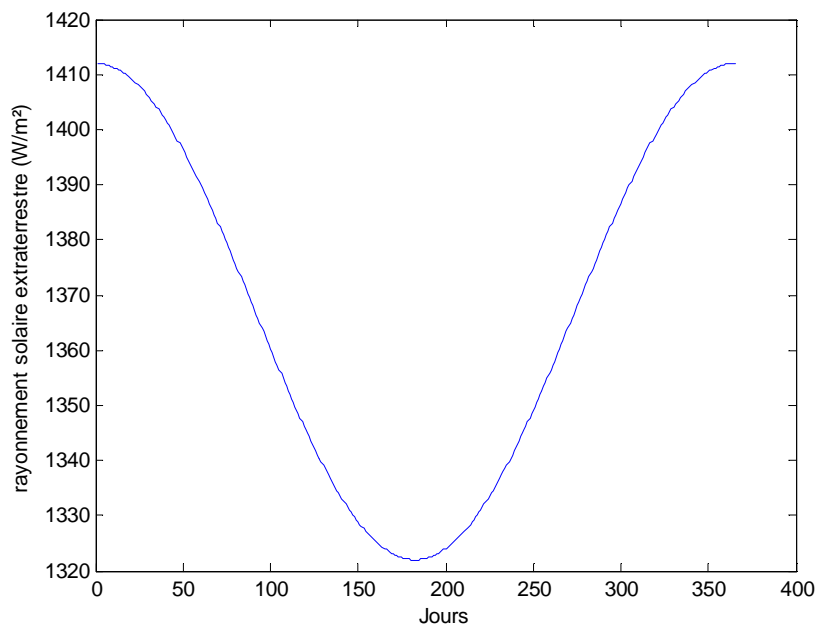
Fig. 3 : Parcours du rayonnement solaire. a Direct. b Diffus. c Réfléchis

### III.3 La constante solaire : [26]

La constante solaire est la densité de puissance du rayonnement sur l'atmosphère externe de la terre et exprimée en watts par mètre carré par minute ( $W/m^2$ ).

Le Centre Mondiale du Rayonnement a adopté une valeur de  $1367 \text{ W / m}^2$  avec une incertitude de 1%. La constante solaire la plus mise à jour est  $I_0 = 1367 \text{ W / m}^2$ , qui est équivalent à  $I_0 = 1.960 \text{ calories / cm}^2$  de minute ou  $432 \text{ Btu / ft}^2\text{h}$  ou  $4.921 \text{ MJ / m}^2\text{h}$ .

Iqbal (1986) fournit un calcul plus détaillée sur la constante solaire. Comme la distance entre le soleil et la terre change durant l'année, la valeur de la constante solaire change également pendant l'année fig. 4.



**Fig. 4:** la variation de l'irradiation solaire extraterrestre

La meilleure valeur de la constante solaire utilisée est actuellement  $I_0 = 1360 \text{ W / m}^2$  (Frochlich et Werhli 1981).[26]

### III.4 L'irradiation solaire hors atmosphère :

L'irradiation solaire est le taux d'énergie incident sur une unité de surface. Cette énergie est calculée par l'intégration du rayonnement reçu sur une unité de surface dans un temps indiqué soit l'heure ou le jour. Exprimée en  $\text{Wh / m}^2$ .

#### III.4.1 Calcul de l'irradiation solaire globale hors atmosphère :

Soit  $G_0$  l'irradiation globale hors atmosphère reçu sur un plan horizontale par rapport au site donné.

$$G_0 (Wh/m^2) = \left( \frac{24I_0}{\pi} \right) \left( 1 + 0.033 \cos \left( \frac{360N}{365} \right) \right) \cdot \left( \cos \lambda \cos \delta \sin \omega + 2\pi \omega \sin \lambda \sin \frac{\delta}{360} \right) \quad (10)$$

Tel que ;

$I_0$  : est la constante solaire,  $I_0 = 1367(W/m^2)$

$N$  : est le numéro du jour à partir du 1<sup>er</sup> Janvier (on considère 365 jours dans l'année)

$\lambda$  : est la latitude du lieu

$\delta$  : est la déclinaison du soleil

$$\delta = 23.45 \left[ \sin \left( \frac{360(284 + N)}{365} \right) \right] \quad (11)$$

$\omega$  : est l'angle horaire :

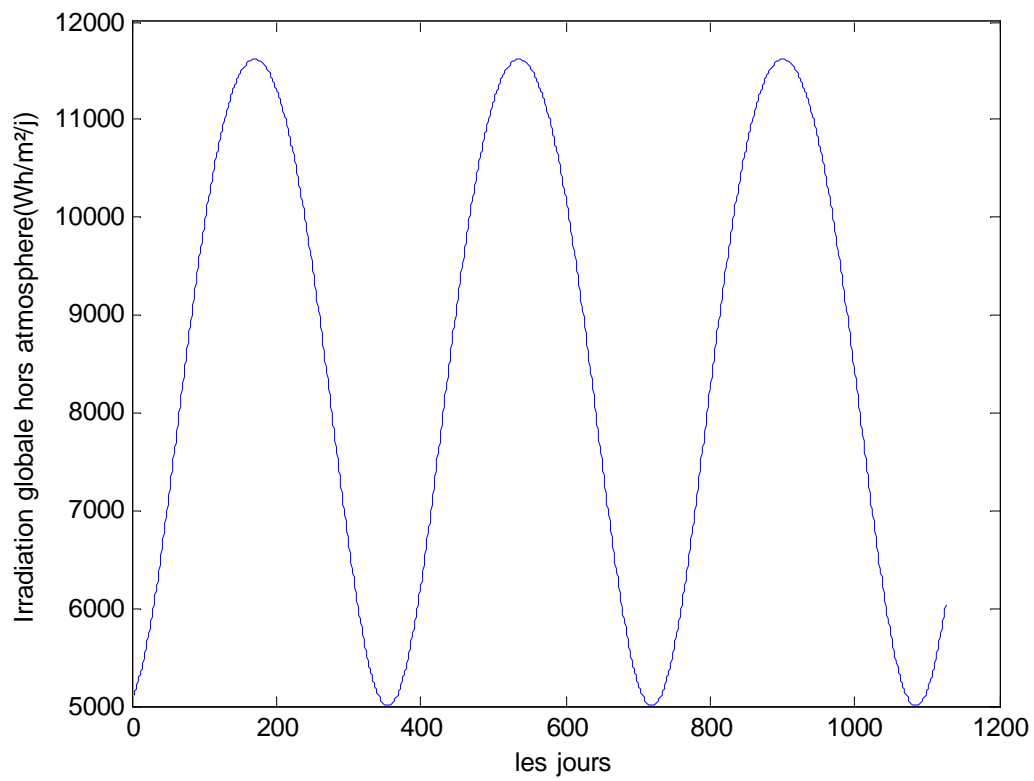
$$\omega = \cos^{-1}(-\tan \lambda \tan \delta) \quad (12)$$

NB : Tous les angles sont en radians.

#### II.4.2 Application pour le site de Ghardaïa :

La latitude du site de Ghardaïa est:  $\lambda = 32.23'N$

La fig. 5 représente les données de l'irradiation solaire globale hors atmosphère pour Ghardaïa.



**Fig.5** : l'irradiation globale hors atmosphère  $G_0$  par rapport au site de Ghardaïa

## *Chapitre.2 Analyse des séries temporelles*

---

## I. Introduction

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps. Elles ont été utilisées en astronomie (*'on the periodicity of sunspots', 1906*), en météorologie (*'time-series regression of sea level on weather', 1968*), en théorie du signal (*'Noise in FM receivers', 1963*), en biologie (*'the autocorrelation curves of schizophrenic brain waves and the power spectrum', 1960*), en économie (*'time-series analysis of imports, exports and other economic variables', 1971*)...etc. [27]

Pour l'analyse des séries temporelles, une approche est utilisée et qui est basé sur l'utilisation des corrélations qui consiste à étudier les corrélations croisées de fonctions de la série chronologique. Ces méthodes sont généralement paramétriques de type moyenne-mobiles (Moving Average MA) ou autorégressives (AR) - voire les deux (ARMA) (nous verrons en détails dans ce chapitre). Toutes ces méthodes consistent à estimer des paramètres peuvent généralement être vues comme des généralisations de la régression linéaire.

L'autre approche également utilisée est celle basée sur l'étude des 'fréquences'. Cette vision est une généralisation des méthodes utilisées en analyse de Fourier. L'idée est ici d'approximer une fonction analytique par une somme pondérée de fonctions sinus ou cosinus.

### I.1 Histoire des séries temporelles

Historiquement, les astronomes sont les premiers qui ont travaillé sur des séries chronologiques (Kepler a pu énoncer ses lois sur le mouvement des planètes). Par exemple, les grecs ou les hébreux ont observés des phénomènes cycliques (*D'après Against the Gods, the remarkable story of risk*), mais ils n'ont jamais pensé à faire de la prévision, car à cet époque, prévoir des phénomène futurs, pouvait faire croire à une tentative de rivaliser avec les dieux : Halley découdra le cycle d'une même comète d'où son nom « comète de Halley » qui fut aperçue en 1531, en 1607 et en 1682 (cette comète avait été observée d'ailleurs depuis 240 avant J.C.), et il prévoit qu'on la reverra en 1758 (et ce fut effectivement le cas, tous les 76 ans).

C'est grâce aux différents outils mathématiques développés au XV<sup>III</sup>e et XIX<sup>e</sup> siècles qui ont permit de mettre en place les premières techniques d'étude des séries chronologiques, parmi lesquelles, l'analyse harmonique. Et leur but était de détecter des saisonnalités cachées au sein de leurs données. Ainsi, Lagrange a utilisé ces méthodes pour détecter la périodicité

cachée en 1772 et en 1778. Un demi-siècle plus tard, en 1847, Buys et Ballot, dans 'Les changements périodiques de la températures ont proposé des méthodes pour étudier la périodicité de données astronomiques. Toutefois, il a fallu attendre 1889 pour que Sir Arthur Shuster introduise le périodogramme, qui constitue la base des méthodes spectrales d'analyse de séries chronologiques. Wolf a considéré la série chronologique du nombre de taches solaires et via le periodogramme a pu montrer qu'il y'avait une périodicité de 11 années dans le cycle solaire. Aussi en 1924, Whittaker et Robinson ont aussi utilisé cette théorie sur la brillance de l'étoile TUrsa Major et ont pu modéliser ce phénomène de la brillance par deux fonctions dont les périodes sont respectivement 24 et 29. [27]

## I.2 Développement de l'analyse des séries temporelles :

L'étude des séries temporelles semble avoir atteint sa maturité au cours des années 70 où des développements significatifs sont apparus.

En 1965, Cooley et Tukey ont bien contribué dans l'étude spectrale des séries en publiant leur article '*an algorithm for the machine calculation of complex Fourier series*', introduisant la Transformée de Fourier Rapide ( Fast Fourier Transform FFT), qui est un algorithme permettant de calculer rapidement des périodogrammes.

A la même époque, en 1970, Box et Jenkins ont publié leur ouvrage '*Time series analysis, forecasting and control* ', montrant que l'étude des séries temporelles à l'aide de processus de type ARMA ( ARIMA, AR, MA, ARCH, GARCH, ARMAX,...) pouvait s'appliquer à de nombreux domaines tel que : la météorologie, l'astronomie, l'économétrie,..., et pouvait être facilement implémentée informatiquement. C'est cette dernière méthode que nous adopterons pour la réalisation de notre travail. Mais Avant d'introduire la notion de processus temporel et plus particulièrement la classe des processus ARMA (ARIMA , SARIMA) qui sont particulièrement utiles pour décrire le comportement des séries temporelles univariées, il est important que l'on fasse au préalable un certain nombre de rappels en probabilité et en statistiques.

## II. Rappels de Probabilité et de Statistiques :[28]

### II.1 Introduction :

Soit  $X$  une variable stochastique réelle ou variable aléatoire réelle (v.a.r. en abrégé) Continue. Pour une réalisation  $x$ , la loi de probabilité, est définie par la fonction de densité  $f_X(x)$  , supposée continue, telle que  $f_X(x) \geq 0$  :

$$\forall (a, b) \in \mathbb{R}^2 \quad P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (13)$$

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \quad (14)$$

Tel que  $F$  est la fonction de répartition ou fonction de distribution cumulative associée à  $X$  :

$$F(a) \equiv P(X \leq a) \equiv P(X < a) = \int_{-\infty}^a f_X(x) dx \quad (15)$$

Quelques propriétés de la fonction de répartition  $f_X(x)$  :

- $F$  est positive non décroissante. Si  $x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$
- $0 \leq F(x) \leq 1$
- $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow \infty} F(x) = 1$
- Cas continu :  $F(x) = \int_{-\infty}^x f_X(x) dx$
- Cas discret :  $F(x) = \sum_{x_i \leq x} Pr(X = x_i) = \sum_{x_i \leq x} p_i$
- $f_X$  est continue à droite :  $\lim_{|\varepsilon| \rightarrow 0} F(x + |\varepsilon|) = f_X(x)$

$$\lim_{|\varepsilon| \rightarrow 0} F(x - |\varepsilon|) = F(x) - P(X = x)$$

- La fonction de répartition est constante sur un intervalle  $[x_i, x_{i+1}[$  . et c'est une fonction en escalier.

## II.2 La population des moments théoriques

**Définition :** Pour une variable aléatoire réelle continue  $X$ , de densité  $f_X(x)$ ;

La population des moments associée à tout ordre  $k \in \mathbb{N}$ , est définie comme suit :

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) dx \quad (16)$$

Et la population des moments centrés associée à tout ordre  $k \in \mathbb{N}$ , notée  $\mu_k$ , est définie par :

$$\mu_k = E[(X - \mu)^k] = \int_{-\infty}^{\infty} (X - \mu)^k f_X(x) dx \quad (17)$$



**Propriété 1:** On considère  $Y$  une v.a.r. transformée, telle que  $Y = g(X)$ . Soit  $f_X(\cdot)$  la fonction de densité de la v.a.r.  $X$ . La population des moments et la population des moments centrés d'ordre  $k$ , de la transformée  $g(X)$  sont alors définies par l'espérance respective des transformées :

$[g(X)]^k$  et  $[g(X) - \mu^Y]^k$ , avec  $\mu^Y = E[g(x)], \forall k \in \mathbb{N}$  :

$$E\{[g(X)]^k\} = \int_{-\infty}^{\infty} [g(X)]^k f_X(x) dx \quad (18)$$

$$\mu_k^Y = E\{[g(X) - \mu^Y]^k\} = \int_{-\infty}^{\infty} [g(X) - \mu^Y]^k f_X(x) dx \quad (19)$$

### II.2.1 l'espérance mathématique:

L'espérance mathématique d'une variable aléatoire  $X$ , notée  $E(X)$ , correspond au moment d'ordre un ( $k = 1$ ) et est l'équivalent en probabilité à la moyenne d'une série statistique en statistique.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \mu \quad (20)$$

#### *Quelque propriété de l'espérance mathématique:*

- Constante : L'espérance d'une variable aléatoire constante est égale à cette constante; par exemple, si  $b$  est une constante, alors  $E(b) = b$ .
- Monotonie : Si  $X$  et  $Y$  sont des variables aléatoires tels que  $X \leq Y$ , alors  $E(X) \leq E(Y)$
- Linéarité :  $E(X + Y) = E(X) + E(Y)$

$$E(aX) = aE(X)$$

On utilise souvent comme estimateur de l'espérance la moyenne empirique, qui est un estimateur:

- Sans biais
- Convergent selon la loi des grands nombres et même fortement convergent selon la loi forte des grands nombres
- Distribué normalement asymptotiquement selon le théorème central limite

### II.2.2 La variance :

Pour ( $k = 2$ ), on définit le moment centré d'ordre deux et qui représente la variance notée  $V(X)$  ;

$$V(X) = \mu_2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (X - \mu)^2 f_X(x) dx \quad (21)$$

Notons que la racine carrée de la variance est appelé "Ecart-type "

Résultat découlant de la propriété 1 :

$$E(X^2) = V(X) + [E(X)]^2 \quad (22)$$

### II.2.3 Le Kurtosis :

Le Kurtosis est une mesure de "l'épaisseur" des queues de distributions.

$$Kurtosis = \mu_4 = E[(X - \mu)^4] \quad (23)$$

En règle générale, on exprime ces deux mesures en contrôlant par une fonction puissance de la variance ( $X = \sigma^2$ ). On définit ainsi une nouvelle mesure :

$$\text{Le degré d'excès de Kurtosis} = \frac{\mu_4}{\sigma^4} - 3 \quad (24)$$

### II.2.4 Le Skewness :

La Skewness est une mesure de l'asymétrie de la distribution.

Pour des fonctions de distributions symétriques, telles que :

$$f_X(\mu - x) = f_X(\mu + x) \Rightarrow \text{La valeur de la Skewness est nulle, } S_u = 0.$$

En revanche, pour des fonctions de distributions asymétriques, la valeur de la Skewness est positive, si la partie épaisse de la distribution se situe dans la direction positive.

### III. Processus aléatoire stationnaire

#### III.1 La stationnarité au sens stricte:

On dit qu'un processus aléatoire temporel  $(X_t, t \in T)$  est complètement stationnaire si, pour toute suite d'instants  $\{t_1, t_1, \dots, t_N\}$ , il existe un entier  $k$  quelconque tel que la f.d.p (la fonction de distribution de probabilités) jointe  $f$  de  $\{X_{t_1}, X_{t_1}, \dots, X_{t_N}\}$  est identique à la f.d.p. jointe de  $\{X_{t_1+k}, X_{t_1+k}, \dots, X_{t_N+k}\}$

$$f(X_{t_1}, X_{t_1}, \dots, X_{t_N}) = f(X_{t_1+k}, X_{t_1+k}, \dots, X_{t_N+k}) \quad (25)$$

Cette définition signifie que déplacer l'origine du temps de  $k$  périodes n'a aucune incidence sur la distribution de probabilités de la variable aléatoire temporelle  $X_t$ . Autrement dit ; le processus est à l'équilibre.

Le problème posé par cette définition est qu'elle est trop stricte, i.e. qu'elle est invérifiable de façon pratique. En effet, il faudrait pouvoir calculer une f.d.p. jointe pour chaque instant de la série d'observations. Or, à chaque instant on ne dispose que d'une seule donnée d'observation.

C'est pourquoi on restreint dans les travaux appliqués la notion de stationnarité à la stationnarité de second ordre encore qualifiée de stationnarité au sens large.

#### III.2 La stationnarité au sens large :

On dit qu'un processus aléatoire temporel  $(X_t, t \in T)$  est stationnaire de second ordre s'il vérifie les propriétés suivantes:

$$\begin{cases} E[X_t] = \mu & \forall t \\ V[X_t] = \sigma_x^2 & \forall t \\ cov[X_t, X_{t-k}] = \gamma_k & \forall t \end{cases} \quad (26)$$

Les deux premières propriétés signifient que la moyenne et la variance du processus sont indépendantes de l'instant auquel ils sont mesurés. La troisième propriété signifie que la covariance ne dépend que du seul retard  $k$ .

### III.3 Le Théorème de Wold :

Le théorème de Wold (1938) est le théorème fondamental de l'analyse des séries temporelles Stationnaires.

#### III.3.1 Enoncé du théorème de la décomposition de Wold :

Tout processus stationnaire d'ordre deux  $(x_t, t \in \mathbb{Z})$  peut être représenté sous la forme :

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + k_t \quad (27)$$

Où les paramètres  $\psi_j$  satisfont :  $\psi_0 = 1$ ,  $\psi_j \in \mathbb{R}$ ,  $\forall j \in \mathbb{N}^*$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$  et où  $\varepsilon_t$  est un bruit blanc i.i.d  $(0, \sigma_\varepsilon^2)$ . On dit que la somme des chocs passés correspond à la composante linéaire stochastique de  $x_t$ . Le terme  $k_t$  désigne la composante linéaire déterministe telle que :  $cov(k_t, \varepsilon_{t-j}) = 0, \forall j \in \mathbb{Z}$ .

Ainsi, d'après le théorème de Wold, si l'on omet la composante déterministe  $k_t$ , tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés, ces chocs étant représentés par un bruit blanc de variance finie. L'implication forte de ce théorème est que, si l'on connaît les pondérations  $\psi_j$ ,  $\forall j \in \mathbb{N}$ , et si l'on connaît la variance  $\sigma_\varepsilon^2$  du bruit blanc, on est en mesure de proposer une représentation de n'importe quel processus stationnaire. Cette représentation est aussi qualifiée de *représentation moyenne mobile infinie MA( $\infty$ )*.

### III.4 Définition de l'opérateur retard :

L'opérateur retard (noté "L" pour lag ou "B" suivant les ouvrages) est défini de la façon suivante:

On considère un processus stochastique  $(x_t, t \in \mathbb{Z})$ , l'opérateur retard noté L, est défini par la relation :

$$Lx_t = x_{t-1} \quad \forall t \in \mathbb{Z} \quad (28)$$

Cet opérateur permet ainsi de définir une application qui à toute variable  $x_t$  associe la variable retardée  $x_{t-1}$ . Cet opérateur possède les propriétés suivantes :

- ✓ **Propriété 1** :  $L^j x_t = x_{t-j}$ ,  $\forall j \in \mathbb{Z}$ , en particulier on a  $L^0 x_t = x_t$
- ✓ **Propriété 2** : Si  $x_t = c$ ,  $\forall t \in \mathbb{Z}$  avec  $c \in \mathbb{R}$ ,  $L^j x_t = L^j c = c$ ,  $\forall j \in \mathbb{Z}$
- ✓ **Propriété 3** :  $L^i(L^j x_t) = L^{i+j} x_t = x_{t-i-j}$   $\forall (i, j) \in \mathbb{Z}^2$

La preuve est la suivante :

$$L^i(L^j x_t) = L^i(x_{t-j}) = x_{t-i-j}$$

- ✓ **Propriété 4** :  $L^{-i} x_t = x_{t+i}$ ,  $\forall i \in \mathbb{Z}$
- ✓ **Propriété 5** :  $(L^i + L^j)x_t = L^i x_t + L^j x_t = x_{t-i} + x_{t-j}$ ,  $\forall (i, j) \in \mathbb{Z}^2$
- ✓ **Propriété 6** : Si  $|a| < 1$

$$(1 - aL)^{-1} x_t = \frac{x_t}{(1 - aL)} = \lim_{j \rightarrow \infty} (1 + aL + a^2 L^2 + \dots + a^j L^j) x_t$$

Cette dernière propriété est particulièrement utile pour inverser des polynômes d'ordre 1 définis en l'opérateur retard.

### III.5 Théorème de Wold et polynômes retard

Partant de la définition de cet opérateur retard, on peut alors construire des polynômes à coefficients réels en cet opérateur. Par exemple, considérons un polynôme  $\psi(L)$  d'ordre  $q$  :

$$\psi(L) = \psi_0 L^0 + \psi_1 L^1 + \psi_2 L^2 + \dots + \psi_q L^q \quad (29)$$

L'application de ce polynôme à un processus aléatoire  $(\varepsilon_t, t \in \mathbb{Z})$  nous donne donc :

$$\begin{aligned} \psi(L)\varepsilon_t &= \psi_0 L^0 \varepsilon_t + \psi_1 L^1 \varepsilon_t + \psi_2 L^2 \varepsilon_t + \dots + \psi_q L^q \varepsilon_t \\ &= \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_q \varepsilon_{t-q} \end{aligned} \quad (30)$$

Si  $\psi_0 = 1$  et  $q \rightarrow \infty$  on retrouve ici une forme moyenne mobile infinie identique à celle du théorème de Wold. Le nouvel énoncé de ce théorème (identique, bien entendu, au précédent) est le suivant :

#### **Théorème (Décomposition de Wold) :**

Tout processus stationnaire d'ordre deux  $(x_t, t \in \mathbb{Z})$  peut être représenté sous la forme :

$$x_t = \psi(L)\varepsilon_t + k_t \quad (31)$$

Où le polynôme de degré infini  $\psi(L)$  est défini par  $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ , les paramètres  $\psi_j$  satisfaisant  $\psi_0 = 1$ ,  $\psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^*$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$  et où  $\varepsilon_t$  est un bruit blanc i.i.d  $(0, \sigma_\varepsilon^2)$  et où la composante linéaire déterministe  $k_t$  vérifie :  $cov(k_t, \varepsilon_{t-j}) = 0, \forall j \in \mathbb{Z}$ .

### III.6 Les processus ARMA

#### III.6.1 Introduction :

Jusqu'à présent nous avons vu que tout processus stationnaire pouvait s'écrire sous la forme d'une somme pondérée infinie de chocs passés (théorème de Wold). Pour toute cette classe de processus la décomposition de Wold est donc une première représentation possible. Toutefois, cette représentation n'est jamais optimale parmi toutes les représentations possibles d'un même processus. En effet, lorsque que l'on cherche à modéliser une série temporelle, on applique toujours un principe de parcimonie qui veut que, à qualité d'ajustement égale, l'on adopte en priorité la représentation nécessitant l'estimation du minimum de paramètres. Or par définition, si l'on devait appliquer la décomposition de Wold, cela supposerait que l'on estime une infinité de paramètres (les  $\psi_j$  et le  $\sigma_\varepsilon^2$ ). Donc dans la pratique, il convient de rechercher d'autres représentations possibles pour les processus temporels. Parmi les représentations les plus utilisées figurent les représentations ARMA pour AutoRegressive Moving Average. Cette représentation consiste en l'adjonction d'une composante autorégressive d'ordre fini (AR) et d'une composante moyenne mobile d'ordre fini (MA).

Nous allons donc commencer par définir la classe des processus AR, MA et ARMA et par la suite nous introduirons des processus ARMA saisonniers : les SARMA .

#### III.6.2 Le processus AR(p) :

On dit que le processus aléatoire temporel  $(X_t, t \in T)$  est de type AR(p) i.e. autorégressif d'ordre p si l'on peut écrire:

$$\left\{ \begin{array}{l} X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t \\ \text{où} \\ (\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p) \in \mathbb{R}^{p+1} \\ E[\varepsilon_t] = 0 \quad ; \quad V[\varepsilon_t] = \sigma_\varepsilon^2 \quad ; \quad \text{cov}[\varepsilon_t, \varepsilon_s] = 0 \quad \forall t \neq s \end{array} \right. \quad (32)$$

Supposons que  $\varphi_0 = 0$ . On peut réécrire le processus AR(p) à l'aide du polynôme retard :

Tel que :

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p} = \varepsilon_t = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) X_t \quad (33)$$

Et le polynôme retard a pour expression :

$$\varphi(L) = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) \quad (34)$$

Dont l'équation caractéristique associée s'écrit :

$$1 - \varphi_1 Z - \varphi_2 Z^2 - \dots - \varphi_p Z^p = 0 \quad (35)$$

Le polynôme retard est inversible si les racines de l'équation caractéristique sont toutes de module supérieur à l'unité. Dans ce cas, on peut écrire:

$$X_t = \varphi^{-1}(L) \varepsilon_t \quad (36)$$

A partir de la On peut en déduire les deux premiers moments du modèle AR(p)

### III.6.2.1 La moyenne du processus AR(p) :

$$E(X_t) = E[\varphi^{-1}(L) \varepsilon_t] = 0 \quad \text{car} \quad E[\varepsilon_t] = 0 \quad \forall t \quad (37)$$

### III.6.2.2 La variance du processus AR(p) :

Pour le calcul de la variance, on multiplie chaque terme du processus par  $X_t$

$$X_t * X_t = \varphi_1 X_{t-1} * X_t + \varphi_2 X_{t-2} * X_t + \dots + \varphi_p X_{t-p} * X_t + \varepsilon_t * X_t \quad (38)$$

Puis on procède au calcul de l'espérance de chacun des termes de l'égalité :

$$\begin{aligned} E[X_t * X_t] &= E[\varphi_1 X_{t-1} * X_t] + E[\varphi_2 X_{t-2} * X_t] + \dots \\ &\dots + E[\varphi_p X_{t-p} * X_t] + E[\varepsilon_t * X_t] \end{aligned} \quad (39)$$

Puisque  $E[\varepsilon_t * X_t] = E[\varepsilon_t * (\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t)] = \sigma_\varepsilon^2$ , l'expression s'écrit alors :

$$V[X_t] = \varphi_1 cov[X_{t-1}, X_t] + \varphi_2 cov[X_{t-2}, X_t] + \dots + \varphi_p cov[X_{t-p}, X_t] + \sigma_\varepsilon^2 \quad (40)$$

Soit par convention d'écriture :

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \dots + \varphi_p \gamma_p + \sigma_\varepsilon^2 \quad (41)$$

On procède d'une manière similaire pour calculer la covariance à l'ordre  $k$  ( $k \neq 0$ ), à savoir que l'on multiplie chaque terme de l'égalité (32) par  $X_{t-k}$

$$X_t * X_{t-k} = \varphi_1 X_{t-1} * X_{t-k} + \varphi_2 X_{t-2} * X_{t-k} + \dots + \varphi_p X_{t-p} * X_{t-k} + \varepsilon_t * X_{t-k} \quad (42)$$

Puis on reprend l'espérance de chaque terme de l'égalité,

$$\begin{aligned} E[X_t * X_{t-k}] &= E[\varphi_1 X_{t-1} * X_{t-k}] + E[\varphi_2 X_{t-2} * X_{t-k}] + \dots \\ &\dots + E[\varphi_p X_{t-p} * X_{t-k}] + E[\varepsilon_t * X_{t-k}] \end{aligned} \quad (43)$$

Ceci nous ramène à l'expression de la covariance :



$$\begin{aligned} cov[X_t, X_{t-k}] &= \varphi_1 cov[X_{t-1}, X_{t-k}] + \varphi_2 cov[X_{t-2}, X_{t-k}] + \dots \\ &\dots + \varphi_p cov[X_{t-p}, X_{t-k}] \end{aligned} \quad (44)$$

Par convention d'écriture :

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \dots + \varphi_p \gamma_{k-p} \quad (45)$$

L'écriture de l'ACF (fonction d'autocorrélation) est alors immédiate, puisqu'il suffit de diviser chaque terme de la covariance par la variance du processus:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad (46)$$

### La fonction d'autocorrélation partielle FACP:

l'autocorrélation partielle d'ordre  $k$  désigne la corrélation entre  $X_t$  et  $X_{t-k}$  obtenue lorsque l'influence des variables  $X_{t-k-i}$  avec  $i < k$ , a été retirée. FACP est un moyen utile pour déterminer l'ordre  $p$  dans un modèle autorégressif AR.[30]

L'évaluation des coefficients  $\varphi_{ii}$  de FACP se fait comme suit :

- ✓ On estime d'abord le AR(1) suivant par la méthode des moindres carrés ordinaires MCO :

$$X_t = \varphi_{11} X_{t-1} + \varepsilon_t \quad (47)$$

Tel que  $\varphi_{11}$  est le coefficient d'ACP d'ordre 1

- ✓ Puis, on estime le AR(2) suivant :

$$X_t = \varphi_{21} X_{t-1} + \varphi_{22} X_{t-2} + \varepsilon_t \quad (48)$$

On obtient le  $\varphi_{22}$  qui est le coefficient d'ACP d'ordre 2

Et ainsi de suite pour générer les coefficients d'ACP des autres retards.

### Remarque :

On peut caractériser le comportement d'un processus AR( $p$ ) à travers ses fonctions ACF et PACF : tel que la fonction ACF décroît vers 0 rapidement si  $\varphi_1$  est proche de 0 ou

lentement si  $\varphi_1$  est proche de 1. Son allure est de type exponentielle si  $\varphi_1 > 0$  ou sinusoïdale si  $\varphi_1 < 0$ . La fonction PACF (fonction d'autocorrélation partielle) s'annule pour tout retard  $k > p$ .

### III.6.3 Le processus MA

On dit que le processus aléatoire temporel  $(X_t, t \in T)$  est de type MA(q) i.e. Moving Average (moyenne mobile) d'ordre q si l'on peut écrire:

$$\begin{cases} X_t = \theta_0 + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \\ \text{où} \\ (\theta_0, \theta_1, \theta_2, \dots, \theta_q) \in \mathbb{R}^{q+1} \\ E[\varepsilon_t] = 0 \quad \forall t ; \quad V[\varepsilon_t] = \sigma_\varepsilon^2 \quad ; \quad cov[\varepsilon_t, \varepsilon_s] = 0 \quad \forall t \neq s \end{cases} \quad (49)$$

La condition d'inversibilité du MA(q) consiste à écrire que les racines de l'équation caractéristique  $1 + \theta_1Z + \dots + \theta_qZ^q = 0$  sont en dehors du cercle unité.

#### III.6.3.1 La moyenne du processus MA(q) :

$$\begin{aligned} E[X_t] &= E[\theta_0 + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}] = \theta_0 \\ \text{car } E[\varepsilon_t] &= 0 \quad \forall t \end{aligned} \quad (50)$$

#### III.6.3.2 La variance du processus MA(q) :

$$\begin{aligned} V[X_t] &= E[(X_t - E[X_t])^2] \\ &= E[(X_t - \theta_0)^2] \\ &= E\left[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q})^2\right] \\ &= E[\varepsilon_t^2] + \theta_1^2 E[\varepsilon_{t-1}^2] + \dots + \theta_q^2 E[\varepsilon_{t-q}^2] \quad \text{car } cov[\varepsilon_t, \varepsilon_s] = 0 \quad \forall t \neq s \\ &= (1 + \theta_1^2 + \dots + \theta_q^2)\sigma_\varepsilon^2 \end{aligned} \quad (51)$$



En supposant  $\vartheta_0 = 0$  ; on pourra écrire le processus ARMA en fonction des polynômes retards :

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) X_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \quad (56)$$

$$\text{Ce qui revient à écrire : } \varphi(L) X_t = \theta(L) \varepsilon_t \quad (57)$$

On peut noter que le processus est stationnaire et inversible si toutes les racines des polynômes respectivement  $\varphi$  et  $\theta$  sont à l'extérieur du cercle unité.

Par la suite, on suppose ces deux conditions vérifiées. On procède au calcul des deux premiers moments du processus ARMA :

#### III.6.4.1 La moyenne du processus ARMA :

$$E[X_t] = E[\varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}] = 0$$

Car

$$E[\varepsilon_t] = 0 \quad \forall t \quad (58)$$

#### III.6.4.2 La variance du processus ARMA :

La variance n'a pas d'expression simple, aussi on ne propose que l'expression de la covariance à l'ordre  $k$ . On commence par multiplier chaque terme par  $X_{t-k}$  :

$$\begin{aligned} X_t * X_{t-k} &= \varphi_1 X_{t-1} * X_{t-k} + \varphi_2 X_{t-2} * X_{t-k} + \dots + \varphi_p X_{t-p} * X_{t-k} + \varepsilon_t * X_{t-k} \\ &\quad + \theta_1 \varepsilon_{t-1} * X_{t-k} + \theta_2 \varepsilon_{t-2} * X_{t-k} + \dots + \theta_q \varepsilon_{t-q} * X_{t-k} \end{aligned} \quad (59)$$

En procédant au calcul de l'espérance de chaque terme de l'égalité, on obtient après simplification l'expression de la covariance suivante :

$$\gamma_k = \varphi_1 \gamma_{k-1} + \dots + \varphi_p \gamma_{k-p} \quad \text{pour } k > q \quad (60)$$

On déduit le coefficient d'autocorrélation :

$$\rho_k = \varphi_1 \rho_{k-1} + \dots + \varphi_p \rho_{k-p} \quad \text{pour } k > q \quad (61)$$

**Remarque :**

Il apparaît que les fonctions ACF et PACF d'un processus ARMA (p,q) combinent des caractéristiques propres aux processus AR et MA.

La fonction ACF décroît vers 0 rapidement et se comporte comme l'ACF d'un processus AR(p) après  $q - p$  retards. Son allure est de type exponentiel ou sinusoïdal selon la valeur des coefficients. La fonction PACF décroît vers 0 rapidement et se comporte comme la PACF d'un processus MA(q) après  $p - q$  retards. Son allure est de type exponentiel ou sinusoïdal selon la valeur des coefficients. [27]

**IV Processus aléatoire non stationnaire :****IV.1 Introduction :**

Précédemment, nous avons introduit la notion de stationnarité du second ordre ou stationnarité faible. C'est-à-dire que l'ensemble de ses moments d'ordre un et d'ordre deux sont indépendants du temps. Par opposition, un processus non stationnaire est un processus qui ne satisfait pas l'une ou l'autre de ces deux conditions.

Ainsi, l'origine de la non stationnarité peut provenir d'une dépendance du moment d'ordre un (l'espérance) par rapport au temps et/ou d'une dépendance de la variance ou des autocovariances par rapport au temps.

Par conséquent avant de passer à la modélisation d'un tel processus, il va falloir avant toute chose le "stationnariser", c'est à dire trouver une transformation stationnaire de ce processus. Puis, on modélise et l'on estime les paramètres associés à la composante stationnaire.

Il existe différentes sources de non stationnarité et qu'à chaque origine de la non stationnarité est associée une méthode propre de stationnarisation. Nous allons présenter deux classes de processus non stationnaires, selon la terminologie de Nelson et Plosser (1982) : les processus TS (Time Stationary) et les processus DS (Differency Stationary). [29]

**IV.2 Le processus TS**

On dit que  $(x_t, t \in \mathbb{Z})$  est un processus TS ou qu'il est caractérisé par une non stationnarité déterministe s'il peut s'écrire sous la forme :

$$X_t = f(t) + Z_t \quad (62)$$

Où  $f(t)$  est une fonction du temps et  $Z_t$  est un processus stochastique stationnaire.

Dans ce cas, le processus  $X_t$  s'écrit comme la somme d'une fonction déterministe du temps et d'une composante stochastique stationnaire de type ARMA. Dès lors, il est évident que le processus ne satisfait plus la définition de la stationnarité du second ordre.

Car :

$E(X_t) = f(t) + Z$ , ( $Z = E(Z_t)$ ) Dépend du temps, ce qui viole l'une des conditions de la définition d'un processus stationnaire.

Le processus peut être stationnarisé en retranchant à  $x_t$  la valeur estimée de  $f(t)$  par la méthode des moindres carrés ordinaires MCO.

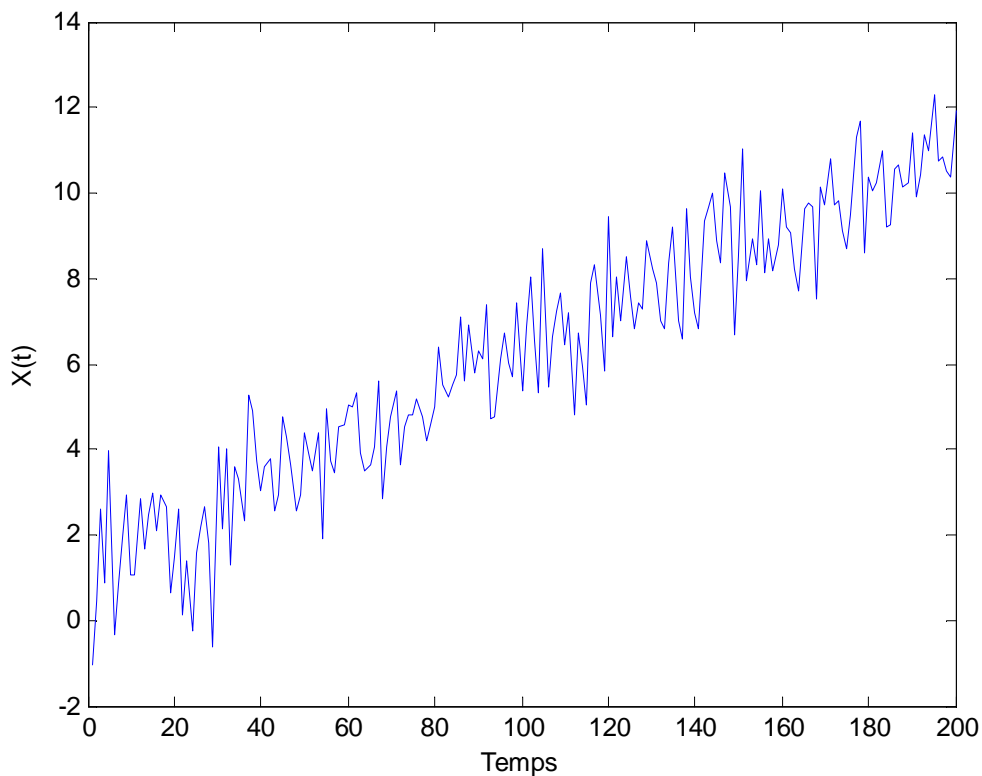
- **Exemple de processus TS:**

Soit le processus suivant ( $x_t, t \in \mathbb{Z}$ ) :

$$X_t = 1 + 0.05t + \varepsilon_t \quad (63)$$

Avec  $\varepsilon_t \sim N(0,1)$  un bruit blanc.

Le processus  $x_t$  correspond à la somme d'une fonction linéaire du temps,  $f(t) = 1 + 0.05t$  et d'un bruit blanc  $\varepsilon_t$ .



**Fig.6** : Processus Non Stationnaire : Trend Déterministe

En calculant la moyenne de  $x_t$  :

$$E(x_t) = 1 + 0.05t \quad \text{Car} \quad E(\varepsilon_t) = 0$$

On voit bien que l'une des conditions de stationnarité (le moment d'ordre un) n'est pas vérifiée ; la moyenne du processus croît avec le temps.

### IV.3 Le processus DS

On dit qu'un processus  $X_t$  est DS (differency stationary) ou caractérisé par une non stationnarité stochastique si le processus différencié une fois  $(1 - L)X_t$  est stationnaire. On parle aussi de processus intégré d'ordre un, qu'on notera :  $X_t \sim I(1)$

De manière générale, on dit que le processus  $X_t$  est un processus intègre d'ordre  $d$ , avec  $d$  le degré d'intégration, si le processus différencié  $d$  fois  $(1 - L)^d X_t$  est stationnaire. On note  $X_t \sim I(d)$ .

Nous avons deux types de processus DS:

➤ **Le processus DS avec dérive  $\beta$**  tel que :

$$X_t = X_{t-1} + \beta + \varepsilon_t \quad (64)$$

Qui représente une non stationnarité de nature stochastique qu'on peut démontrer par récurrence :

$$X_1 = X_0 + \beta + \varepsilon_1$$

$$X_2 = X_1 + \beta + \varepsilon_2 = X_0 + \beta + \varepsilon_1 + \beta + \varepsilon_2 = X_0 + 2\beta + \varepsilon_1 + \varepsilon_2$$

⋮  
⋮

$$X_t = X_0 + \beta t + \sum_{i=1}^t \varepsilon_i \quad (65)$$

$$\text{Ou } \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

En calculant le moment d'ordre un du processus DS avec dérive, on trouve :

$$E[X_t] = X_0 + \beta t \quad (66)$$

Qui dépend du temps ; plus  $t \rightarrow \infty$  et plus  $E[X_t] \rightarrow \infty$

Ce qui prouve que le processus DS est un processus non stationnaire.

- **Le processus DS sans dérive ( $\beta = 0$ )** qui est aussi appelé marche aléatoire ou marche au hasard qui s'écrit comme suit:

$$X_t = X_{t-1} + \varepsilon_t \quad (67)$$

Que nous montrerons sa non stationnarité par récurrence :

$$X_1 = X_0 + \varepsilon_1$$

$$X_2 = X_1 + \varepsilon_2 = X_0 + \varepsilon_1 + \varepsilon_2$$

⋮  
⋮

$$X_t = X_0 + \sum_{i=1}^t \varepsilon_i \quad (68)$$

Ou  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

En calculant sa variance, on obtient :

$$Var(X_t) = Var\left(\sum_{i=1}^t \varepsilon_i\right) = \sum_{i=1}^t Var(\varepsilon_i) = \sum_{i=1}^t \sigma_\varepsilon^2 = t \sigma_\varepsilon^2 \quad (69)$$

Nous constatons que la variance du processus DS sans dérive dépend du temps ; plus  $t \rightarrow \infty$  et plus  $Var[X_t] \rightarrow \infty$

Afin de stationnariser le processus DS, il suffit d'appliquer une différence première comme suit

$$X_t - X_{t-1} = \beta + \varepsilon_t \quad (\text{avec dérive}) \quad (70)$$

$$X_t - X_{t-1} = \varepsilon_t \quad (\text{sans dérive}) \quad (71)$$

#### IV.4 Détection de la non stationnarité et sa nature :

La représentation graphique de la série peut (dans certains cas) nous indiquer si elle n'est pas stationnaire, mais elle ne nous permet pas de discriminer entre les deux types de non stationnarité. Quand la série est non stationnaire (que ce soit DS ou TS), le coefficient



d'autocorrélation d'ordre 1 (empirique) est très élevé (proche de 1), l'autocorrélogramme simple décroît alors lentement.

Rappelons que l'autocorrélogramme d'un processus stationnaire présente une décroissance "rapide", puisque les autocorrélations simples  $\rho(h) \rightarrow 0$  quand  $h \rightarrow \infty$ .

De même, on observe un pic à l'origine dans la représentation de la densité spectrale pour un processus non stationnaire.

Ainsi, la représentation graphique, l'autocorrélogramme simple et la densité spectrale peuvent nous indiquer la présence de non stationnarité, mais ne peuvent pas nous renseigner sur son type (DS ou TS).

Les deux types de non stationnarité, DS et TS, n'ayant pas les mêmes conséquences statistiques sur la nature du processus, et nécessitant des méthodes différentes pour rendre le processus stationnaire, on peut mettre en œuvre une méthode, que l'on peut qualifier de méthode heuristique, pour se donner une idée de la nature de la non stationnarité. Il s'agit de stationnariser la série, d'une part par différenciation, d'autre part par retrait d'une composante déterministe, et d'observer, à l'aide de la représentation graphique et de l'autocorrélogramme, les résultats obtenus. Cette méthode est basée sur les conséquences de l'application d'une méthode de stationnarisation à tort.

Il existe aussi plusieurs tests de non stationnarité (racine unitaire) de la série temporelle tel que le test de **Dickey-Fuller**, **Dickey-Fuller augmenté**, **le test de Philips et Perron**, **le test de Schmidt et Shin (KPSS)**, mais nous prendrons soin de présenter en détails le test de Dickey-Fuller qui est le plus répandu et qui est inspiré des travaux de JD. Hamilton, Time series analysis, Princeton University Press, 1994.

#### **IV.5 Test de Dickey-Fuller augmenté ( test de racine unitaire) :[27]**

Dickey et Fuller [1979] sont les premiers à proposer un test de non stationnarité qui permet de tester l'hypothèse nulle de racine unitaire dans un processus AR(1), i.e. de tester si la non stationnarité qui caractérise la série est de nature stochastique. Ils testent trois alternatives de processus autorégressifs: un processus de marche au hasard sans dérive, un processus de marche au hasard avec dérive et un processus combinant un terme autorégressif et une tendance déterministe. Les deux premiers modèles expriment une non stationnarité purement stochastique, alors que le dernier est caractérisé par une non stationnarité de nature mixte. Ces différents modèles supposent que le terme d'erreur est un bruit blanc. Par la suite,

les articles de Dickey-Fuller [1981] et Said-Dickey [1984] étendent les résultats des tests en considérant que l'erreur suit un processus ARMA(p,q).

Les trois modèles proposés sont:

$$\left\{ \begin{array}{l} \text{Modele [1]} \quad \Delta X_t = \Phi X_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta X_{t-i} + \varepsilon_t \\ \text{Modele [2]} \quad \Delta X_t = \Phi X_{t-1} + a + \sum_{i=1}^{p-1} \alpha_i \Delta X_{t-i} + \varepsilon_t \\ \text{Modele [3]} \quad \Delta X_t = \Phi X_{t-1} + a + \beta t + \sum_{i=1}^{p-1} \alpha_i \Delta X_{t-i} + \varepsilon_t \end{array} \right. \quad (72)$$

Avec  $p$  le nombre de retards à ajouter dans la régression afin de prendre en compte l'autocorrélation et donc de "blanchir" les résidus. On parle de correction paramétrique de l'autocorrélation. Pour le choix de  $p$  à introduire dans la régression, on peut se baser sur l'autocorrélogramme partiel de  $(1 - L)X_t$ .

#### IV.5.1 La stratégie du test de racine unitaire :

on cherche à tester la racine unitaire sous  $H_0$  contre une racine en dehors du cercle unité sous  $H_1$ . Ceci consiste donc à tester :

$$H_0: \quad \Phi = 0$$

Contre

$$H_1: \quad \Phi < 0$$

Dans les trois modèles [1], [2] et [3].

La statistique de test (tabulée par Dickey et Fuller) pour l'estimateur de  $\Phi$  est donnée par :

$$t_{\hat{\Phi}} = \frac{\hat{\Phi}_n - 1}{\hat{\sigma}_{\Phi_n}} \quad (73)$$

Les paramètres sont estimés par la méthode des MCO.

- Si la valeur empirique de la statistique de test est inférieure à la valeur critique de la table, alors on rejette l'hypothèse nulle de non stationnarité de type stochastique.

- Si la valeur empirique de la statistique de test est supérieure à la valeur critique de la table, alors on accepte l'hypothèse nulle de non stationnarité de type stochastique i.e. qu'il y a présence d'une racine unitaire.

#### IV.6 Les processus ARIMA (p,d,q) :

Soit  $X_t$  une série non stationnaire et qui ne présente pas de saisonnalité, alors il convient de la modéliser à l'aide d'un processus ARIMA (p, d, q) (Autoregressive Integrated Moving Average) où d désigne l'ordre de différenciation (intégration).

Un processus ARIMA (p,d,q) de la série  $Y_t$  est un processus de la forme suivante:

$$\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad (74)$$

où

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

Et  $\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$ ,  $L$  est l'opérateur retard, "d" est le degré d'intégration pour que la série  $X_t$  devient stationnaire ( $d \geq 0$ ) et  $(\varphi_1, \varphi_2, \dots, \varphi_p)$  et  $(\theta_1, \theta_2, \dots, \theta_q)$  sont les coefficients à estimer.

On peut poser :  $Y_t = \nabla^d X_t = (1-L)^d X_t$ , Alors pour estimer les paramètres du processus ARIMA (p,d,q) de la série  $X_t$  non stationnaire, on estimera les coefficients du processus ARMA (p, q) de la série  $Y_t$  qui est stationnaire.

#### IV.7 Les processus SARIMA (p,d,q)(P,D,Q) :

Ces processus sont une généralisation des modèles ARIMA (p,d,q), contenant une partie saisonnière.

$$(1-L^s)\varphi(L)\Phi(L^s)(1-L)^d X_t = \theta(L)\vartheta(L^s)\varepsilon_t + \varphi_0 \quad (75)$$

où

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

$$\Phi(L^s) = 1 - \phi_1 L^s - \phi_2 L^{2s} - \dots$$

$$\vartheta(L^s) = 1 - \vartheta_1 L^s - \vartheta_2 L^{2s} - \dots$$

- $p$  est le nombre de termes autorégressifs
- $d$  est le nombre de différences
- $q$  est le nombre de termes moyens mobiles

Les polynômes  $\varphi(L)$  et  $\theta(L)$  caractérisent respectivement les parties AR (autorégressives) et MA (moving average) du modèle. Et les polynômes  $\varphi(L^s)$  et  $\theta(L^s)$  caractérisent respectivement les parties AR et MA saisonnières.

$(1 - L)^d$  : Ce terme sert à différencier le signal, quand c'est nécessaire afin de le rendre stationnaire avant de passer à l'identification. " $d$ " est le degré de différenciation, en général le maximum est de 3.

Si  $d = 1$  alors,  $(1 - L)^1 X = X_t - X_{t-1}$

$L$  est l'opérateur retard.

$$\nabla = (1 - L) \quad (76)$$

$$\nabla_{365} = (1 - L^{365}) \quad (77)$$

$\varepsilon_t$  est un bruit blanc :  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

Remarquons que ces processus ont à peu près la même définition que les processus ARMA, la seule différence est qu'on permet au polynôme autorégressif d'avoir la racine unité.

## ***Chapitre.3 La méthodologie de Box & Jenkins***

---

## I. Introduction

Les travaux de G. BOX et G. JENKINS (dans les années soixante dix) ont beaucoup contribué dans la théorie et la pratique des modèles des série temporelles. L'objectif auquel ils se proposent de répondre dans leur ouvrage, "Time Series Analysis ; Forecasting and Control", est de construire un modèle aléatoire de type ARMA permettant de reproduire au mieux les réalisations d'une série temporelle. Trois étapes sont nécessaires à la construction des modèles: une phase d'identification, une phase d'estimation, une phase de validation. A l'issue de ces trois phases, une fois déterminé le meilleur modèle ARMA, on utilise ce modèle à des fins de prévision.

## II. L'identification

Cette étape est la plus difficile dans le cycle de trois étapes définie par Box-Jenkins. Alors, plusieurs modèles candidats seront retenus et éliminés successivement dans les étapes deux et trois du cycle jusqu'à en retenir le bon. Pour cela plusieurs critères sont nécessaires pour identifier les modèles.

Pour commencer, il n'ya pas de règles exactes qui nous permettent de connaître les degrés maximales des polynômes du modèle ARMA (p,q) de la série stationnaire. D'après Box et Jenkins (1976) La méthode d'identification est fondée sur la comparaison des moments empiriques de la série considérée aux moments théoriques associés aux différentes représentations potentielles. On se concentre ici sur les moments d'ordre deux résumés par la fonction d'autocorrélation (FAC) et la fonction d'autocorrélation partielle (FACP).

D'une manière générale, on peut déduire que :

- Pour un modèle AR(p) : la FAC de la série stationnaire ou devenu stationnaire décroît exponentiellement ou sinusoidalement et la FACP s'annule après un retard  $k=p$ .
- Pour un modèle MA(q) : la FACP de la série stationnaire ou devenu stationnaire décroît exponentiellement ou sinusoidalement et la FAC s'annule après un retard  $k=q$ .

Le tableau.1, résumé quelques cas d'identification concernant les modèles AR(1), AR(2), MA(1), MA(2) et ARMA(1,1), ainsi que les estimations préliminaires de leurs coefficients.[32]

Rappelons que  $\rho_k$  est le coefficient d'autocorrélation de la série stationnaire et  $\phi_{kk}$  est son coefficient d'autocorrélation partiel.

ordre	(1,d,0)	(0,d,1)
✓ Comportement de $\rho_k$	Décroit exponentiellement	$\rho_1 \neq 0$
✓ Comportement de $\phi_{kk}$	$\phi_{11} \neq 0$	Décroit exponentiellement
✓ Estimations préliminaire des paramètres du modèle	$\varphi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$
✓ Région d'admissibilité	$-1 < \varphi_1 < 1$	$-1 < \theta_1 < 1$
ordre	(2,d,0)	(0,d,2)
✓ Comportement de $\rho_k$	Une sinusoïde ou exponentiel atténuée	$\rho_1$ et $\rho_2$ toujours $\neq 0$
✓ Comportement de $\phi_{kk}$	$\varphi_1$ et $\varphi_2 \neq 0$	Une sinusoïde ou exponentiel atténuée
✓ Estimations préliminaire des paramètres du modèle	$\begin{cases} \varphi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\ \varphi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$	$\begin{cases} \rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \end{cases}$
✓ Région d'admissibilité	$\begin{cases} -1 < \varphi_2 < 1 \\ \varphi_2 + \varphi_1 < 1 \\ \varphi_2 - \varphi_1 < 1 \end{cases}$	$\begin{cases} -1 < \theta_2 < 1 \\ \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \end{cases}$
ordre	(1,d,1)	
✓ Comportement de $\rho_k$	décroissance exponentielle après le premier retard	
✓ Comportement de $\phi_{kk}$	dominé par une décroissance exponentielle après le premier retard	
✓ Estimations préliminaire des paramètres du modèle	$\rho_1 = \frac{(1-\theta_1\varphi_1)(\varphi_1-\theta_1)}{1+\theta_1^2-2\varphi_1\theta_1}$	$\rho_2 = \rho_1\varphi_1$
✓ Région d'admissibilité	$-1 < \varphi_1 < 1$	$-1 < \theta_1 < 1$

**Tableau. 1:** comportement de la FAC et la FACP des processus ARMA (p,q)

Pour un processus ARMA (p,q) stationnaire d'ordres élevés, la FAC se comporte comme celle d'un processus AR(p) après q – p retards, et la FAP se comporte comme celle d'un processus MA(q) après p – q retards.

Le tableau suivant résume les propriétés des différents processus AR(p), MA(q) et ARMA(p,q). [32]

	Processus AR(p)	Processus MA(q)	Processus mixte ARMA(p,q)
✓ Modèle en termes de précédents $X_t$	$\varphi(B)X_t = \varepsilon_t$	$\theta^{-1}(B)X_t = \varepsilon_t$	$\theta^{-1}(B)\varphi(B)X_t = \varepsilon_t$
✓ Modèle en termes de précédents $\varepsilon_t$	$X_t = \varphi^{-1}(B)\varepsilon_t$	$X_t = \theta(B)\varepsilon_t$	$X_t = \varphi^{-1}(B)\theta(B)\varepsilon_t$
✓ Conditions de stationnarité	Les racines de $\varphi(B) = 0$ se trouvent en dehors du cercle unité	Toujours stationnaire	Les racines de $\varphi(B) = 0$ se trouvent en dehors du cercle unité
✓ Conditions d'inversibilité	Toujours inversible	Les racines de $\theta(B) = 0$ se trouvent en dehors du cercle unité	Les racines de $\theta(B) = 0$ se trouvent en dehors du cercle unité
✓ Fonction d'autocorrélation	Infinie (exponentielle et/ou onde sinusoïdale atténuée)  diminue	finie  S'annule	Infinie (exponentielle et/ou onde sinusoïdale atténuée après q-p retards)  diminue
✓ Fonction d'autocorrélation partielle	finie  S'annule	Infinie (dominée par une exponentielle et/ou onde sinusoïdale atténuée)  diminue	Infinie (dominée par une exponentielle amortie et/ou onde sinusoïdale atténuée après p-q retards)  diminue

**Tableau.2** : propriétés des processus AR(p), MA(q) et ARMA (p,q)



### III. Estimation

L'estimation des paramètres d'un modèle ARMA(p, q) lorsque les ordres p et q sont supposés connus peut s'effectuer comme suit :

#### III.1 Système d'équations de Yule-Walker :

Pour les processus AR(p), on peut retrouver les estimateurs des paramètres  $\varphi_1, \varphi_2, \dots, \varphi_p$  et  $\sigma_\varepsilon^2$  du modèle par la résolution des équations de Yule-Walker qui sont déduites directement de la fonction d'autocorrélation que nous avons déjà présenté au chapitre précédent :

$$\rho_k = \begin{cases} \sum_{j=1}^p \varphi_j \rho_{k-j} & \text{si } k > 0 \\ \sum_{j=1}^p \varphi_j \rho_{-j} + \sigma_\varepsilon^2 & \text{si } k = 0 \end{cases} \quad (78)$$

Dont l'écriture matricielle est la suivante :

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_{-1} & \dots & \rho_{1-p} \\ \rho_1 & \rho_0 & \dots & \rho_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix} \quad (79)$$

Tel que, les  $\rho_i$  ( $i = 0; p \in \mathbb{N}$ ) sont les estimateurs des coefficients de corrélation de la série stationnaire  $X_t$ .

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\sigma_{X_t} \sigma_{X_{t-k}}} = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sqrt{\sum_{t=k+1}^n (X_t - \bar{X})^2} \sqrt{\sum_{t=k+1}^n (X_{t-k} - \bar{X})^2}} \quad (80)$$

Le signal étant réel et  $\rho_k = \rho_{-k}$  ; on retrouve alors une matrice d'autocorrélation qui est symétrique et appelé aussi de Toeplitz.

Il suffit donc de résoudre les p équations à p inconnues pour pouvoir trouver les coefficients  $\varphi_1, \varphi_2, \dots, \varphi_p$  et pour la variance  $\sigma_\varepsilon^2$  on pose  $k = 0$ .

### III.2 Méthode du maximum de vraisemblance :

C'est la méthode la plus couramment utilisée pour estimer les différents paramètres  $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$  du modèle ARMA (p,q) dont la représentation est la suivante :

$$\varphi(L)X_t = \theta(L)\varepsilon_t + \varphi_0 \quad (80)$$

D'où :

$$\varepsilon_t = \sum_{i=1}^p \varphi_i X_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} - \varphi_0 \quad (81)$$

Le principe de la vraisemblance revient à déterminer la valeur des paramètres  $(\varphi_0, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$  fonction des observations  $(X_1, X_2, \dots, X_T)$  ; qui assure la plus grande probabilité d'apparition de ces observations  $(X_1, X_2, \dots, X_T)$ .

Pour commencer ; on fait l'hypothèse de la normalité des résidus  $\varepsilon_t$  avant de spécifier une forme fonctionnelle à la vraisemblance du modèle tel que ;  $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$  dont la fonction de distribution est :

$$f(\varepsilon) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}} \quad (82)$$

Et la fonction vraisemblance associée au vecteur de réalisation  $(X_1, X_2, \dots, X_T)$  est :

$$\begin{aligned} f(X_1, X_2, \dots, X_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2) &= \prod_{t=1}^T \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(X_t - E(X_t))^2}{2\sigma_\varepsilon^2}} \\ &= \left( \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^T \exp \left[ \sum_{t=1}^T -\frac{(X_t - E(X_t))^2}{2\sigma_\varepsilon^2} \right] \end{aligned} \quad (83)$$

Afin de maximiser la fonction de vraisemblance, on doit d'abord considérer son logarithme pour faciliter les calculs :

$$\begin{aligned} \ln \left( f(X_1, X_2, \dots, X_T; \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2) \right) &= T \ln \left( \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right) - \left[ \sum_{t=1}^T \frac{(\varepsilon_t)^2}{2\sigma_\varepsilon^2} \right] \\ &= -\frac{T}{2} \ln(\sigma_\varepsilon^2) - \frac{T}{2} \ln(2\pi) - \left[ \sum_{t=1}^T \frac{(\varepsilon_t)^2}{2\sigma_\varepsilon^2} \right] \quad (84) \end{aligned}$$

Enfin, pour maximiser la fonction vraisemblance, cela revient à minimiser le terme d'erreur :  $(\varepsilon_t)^2$

#### IV. Tests de validation du modèle :

Après avoir estimé les paramètres du modèle, on doit s'assurer de sa validité avant de faire de la prévision, pour cela, plusieurs critères sont à vérifier à travers des tests qui ont été développés.

##### IV.1 La parcimonie :

D'après Box & Jenkins, l'une des règles de sélection d'un modèle est la parcimonie qui consiste à déterminer et estimer le minimum de paramètres possible. Dans ce processus de sélection d'un modèle ARMA (p,q) ; une augmentation de p ou de q fait diminuer la somme des résidus au carré mais fait aussi diminuer le nombre de degré de liberté ce qui va dans le sens opposé de la parcimonie. Pour cela on doit faire un arbitrage entre ces deux facteurs à l'aide de deux critères de comparaison ; AIC (Akaike information criterion)( Akaike,1974) et BIC (Baisian information critérium)(Schwarz,1978) qui sont les premiers critères apparaissant dans la littérature et largement diffusés et appliqués.

Le principe est d'avoir le minimum possible pour un meilleur modèle.

- **Le critère d'information d'Akaike (AIC) :** [32]

La formulation mathématique d'AIC est définie comme suit :

$$AIC(p + q) = T \ln(\sigma_\varepsilon^2) + 2(p + q) \quad (85)$$

Où p et q sont le nombre de paramètres d'AR et de MA à estimer. Le modèle qui donne le minimum AIC est choisi comme modèle parcimonieux, de même pour le BIC.

- **Le Critère d'information Bayésien (BIC):** [32]

$$BIC(p + q) = T \ln \sigma_{\varepsilon}^2 + (p + q) \ln T \quad (86)$$

T : le nombre d'observation

$\sigma_{\varepsilon}^2$  : Variance des résidus

## IV.2 Test sur les résidus

### IV.2.1 Test d'autocorrélations des résidus

Il est aussi important que le modèle ainsi retenu ne présente pas d'autocorrélation des erreurs résiduelles et ceci revient à dire que La fonction d'autocorrélation des résidus (RACF) devrait obéir à celle d'un bruit blanc. Il y a deux méthodes utilisées :

- La première est de dessiner le corrélogramme en traçant  $r_k(\varepsilon)$  en fonction du retard k. Si une partie du RACF est significativement différente de zéro, ceci peut signifier que le modèle actuel est insatisfaisant.

- La seconde est la statistique Q (r) qui est la plus récente suggérée par Ljung et Box (1978). Le principe est de vérifier l'hypothèse d'adéquation du modèle en choisissant un niveau de signification ensuite en comparant la valeur de  $\chi^2$  calculé à  $\chi^2$  tabulée de la valeur critique.

Si la valeur calculée de  $\chi^2$  est plus petite que la valeur critique  $\chi^2$  tabulée, le modèle est adéquat sur la base des données disponibles.

La statistique de Q (r) est construite comme suit: [33]

$$Q(r) = T' (T' + 2) \sum_{k=1}^m (T' - k)^{-1} r_k(\varepsilon)^2 \xrightarrow{T \rightarrow \infty} \chi^2 (m - p - q) \quad (87)$$

Tel que :

$$r_k(\varepsilon) = \frac{\sum_{i=k+1}^T \varepsilon_i \varepsilon_{i-k}}{\sum_{i=1}^T \varepsilon_i^2} \quad (88)$$

Q(r) : statistique de Ljung-Box au retard m

$r_k(\varepsilon)$  : ACF de  $\varepsilon_k$  au retard k

m : un nombre présélectionné des autocorrélations

$T' = T-d$  : le nombre d'observations restantes après d différenciations afin de stationnariser la série

#### IV.2.2 Test de normalité des résidus

Pour vérifier si le processus des résidus  $\{\varepsilon_t, t \in \mathbb{Z}\}$  est un bruit blanc gaussien, plusieurs tests peuvent être utilisés, mais le test le plus courant est celui de Jaque et Bera. Ce dernier est fondé sur la notion de skewness  $S_k$  (moment d'ordre 3 et asymétrie) et de Kurtosis  $K_u$  (moment d'ordre 4 et queue de distribution).

La statistique du test de Jaque et Bera est comme suit :

$$S = \frac{T}{6} S_k + \frac{T}{24} (K_u - 3)^2 \xrightarrow{T \rightarrow \infty} \chi^2(2) \quad (89)$$

Si  $S \geq \chi_{1-\alpha}^2(2)$  ; on rejette l'hypothèse  $H_0$  de normalité des résidus au seuil de  $\alpha$  %

### V. La prévision :

Soit le processus stationnaire  $(X_t)$  observé entre 1 et T qui s'écrit ;  $\varphi(L)X_t = \theta(L)\varepsilon_t$  et on cherche à faire de la prévision à l'horizon h ce qui signifie prévoir  $X_{T+1}, X_{T+2}, \dots, X_{T+h}$ .

#### V.1 La prévision dans le cas d'un modèle AR(p)

Soit le processus centré autorégressif d'ordre p suivant :

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t \quad (90)$$

Supposons l'origine des prévisions est a la date T, ainsi la prévision à la date T+1 est :

$$X_{T+1}^T = X_{T+1} = \varphi_1 X_T + \dots + \varphi_p X_{T-p} \quad (91)$$

Par analogie au rang  $h$  ;

$$X_{T+h} = \varphi_1 X_{T+h-1} + \dots + \varphi_p X_{T+h-p} + \varepsilon_{T+h} \quad (92)$$

Donc :

$$X_{T+h}^* = E(X_{T+h} \setminus X_T, X_{T-1}, \dots) \quad (93)$$

L'indice "\*" désigne la prévision (ou la valeur prédite)

Ainsi d'une façon récursive, on aura :

$$X_{T+h}^* = \begin{cases} \varphi_1 X_{T+h-1}^* + \dots + \varphi_{h-1} X_{T+1}^* + \varphi_h X_T + \dots + \varphi_p X_{T+h-p} & \text{pour } h \leq p \\ \varphi_1 X_{T+h-1}^* + \dots + \varphi_p X_{T+h-p}^* & \text{pour } h > p \end{cases} \quad (94)$$

On remarque que ces prévisions s'obtiennent donc par itérations successives et sont calculables à partir des observations.

## V.2 La prévision dans le cas d'un modèle MA(q)

Le processus MA(q) satisfait la relation ;  $X_t = \theta(L)\varepsilon_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

La prévision à la date  $T+1$  à partir de l'instant  $T$  est :

$$\begin{aligned} X_{T+1}^* &= E(X_{T+1} \setminus X_T, X_{T-1}, \dots) = E(X_{T+1} \setminus \varepsilon_T, \varepsilon_{T-1}, \dots) \\ &= \varepsilon_{T+1} + \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T+1-q} \\ &= 0 + \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T+1-q} \end{aligned} \quad (95)$$

Ainsi la prévision à l'horizon  $h$  s'écrit:

$$\begin{aligned} X_{T+h}^* &= E(X_{T+h} \setminus X_T, X_{T-1}, \dots) = E(X_{T+h} \setminus \varepsilon_T, \varepsilon_{T-1}, \dots) \\ X_{T+h}^* &= \begin{cases} \theta_h \varepsilon_T + \dots + \theta_q \varepsilon_{T+h-q} & \text{pour } h \leq q \\ 0 & \text{pour } h > q \end{cases} \end{aligned} \quad (96)$$

Mais l'inconvénient dans ces calculs ; c'est qu'on estime  $X_{T+h}^*$  à partir des erreurs passées qui ne sont pas du tous observables. Afin d'y remédier, nous procéderons a la prévision du modèle MA(q) à travers un processus AR( $\infty$ ) qui fera introduire les valeurs passées de la variable  $X_t$ .

### V.2.1 Prévision à l'aide d'un processus AR( $\infty$ ) pour un MA(q)

L'équation  $X_t = \theta(B)\varepsilon_t$  peut se réécrire  $\theta^{-1}(L)X_t = \varepsilon_t$

D'où :

$$X_t = \sum_{k=1}^{\infty} A_k X_{t-k} + \varepsilon_t \quad (97)$$

La prévision à l'horizon h est :

$$X_{T+h}^* = \sum_{k=1}^{h-1} A_k X_{T+h-k}^* + \sum_{k=h}^{\infty} A_k X_{T+h-k} \quad (98)$$

En supposant que  $X_t$  n'est pas observable pour  $t < 0$ , alors on pourra considérer l'écriture suivante :

$$X_{T+h}^* = \sum_{k=1}^{h-1} A_k X_{T+h-k}^* + \sum_{k=h}^{T+h} A_k X_{T+h-k} + \underbrace{\sum_{k=T+h+1}^{\infty} A_k X_{T+h-k}}_{\text{Négligeable } X_t = 0 \text{ pour } t < 0} \quad (99)$$

Ce qui nous donne une approximation de  $X_{T+h}^* = \hat{X}_{T+h}^*$

$$\hat{X}_{T+h}^* = \sum_{k=1}^{h-1} A_k X_{T+h-k}^* + \sum_{k=h}^{T+h} A_k X_{T+h-k} \quad (100)$$

### V.3 La prévision dans le cas d'un modèle ARMA (p,q)

Soit le processus ARMA(p,q) centré ( $X_t$ ) qui satisfait la relation suivante :

$$\varphi(L)X_t = \theta(L)\varepsilon_t \quad (101)$$

Alors  $X_t$  peut s'écrire comme suit :

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (102)$$

Par analogie :

$$X_{t+h} = \sum_{i=1}^p \varphi_i X_{t+h-i} + \varepsilon_{t+h} + \sum_{j=1}^q \theta_j \varepsilon_{t+h-j} \quad (103)$$

$$X_{T+h}^* = \begin{cases} \varphi_1 X_{T+h-1}^* + \dots + \varphi_{h-1} X_{T+1}^* + \varphi_h X_T + \dots + \varphi_p X_{T+h-p} & \text{pour } h \leq p \\ \varphi_1 X_{T+h-1}^* + \dots + \varphi_p X_{T+h-p}^* & \text{pour } h > p \end{cases} \quad (104)$$

Cette forme de prévision est comme dans le cas d'un modèle AR(p). la prévision se calcule de manière itérative.

#### V.4 La prévision dans le cas d'un processus ARIMA (p,d,q)

Soit le processus  $X_t$  satisfaisant la relation :

$$\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad (105)$$

Pour les ARIMA (p,d,q) , on a besoin des conditions initiales suivantes:

$$Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q}) \quad (106)$$

On posera :  $\psi = \varphi(L)(1-L)^d$ , alors la forme de  $X_t$  devient :

$$X_t = \sum_{i=1}^{p+d} \psi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (107)$$

Donc :



$$X_{t+h} = \sum_{i=1}^{p+d} \psi_i X_{t+h-i} + \varepsilon_{t+h} + \sum_{j=1}^q \theta_j \varepsilon_{t+h-j} \quad (108)$$

Alors la prévision  $X_{T+h}^*$  qui sera faite à partir de la date T est :

$$X_{T+h}^* = \sum_{i=1}^{p+d} \psi_i X_{T+h-i}^* + 0 + \sum_{j=1}^q \theta_j \varepsilon_{T+h-j}^* \quad (109)$$

Te que :

$$\begin{cases} X_{T+h-i}^* = X_{t+h-i} & \text{pour } i \geq h \\ \varepsilon_{t+h-j}^* = \begin{cases} 0 & \text{pour } j < h \\ \varepsilon_{t+h-j} & \text{pour } j \geq h \end{cases} \end{cases} \quad (110)$$

#### V.4.1 Prévision d'un ARIMA (p,d,q) en utilisant l'approximation AR :

Pour faire cette approximation, on fait appel à la propriété suivante :

**Propriété :** soit  $(X_t)$  un proceu ARIMA  $(p,d,q)$  ;  $\varphi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t$  , de valeurs initiales  $Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$  , alors  $(X_t)$  peut s'ecrire sous la forme AR suivante :

$$X_t = \sum_{j=1}^t a_j X_{t-j} + f^*(t)Z + \varepsilon_t \quad (111)$$

Ou les  $a_j$  sont les coefficients (pour  $j \geq 1$ ) de la division selon les puissances croissantes de  $\varphi$  par , et  $f^*(t)$  est un vecteur (ligne) de fonctions de t qui tend vers 0 quand  $t \rightarrow \infty$ .

Par analogie :

$$X_{t+h} = \sum_{j=1}^{t+h} a_j X_{t+h-j} + f^*(t+h)Z + \varepsilon_{t+h} \quad (112)$$

Ainsi la prévision à l'horizon h est :

$$X_{T+h}^* = E(X_{T+h} | X_T, X_{T-1}, \dots, X_0, Z) = \sum_{j=1}^{t+h} a_j X_{T+h-j}^* + f^*(T+h)Z + 0 \quad (113)$$

Vu que  $X_{T+h-j}^* = X_{t+h-j}$  pour  $j \geq h$  et  $f^*(t)$  tend vers 0 quand  $t \rightarrow \infty$  ; on a l'approximation suivante :

$$\hat{X}_{T+h}^* = \sum_{j=1}^{h-1} a_j \hat{X}_{T+h-j}^* + \sum_{j=h}^{T+h} a_j X_{T+h-j} \quad (114)$$

#### V.4.2 Prévision d'un ARIMA (p,d,q) en utilisant l'approximation MA :

De la même façon, un processus ARIMA peut être approximé par un processus MA suite à la propriété suivante :

**Propriété :** soit  $(X_t)$  un proceu ARIMA  $(p,d,q)$  ;  $\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t$  , de valeurs initiales  $Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$  , alors  $(X_t)$  peut s'écrire sous la forme MA suivante :

$$X_t = \sum_{j=1}^t b_j \varepsilon_{t-j} + g^*(t)Z \quad (115)$$

Ou les  $b_j$  sont les coefficients de la division selon les puissances croissantes de  $\theta$  par  $\varphi$  et  $g^*(t)$  est un vecteur (ligne) de fonctions de t.

Par analogie :

$$X_{t+h} = \sum_{j=1}^{t+h} b_j \varepsilon_{t+h-j} + g^*(t+h)Z \quad \text{avec } \lim_{u \rightarrow \infty} g(u) = 0 \quad (116)$$

Ainsi la prévision à l'horizon h est :

$$\hat{X}_{T+h}^* = \sum_{j=h}^{T+h} b_j \varepsilon_{T+h-j} \quad (\varepsilon_{T+h-j} = 0 \text{ pour } j < h) \quad (117)$$

### V.5 Prédiction dans le cas des modèles SARIMA [31]

Soit le processus SARIMA (p,d,q)(P,D,Q) suivant :

$$(1 - L)^d(1 - L^S)X_t\varphi(L)\phi(L^S) = \theta(L)\vartheta(L^S)\varepsilon_t \quad (118)$$

Ou :

$$\nabla^d\nabla_S X_t\varphi(L)\phi(L^S) = \theta(L)\vartheta(L^S)\varepsilon_t \quad (119)$$

Tel que :

$$\begin{cases} \nabla^d = (1 - L)^d \\ \nabla_S = (1 - L^S) \end{cases} \quad (120)$$

Après l'identification et l'estimation des paramètres du processus stationnaire  $W_t = \nabla^d\nabla_S X_t$ , on revient à notre série originale  $X_t$  pour faire la prédiction à l'horizon  $h$ .

Afin de mieux comprendre les étapes de la prédiction, on s'appuiera sur un exemple illustré par Box et Jenkins [31] :

Soit le processus suivant :

$$\nabla\nabla_{12}X_t = (1 - \theta L)(1 - \vartheta L^{12})\varepsilon_t \quad (121)$$

$$\text{Avec } \begin{cases} \hat{\theta} = 0.4 \\ \hat{\vartheta} = 0.6 \end{cases}$$

$$\text{Et } \varepsilon_t \sim (0, \sigma_\varepsilon^2), \quad \sigma_\varepsilon^2 = 1.34 * 10^{-3}$$

Après développement de l'équation (121), on aura :

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + \varepsilon_t - 0.4\varepsilon_{t-1} - 0.6\varepsilon_{t-12} + 0.24\varepsilon_{t-13} \quad (122)$$

La prédiction à l'horizon «  $h$  » selon Box & Jenkins s'écrit comme suit :

$$\begin{aligned} \hat{X}_t^*(h) = & X_{t+h-1} + X_{t+h-12} - X_{t+h-13} + \varepsilon_{t+h} - 0.4\varepsilon_{t+h-1} - 0.6\varepsilon_{t+h-12} \\ & + 0.24\varepsilon_{t+h-13} \end{aligned} \quad (123)$$

Tel que :

$$X_{t+j} = \begin{cases} X_{t+j} & j \leq 0 \\ \hat{X}_t(j) & j > 0 \end{cases} \quad (124)$$

$$\varepsilon_{t+j} = \begin{cases} \varepsilon_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} \quad (125)$$

Ceci revient à dire que ; nous remplacerons les valeurs inconnues de  $Y_{t+j}$  par leurs prévisions  $\hat{Y}_t(j)$  et les résidus inconnu par des zéros. Bien sur, après la première étape de prévision, on doit calculer les nouveaux résidus (mettre a jour) à chaque fois, comme suit :

$$\varepsilon_t = X_t - \hat{X}_{t-1}(1) \quad (126)$$

Par exemple, pour prévoir la valeur de  $\hat{X}_t(3) = X_{t+3}$ , c.à.d. la troisième valeur après l'origine  $t$ , on procédera de la manière suivante :

$$X_{t+3} = X_{t+2} + X_{t-9} - X_{t-10} + \varepsilon_{t+3} - 0.4\varepsilon_{t+2} - 0.6\varepsilon_{t-9} + 0.24\varepsilon_{t-10} \quad (127)$$

Ainsi la prévision de  $X_{t+3}^*$  est :

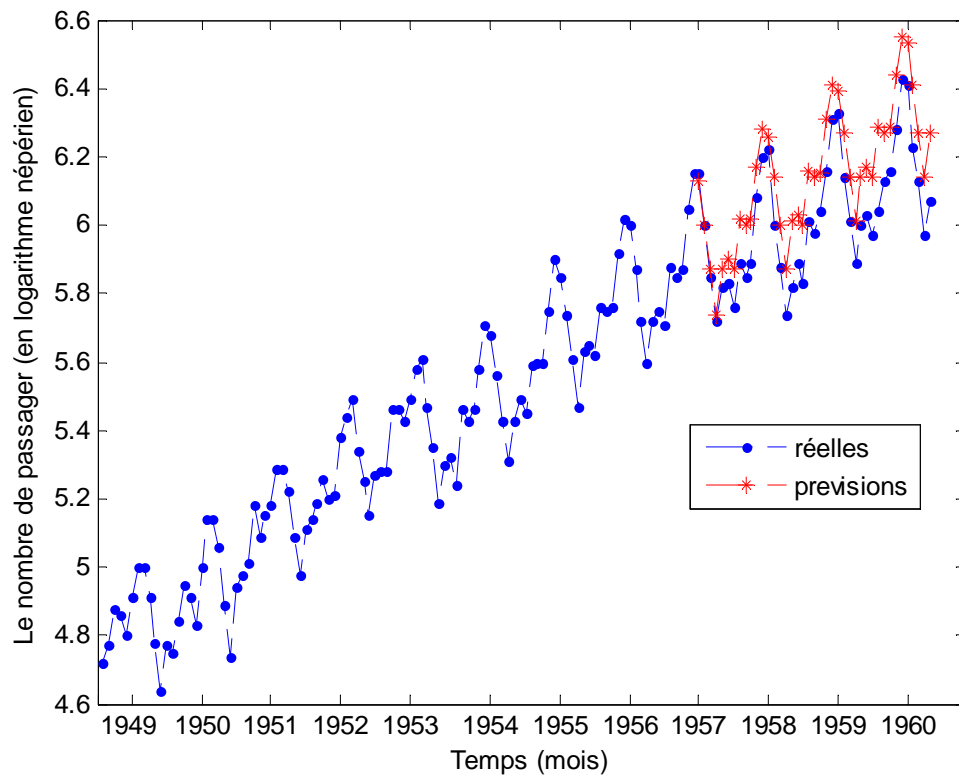
$$\begin{aligned} X_{t+3}^* = \hat{X}_t^*(3) &= \hat{X}_t^*(2) + X_{t-9} - X_{t-10} - 0.6\{X_{t-9} - \hat{X}_{t-10}^*(1)\} \\ &+ 0.24\{X_{t-10} - \hat{X}_{t-11}^*(1)\} \end{aligned} \quad (128)$$

$$\hat{X}_t^*(3) = \hat{X}_t^*(2) + 0.4X_{t-9} - 0.76X_{t-10} + 0.6\hat{X}_{t-10}^*(1) - 0.24\hat{X}_{t-11}^*(1) \quad (129)$$

Pour le calcul de  $\hat{X}_{t-10}^*(1)$  et  $\hat{X}_{t-11}^*(1)$ , on se réfère aussi à l'équation (123).

NB : comme l'ordre de l'équation est égal à 13, alors il nous faut 13 x 2 valeurs de la série temporelle pour faire la prévision.

La Fig.7 illustre les résultats de la prévision de la série citée en exemple dont nous avons entrepris les calculs.



**Fig.7 :** la série X et ses prévisions pour les 36 mois à compter de Juillet 1957

**NB :** D'une manière générale, pour faire la prévision d'une série temporelle, il faudrait que le nombre d'observation dépasse au moins deux fois l'ordre de l'équation caractéristique du modèle SARIMA.

## *Chapitre.4 Analyses et Résultats*

---

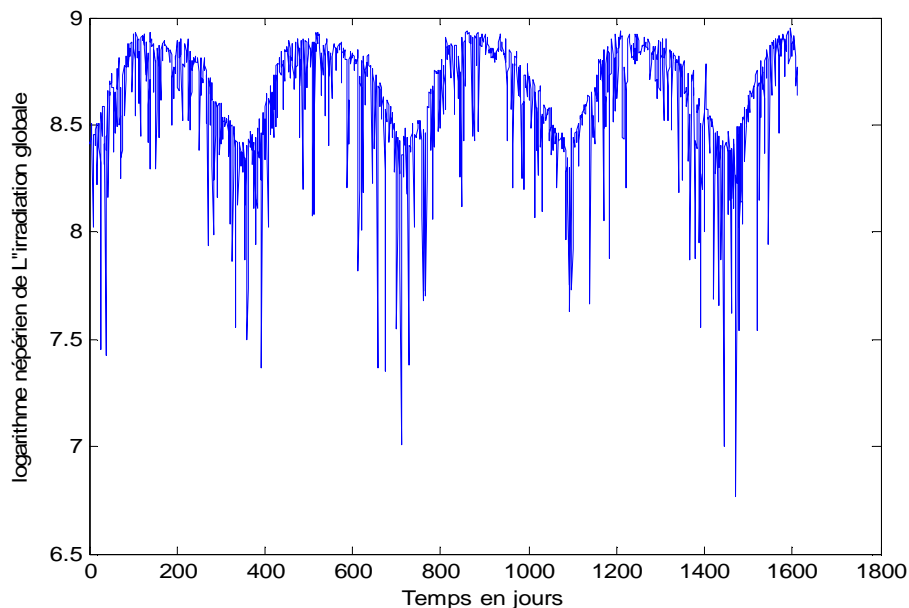
## I. Introduction :

Après avoir formé la série temporelle des moyennes journalières de l'irradiation globale dont les données s'étalent sur une période de 5 ans et demi (de janvier 2005 à Mai 2010), nous procéderons d'abord au traitement stationnaire de cette série qui sera ensuite modélisée.

Avant de procéder à la modélisation, nous effectuerons un changement concernant notre variable temporelle  $E_t$  (l'irradiation globale), on posera :  $G_t = \mathbf{Ln}(E_t)$  (Ln est le logarithme népérien), afin de réduire l'ordre de grandeur de la variable et faciliter les calculs.

## II. Etape d'identification

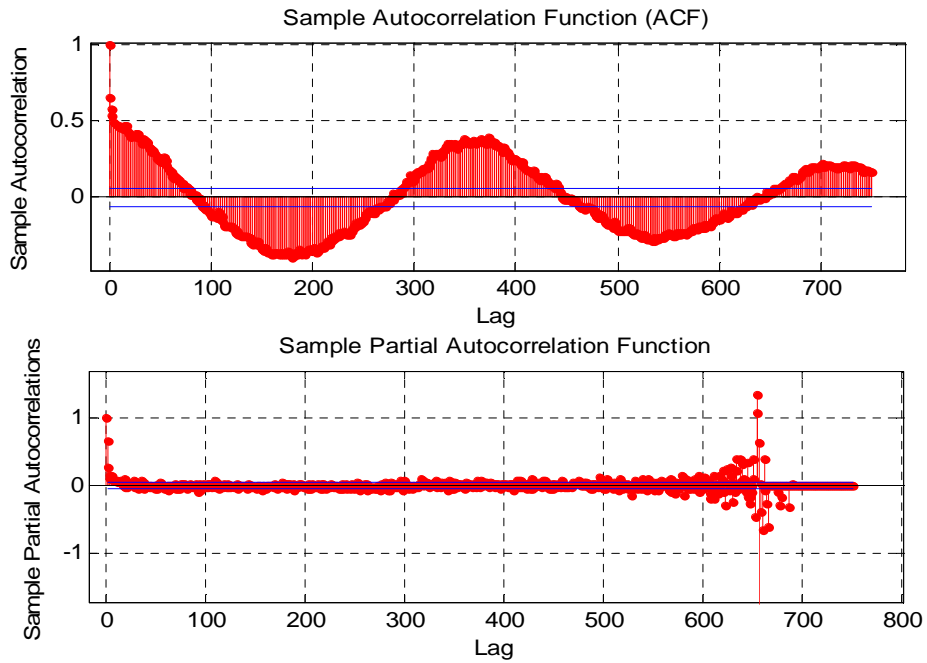
En traçant la courbe de notre variable en fonction du temps (Fig.8), on voit qu'il y a déjà une saisonnalité apparente avec une tendance sinusoïdale, donc il est clair que la série n'est pas stationnaire.



**Fig.8 :** Données de l'irradiation globale journalières du 1 Janvier 2005 au 31 Mai 2009

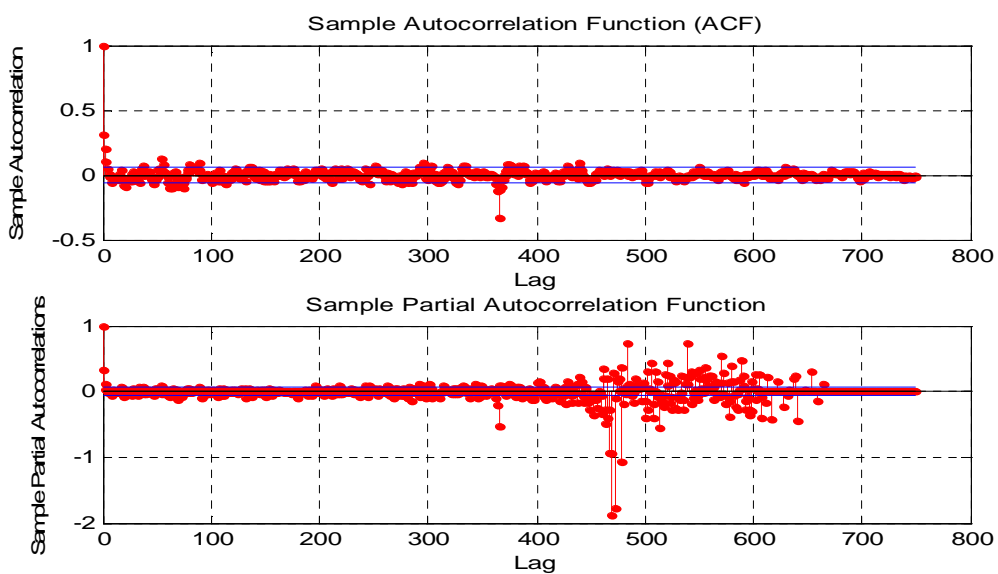
De même En traçant l'autocorrélogramme et l'autocorrélogramme partiel de notre série que nous appelons  $G_t$  (qui est la variable irradiation globale), on remarque une forte autocorrélation avec des pics au point 365, 730, ce qui explique la présence de la saisonnalité d'une période de 365 jours (Fig.9).

On déduit, que nous pouvons déjà multiplier notre signal  $G_t$  par  $(1 - L^s)$ , tel que  $s=365$  afin d'éliminer l'effet de la saisonnalité et stationnariser notre signal avant l'étape d'identification.



**Fig.9** : l'autocorrélogramme et l'autocorrélogramme partiel de  $G_t$

En multipliant notre signal par l'opérateur  $\Delta_{365}$  (le filtre  $(1 - L^s)$ ). Le tracé des autocorrélogramme et autocorrélogramme partiel sont montrés dans la Fig.10.

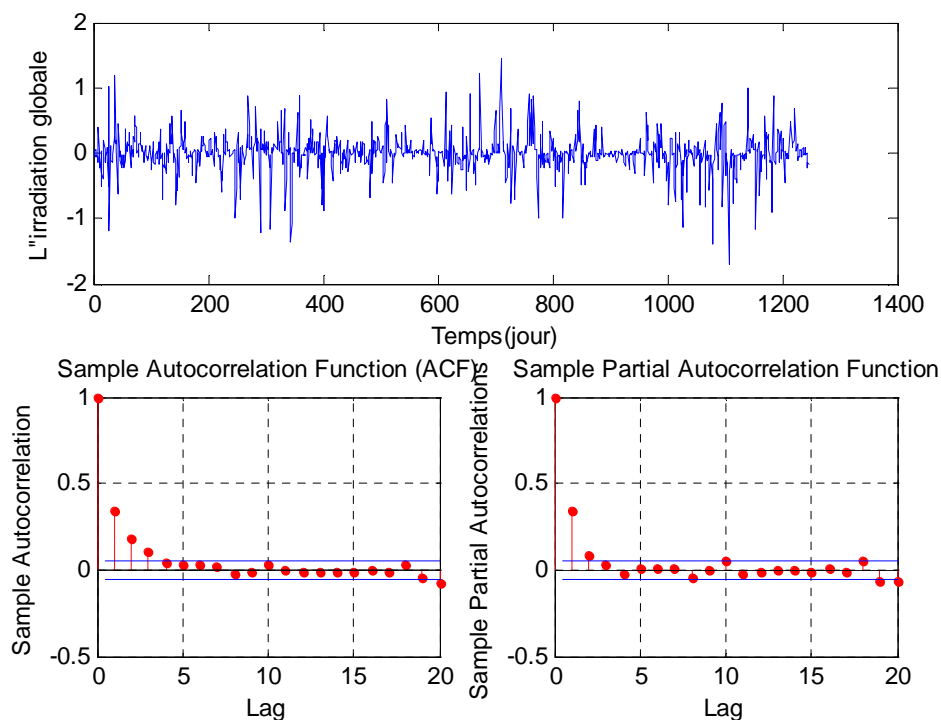


**Fig.10** : l'autocorrélogramme et l'autocorrélogramme partiel de  $\Delta_{365}G_t$



On voit bien que la différentiation  $\Delta_{365}G_t$  réduit nettement les autocorrélations et élimine la saisonnalité. Nous en déduisons que notre signal peut être modélisé par le modèle SARMA (seasonal autoregressive moving average).

A partir de ces autocorrélogrammes, on déduira les degrés maximaux des polynômes  $\varphi(B)$   $\phi(B^s)$   $\theta(B)$   $\vartheta(B^s)$ . Déjà on remarque qu'il ya un paramètre MA saisonnier  $\theta$  ainsi qu'un paramètre AR saisonnier  $\phi$  qui sont significativement différents de 0. Ensuite, On tracera ces autocorrélogrammes pour les 20 premiers retards seulement afin d'identifier le nombre de paramètres AR et MA Fig.11. On voit aussi la serie ainsi différenciée  $X_t = \Delta_{365}G_t$ .



**Fig .11 :**  $X_t$  et ses autocorrélogrammes

Ces autocorrélogrammes montrent deux paramètres AR et trois paramètres MA qui sont significativement différents de 0 mais on doit tester tous les cas possibles, combinant ces différents paramètres, lors de leurs estimation en se basant sur quelques tests statistiques.

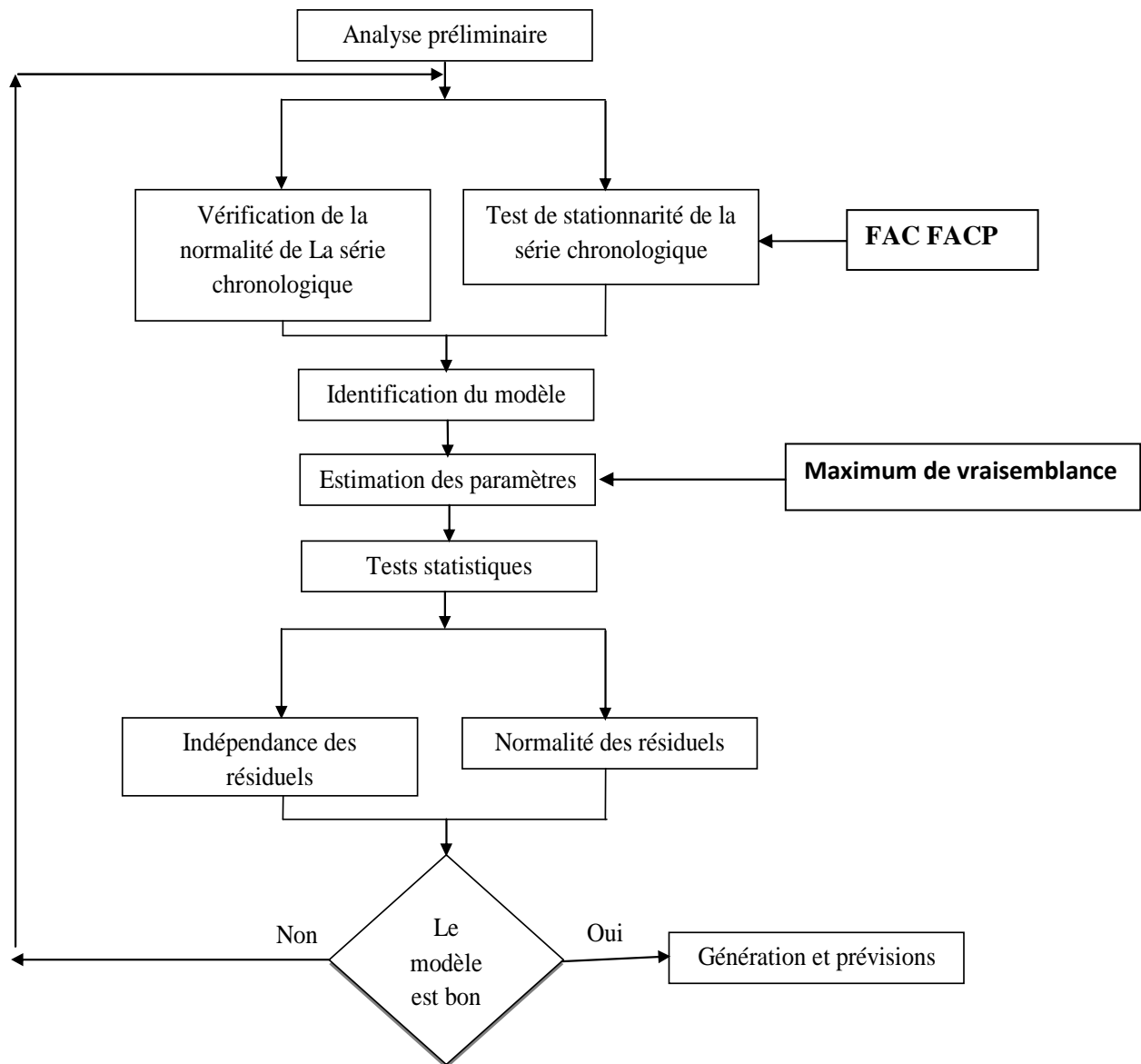
La forme générale du modèle SARMA en tenant compte de tous les paramètres possibles et sur lequel on procédera à l'identification et à l'estimation est le modèle SARIMA(2,0,3)(1,1,1) dont la relation est la suivante:

$$G_t(1 - L^{365})(1 - \varphi_1L - \varphi_2L^2)(1 - \vartheta L^{365}) = (1 - \theta_1L - \theta_2L^2 - \theta_3L^3) * (1 - \vartheta L^{365})\varepsilon_t \tag{130}$$

Tel que :  $X_t = G_t(1 - L^{365})$

Avec :  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  est un bruit blanc.

On résumera les différentes étapes par l'organigramme suivant :



**Fig.12** : Processus itératif de la modélisation

Lors de l'estimation, on examinera la série  $X_t$  au lieu de  $G_t$

### III. Estimation des paramètres :

Afin d'estimer le vecteur des paramètres  $\theta = (\varphi_1, \varphi_2, \theta_1, \theta_2, \theta_3, \vartheta, \emptyset \text{ et } \sigma_\varepsilon^2)$ , on s'appuiera sur la méthode du maximum de vraisemblance que nous avons déjà développée au chapitre.3.

Après avoir développé l'équation (130), on aboutira au résultat suivant :

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \emptyset X_{t-365} + \varphi_1 \emptyset X_{t-366} + \varphi_2 \emptyset X_{t-367} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} \\ - \vartheta \varepsilon_{t-365} + \theta_1 \vartheta \varepsilon_{t-366} + \theta_2 \vartheta \varepsilon_{t-367} + \theta_3 \vartheta \varepsilon_{t-368} \quad (131)$$

D'où :

$$\varepsilon_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \emptyset X_{t-365} + \varphi_1 \emptyset X_{t-366} + \varphi_2 \emptyset X_{t-367} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} \\ + \vartheta \varepsilon_{t-365} - \theta_1 \vartheta \varepsilon_{t-366} - \theta_2 \vartheta \varepsilon_{t-367} - \theta_3 \vartheta \varepsilon_{t-368} \quad (132)$$

Pour  $t = 1, 2, \dots, T$  ( $T$  est le nombre d'échantillon). La log-vraisemblance conditionnelle est alors :

$$L(\theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_\varepsilon^2) - \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2} \quad (133)$$

Pour  $n = (0, -1, \dots, -p + 1)$ , les valeurs initiales de  $X_n = \frac{\varphi_0}{1 - \varphi_1 - \varphi_2 - \dots - \varphi_p}$  (134)

Pour  $n = (0, -1, \dots, -q + 1)$ , les valeurs initiales de  $\varepsilon_n = 0$  (135)

#### III.1 Principe de calcul :

Après avoir initialisé les valeurs de  $X_n$  et  $\varepsilon_n$ , on procédera au calcul de  $\varepsilon_t$  (équation (2),  $t=1, \dots, T$ ) à l'aide d'un premier vecteur de paramètres aléatoire  $\theta = (\varphi_1, \varphi_2, \theta_1, \theta_2, \theta_3, \vartheta, \emptyset \text{ et } \sigma_\varepsilon^2)$ , ensuite on passera à la maximisation de la fonction log-vraisemblance (3) en appliquant les contraintes suivantes :

$$-1 < \varphi < 1 \quad -1 < \theta < 1 \quad -1 < \emptyset < 1 \quad -1 < \vartheta < 1$$

A partir de SARIMA(2,0,3)(1,1,1), on a considéré tous les cas possible que nous résumerons dans le tableau suivant avec les résultats de calcul de quelques tests statistiques

comme le AIC(le critère d'Akaike) et le BIC(le critère Bayezian) ainsi que la statistique de student.

	MVS	AIC	BIC	t-statistic
$\varphi_1 \theta_1 \vartheta$	50.76	-95.51	-80.13	1.516
$\varphi_1 \theta_1 \emptyset \vartheta$	52.15	-96.29	-75.78	1.481
$\varphi_1 \emptyset \vartheta$	49.14	-92.27	-76.89	1.442
$\varphi_1 \vartheta$	47.79	-91.58	-81.32	1.479
$\varphi_1 \theta_1 \theta_2 \emptyset$	6.78	-5.56	14.96	0.703
$\varphi_1 \varphi_2 \emptyset$	6.56	-7.12	8.26	0.697
$\varphi_1 \varphi_2 \theta_1 \emptyset$	6.78	-5.56	14.96	0.703
$\varphi_1 \varphi_2 \theta_1 \theta_2 \vartheta$	50.78	-91.56	-65.92	1.521
$\varphi_1 \theta_1 \theta_2 \vartheta$	<b>50.78</b>	-93.55	-73.04	1.520
$\varphi_1 \varphi_2 \theta_1 \vartheta$	50.77	-93.54	-73.03	1.519
$\varphi_1 \theta_1 \theta_2 \theta_3 \vartheta$	51.45	-92.90	-67.26	1.483
$\varphi_1 \varphi_2 \theta_1 \vartheta$	51.28	-94.56	-74.05	1.498
$\varphi_1 \theta_1 \theta_2 \theta_3 \vartheta$	52.17	-94.34	-68.70	1.485
$\varphi_1 \varphi_2 \theta_1 \emptyset \vartheta$	52.16	-94.324	-68.686	1.484
$\varphi_1 \varphi_2 \theta_1 \theta_2 \emptyset \vartheta$	52.71	-93.43	-62.66	1.493

**Tableau.3** : critères de sélection du modèle

La t-statistique a été proposée par Stone [10] qui est aussi un critère de sélection. il est basée sur le calcul du NRMSE (L'erreur quadratique moyenne normalisée) et du NMBE (L'erreur systématique moyenne normalisée) tel que :

$$t = \left( \frac{(N-1)NMBE^2}{NRMSE^2 - NMBE^2} \right)^{1/2} \quad (136)$$

Plus la valeur de t est petite meilleur est le modèle. Aussi, à condition est que la valeur t soit aussi inférieure à celle affichée dans le tableau représentatif de la loi de Student (qui égale à 1.96 pour un intervalle de confiance de 5% .

Vu que les deux critères AIC et BIC doivent être le plus inférieurs possible ainsi que la t-statistique, et que la MVS( la fonction maximum de vraisemblance) doit être le maximum

possible, ceci nous mène à conclure que le modèle SARIMA(1,0,1)(1,1,1) est le meilleur modèle que nous adopterons pour la prévision et qui s'écrit comme suit avec les paramètres estimés :

$$X_t(1 - \varphi_1 L)(1 - \phi L^{365}) = (1 - \theta_1 L)(1 - \theta L^{365})\varepsilon_t \quad (137)$$

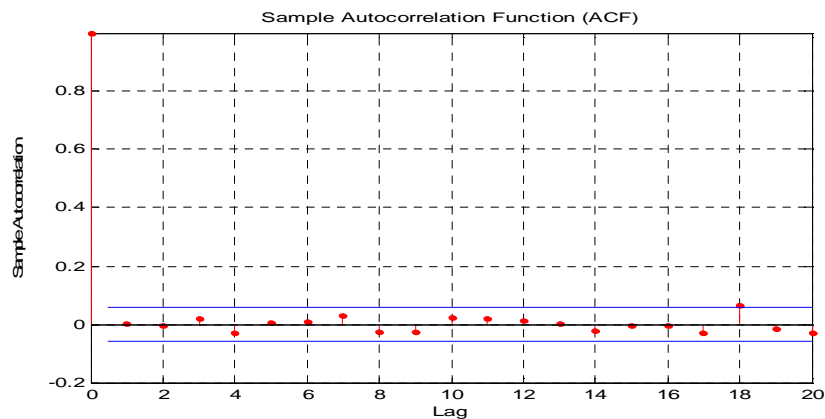
$$X_t(1 - 0.513L)(1 - 0.114L^{365}) = (1 - 0.219L)(1 - 0.705L^{365})\varepsilon_t \quad (138)$$

Avec :  $\varepsilon_t \sim (0, 0.054)$

### Test sur les résidus $\varepsilon_t$ :

Et on doit aussi s'assurer que le résidu  $\varepsilon_t$  est un bruit blanc et ceci revient à dire que La fonction d'autocorrélation des résiduelle (RACF) devrait obéir à celle d'un bruit blanc. Il y a deux méthodes utilisées :

La première est de dessiner le corrélogramme en traçant  $r_k$  (a) en fonction du retard  $k$ . Si une partie du RACF est significativement différente de zéro, ceci peut signifier que le modèle actuel est insatisfaisant.



**Fig.13** : Autocorrélogramme de  $\varepsilon_t$

La seconde est la statistique  $Q(r)$  suggérée par Ljung et Box (1978) qui doit être inférieure à la valeur critique CV de la distribution de Khi-deux. La formule de  $Q(r)$  est la suivante. Elle peut être directement calculée sous MATLAB.

$$Q(r) = n(n+2) \sum_{k=1}^m (n-k)^{-1} r_k(\varepsilon)^2 \quad (139)$$

$$r_k(\varepsilon) = \frac{\sum_{i=k+1}^n \varepsilon_i \varepsilon_{i-k}}{\sum_{i=1}^n \varepsilon_i^2} \quad (140)$$

$Q(r) = 14.5 < CV$  et  $CV=31.41 \rightarrow$  cette condition est vérifiée

Et vu que le corrélogramme de  $\varepsilon_t$  ne présente pas d'autocorrélation des résidus alors on peut adopter ce modèle SARIMA(1,0,1)(1,1,1) pour la prévision des moyennes journalières de l'irradiation globale.

#### IV. La prévision :

En développant notre modèle :

$$X_t(1 - \varphi_1 L)(1 - \varnothing L^{365}) = (1 - \theta_1 L)(1 - \vartheta L^{365})\varepsilon_t \quad (141)$$

On obtient :

$$X_t = \varphi_1 X_{t-1} + \varnothing X_{t-365} - \varphi_1 \varnothing X_{t-366} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \vartheta \varepsilon_{t-365} + \theta_1 \vartheta \varepsilon_{t-366} \quad (142)$$

Pour la prévision, on revient à notre série  $G_t$  tel que : ( $G_t = LN(E_t)$ ) ;

On a :

$$\left\{ \begin{array}{l} X_t = \Delta_{365} G_t = G_t - G_{t-365} \\ X_{t-1} = \Delta_{365} G_{t-1} = G_{t-1} - G_{t-366} \\ X_{t-2} = \Delta_{365} G_{t-2} = G_{t-2} - G_{t-367} \end{array} \right. \quad (143)$$

On replace l'équation (143) dans (142):

$$G_t = G_{t-365} + \varphi_1(G_{t-1} - G_{t-366}) + \varnothing(G_{t-365} - G_{t-720}) - \varphi_1 \varnothing(G_{t-366} - G_{t-721}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \vartheta \varepsilon_{t-365} + \theta_1 \vartheta \varepsilon_{t-366} \quad (144)$$

D'où :

$$G_t = \varphi_1 G_{t-1} + (1 + \varnothing)G_{t-365} - (\varphi_1 + \varnothing \varphi_1)G_{t-366} - \varnothing G_{t-720} + \varnothing \varphi_1 G_{t-721} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \vartheta \varepsilon_{t-365} + \theta_1 \vartheta \varepsilon_{t-366} \quad (145)$$

En tenant compte de l'origine de la prévision est à  $t$  et qui correspond à la date du 31 Mai 2009, la prévision commence à partir de  $t+1$  qui correspond à la date du 1/ Juin 2009. La prévision à l'horizon «  $h$  » ( $h \geq 1$ ) s'écrit comme suit :

$$\begin{aligned} \hat{G}_t^*(h) = & \varphi_1 G_{t+h-1} + (1 + \emptyset) G_{t+h-365} - (\varphi_1 + \emptyset \varphi_1) G_{t+h-366} - \emptyset G_{t+h-720} + \emptyset \varphi_1 G_{t+h-721} \\ & + \varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1} - \vartheta \varepsilon_{t+h-365} + \theta_1 \vartheta \varepsilon_{t+h-366} \end{aligned} \quad (146)$$

Tel que :

$$G_{t+j} = \begin{cases} G_{t+j} & j \leq 0 \\ \hat{G}_t(j) & j > 0 \end{cases} \quad (147)$$

$$\varepsilon_{t+j} = \begin{cases} \varepsilon_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases} \quad (148)$$

$$\varepsilon_t = X_t - \hat{X}_{t-1}(1) \quad (149)$$

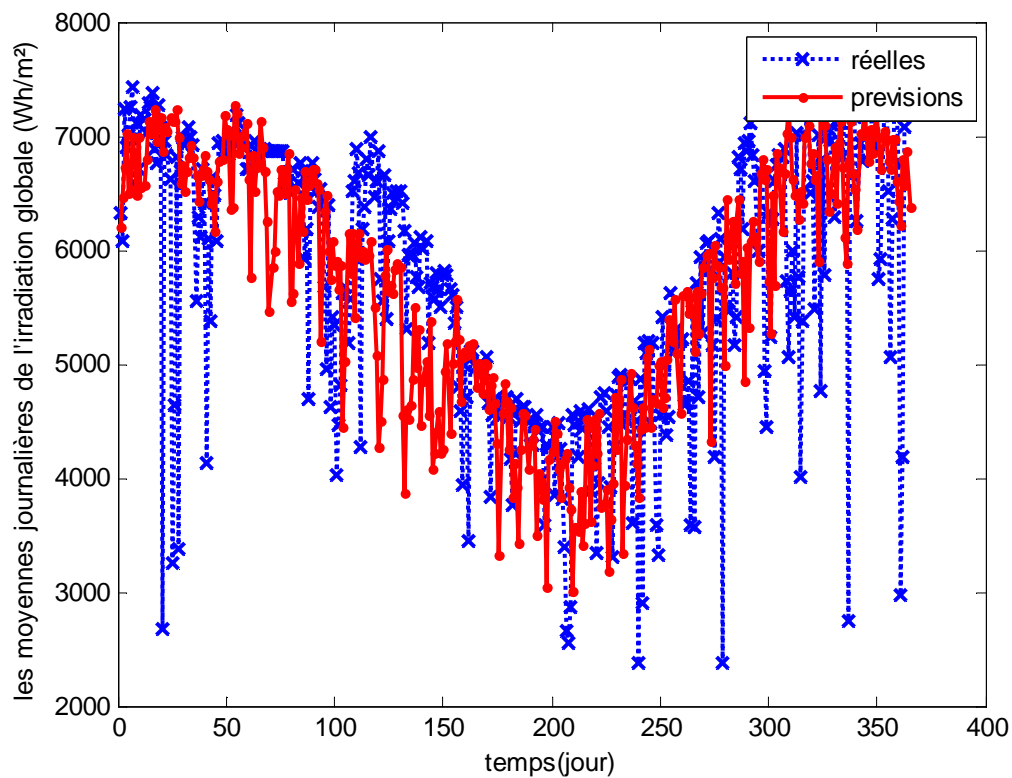
NB : comme l'ordre de l'équation est égal à 366, alors il nous faut 366 x 2 valeurs de la série temporelle pour faire de la prévision et ceci est vérifiée puisque nous avons  $T=1246$  qui est le nombre des observations restante après la différenciation.

#### IV.1 Résultats de la prévision des moyennes journalières de l'irradiation globale ;

Après avoir calculé les prévisions de la série  $G_t$ , on revient à notre série originale  $E_t$  par un simple calcul ;

$$E_t = e^{G_t} \quad (150)$$

La Fig.14 montre les données réelles des moyennes journalières de l'irradiation globale ( $\text{Wh/m}^2$ ) en fonction du temps ainsi que leur prévision calculées via notre modèle SARIMA(1,0,1)(1,1,1).



**Fig.14** : données de l'irradiation globale et ses prévisions du 1 Juin 2009 au 31 mai 2010

#### IV.2 Test de validation du modèle :

Pour évaluer les performances du modèle qui nous permet d'estimer et de prévoir la moyenne journalière de l'irradiation globale, des tests statistiques, qui sont communs pour tous les types de prévisions (LI Danny et Cheung Gary, 2005), tels que le **NRMSE** (Normal Root Mean Square Error), le **NMBE** (Normalized Mean Bias Error) et le **MPE** (Mean Percentage Error) sont calculés afin de comparer les valeurs mesurées à celles calculées via le modèle.

##### 1. Le pourcentage d'erreur moyen (MPE)

$$MPE = \frac{1}{N} \sum_{i=1}^N \left( \frac{G_{i,cal} - G_{i,mes}}{G_{i,mes}} \right) * 100 \quad (151)$$



2. L'erreur systématique moyenne (MBE) et L'erreur systématique moyenne normalisée (NMBE)

$$MBE = \frac{1}{N} \sum_1^N (G_{i,cal} - G_{i,mes}) \quad (152)$$

$$NMBE = \frac{MBE}{\frac{1}{N} \sum_{i=1}^N G_{i,mes}} * 100 \quad (153)$$

3. L'erreur quadratique moyenne (RMSE) et L'erreur quadratique moyenne normalisée (NRMSE)

$$RMSE = \left[ \frac{1}{N} \sum_1^N (G_{i,cal} - G_{i,mes})^2 \right]^{1/2} \quad (154)$$

$$NRMSE = \frac{RMSE}{\frac{1}{N} \sum_{i=1}^N G_{i,mes}} * 100 \quad (155)$$

Tel que  $G_{i,mes}$  et  $G_{i,cal}$  sont respectivement les valeurs mesurées et calculées de l'irradiation solaire globale à l'instant  $i$  et  $N$  est le nombre de valeurs mesurées (calculées).

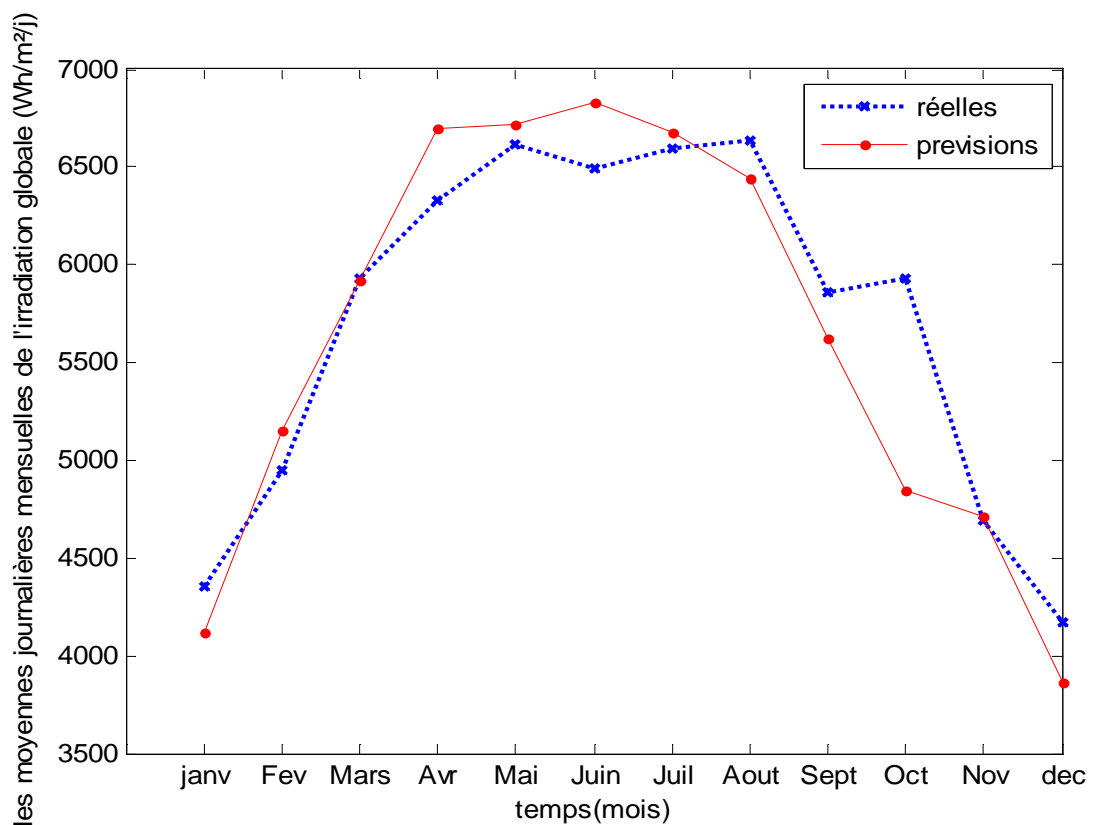
## V. Résultats et discussion :

Le tableau suivant présente les résultats de calculs de quelques tests statistiques qui nous permettront d'évaluer les performances de notre modèle SARIMA développée pour la prévision des moyennes journalières de l'irradiation globale.

RMSE KWh/m <sup>2</sup>	NRMSE %	MBE KWh/m <sup>2</sup>	NMBE %	MPE %	R <sup>2</sup>
<b>1.00</b>	<b>17.53</b>	<b>-0.08</b>	<b>-1.46</b>	<b>1.03</b>	<b>0.64</b>

**Tableau.4** : tests statistiques pour les moyennes journalières de l'irradiation globale

Pour pouvoir valider notre modèle, on cite à titre d'exemple les travaux de M.Koussa et al. [8] (Tableau.6) qui ont fait une étude comparative, basée sur des statistiques, des différents modèles empiriques existant dans la littérature servant à l'estimation de l'irradiation solaire et plus spécialement à Ghardaïa. Sachant qu'ils ont travaillé sur des moyennes journalières mensuelles de l'irradiation solaire. Afin de pouvoir comparer donc nos résultats des statistiques on procédera au calcul des moyennes journalières mensuelles de l'irradiation globale et leurs prévisions du mois de Juin 2009 au mois Mai 2010 calculées à partir des données et prévisions journalières Fig.15.



**Fig.15** : Les moyennes journalières mensuelles de l'irradiation globale reçus sur un plan horizontale à Ghardaïa et leurs prévisions.

Dont les tests statistiques sont représentés sur le tableau suivant :

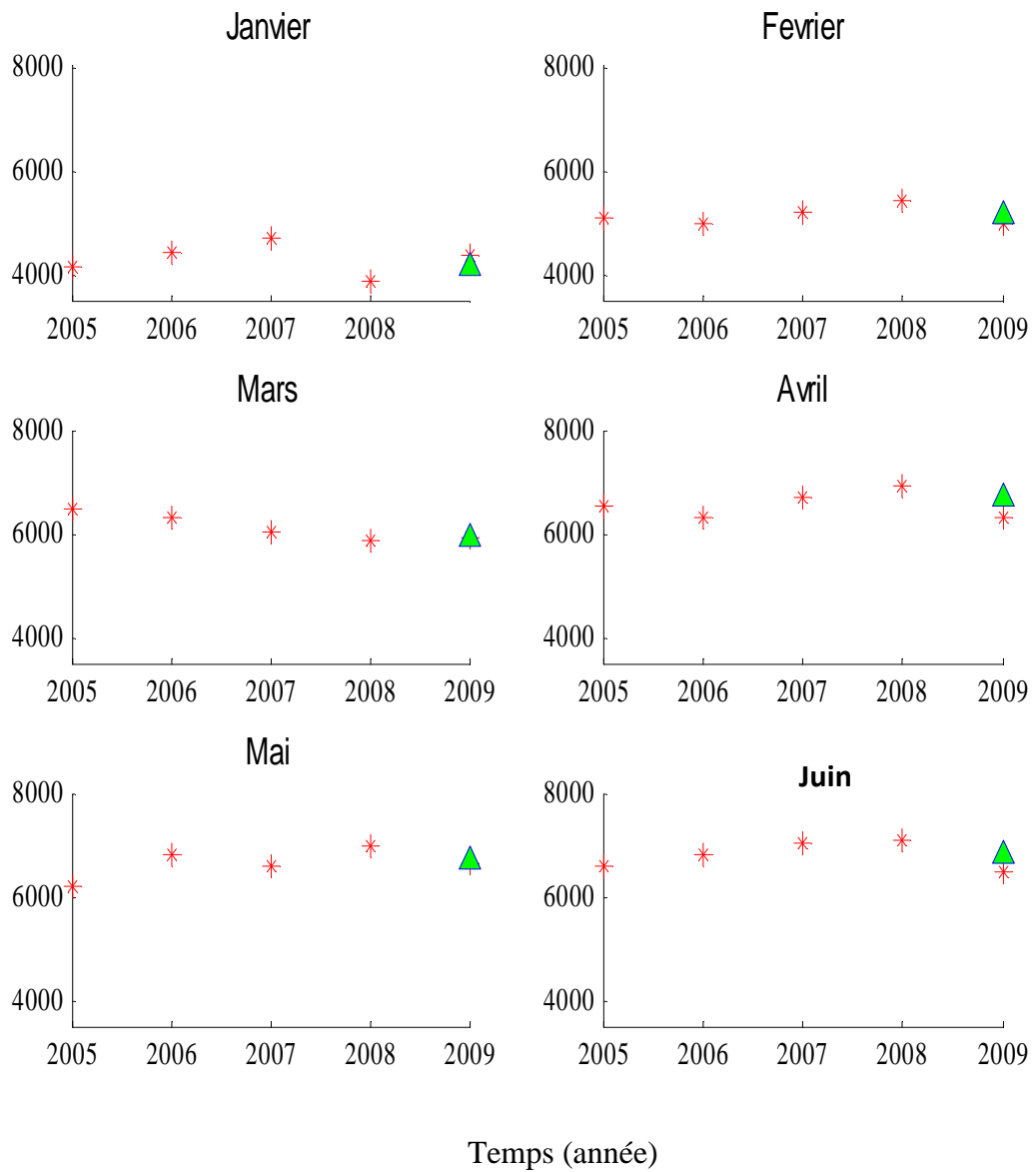
<b>RMSE KWh/m<sup>2</sup></b>	<b>NRMSE %</b>	<b>MBE KWh/m<sup>2</sup></b>	<b>NMBE %</b>	<b>MPE %</b>	<b>R<sup>2</sup></b>
<b>0.97</b>	<b>17.11</b>	<b>-0.08</b>	<b>-1.64</b>	<b>1.12</b>	<b>0.93</b>

**Tableau.5** : tests statistiques calculés pour les moyennes journalières mensuelles de l'irradiation globale

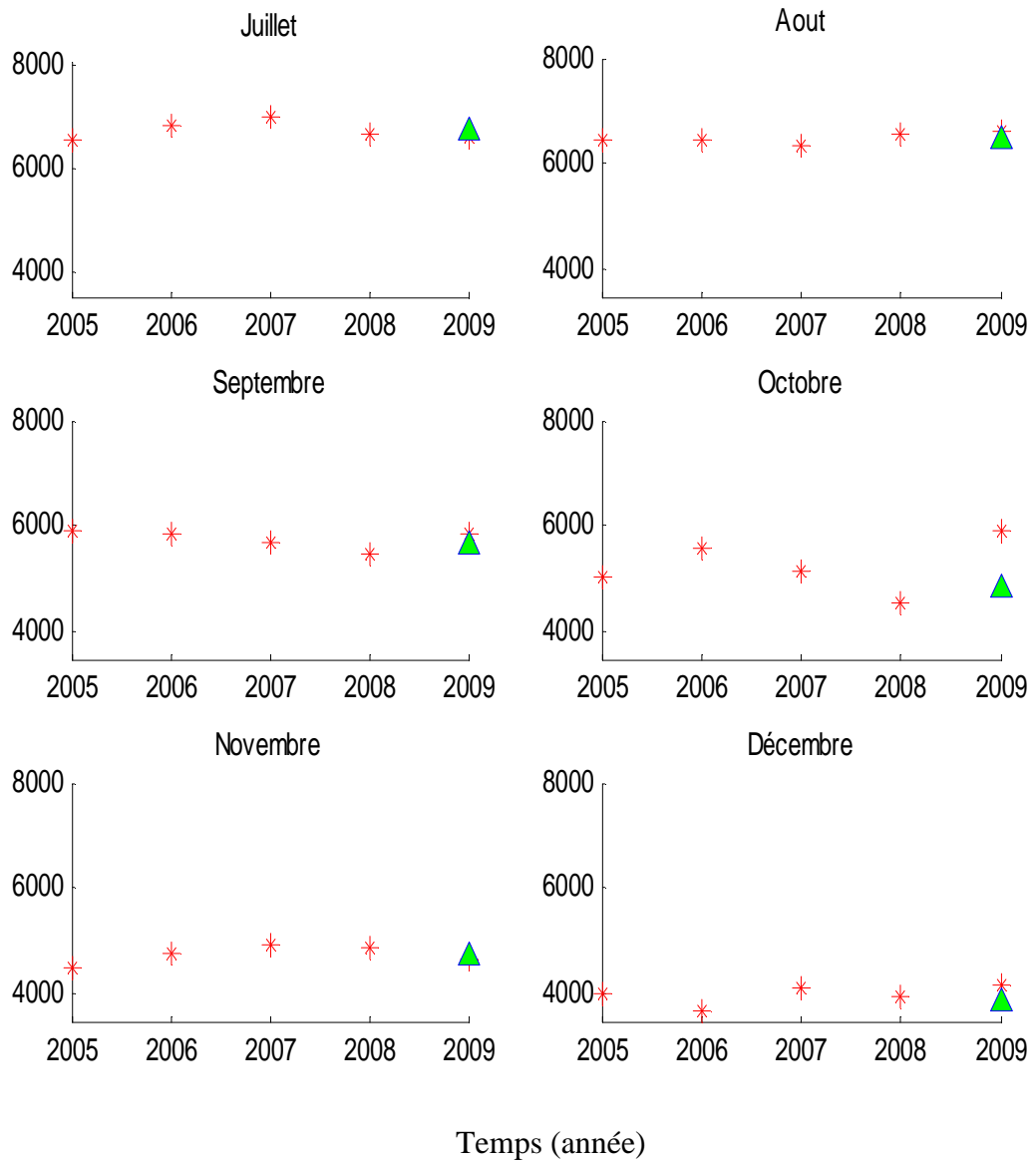
<b>Modèles</b>	<b>MPE (%)</b>	<b>MBE (KWH/m<sup>2</sup>)</b>	<b>RMSE (KWH/m<sup>2</sup>)</b>
Rietveld's	-3.22	-0.162	0.346
McCulloch's	-5.27	0.277	0.394
Hay's	2.94	0.16	0.337
Sayigh's	-10.3	-0.556	0.974
Reddy's	15.6	0.969	1.443
Swartman's	13.6	0.814	1.196
Swartman's	14.7	0.896	1.2
Garg's	0.4	0.02	0.218
Hussain's	-4.8	-0.266	0.349
Sambo's	-22.5	-0.952	1.012
<b>SARIMA</b>	<b>1.12</b>	<b>-0.08</b>	<b>0.97</b>

**Tableau .6** : testes statistiques de quelques modèles empiriques entrepris par M.Koussa [8] et du modèle SARIMA obtenu.

Dans les figures qui suivent (Fig.16, Fig.17), nous résumons les moyennes journalières mensuelles de chaque mois durant toutes les années étudiées ici (2005 à 2010) ainsi que les prévisions correspondantes de l'année 2009.



**Fig.16** : les moyennes journalières mensuelles des irradiations globales calculées pour les différentes années : 2005, 2006, 2007, 2008, 2009, et 2010.



**Fig.17 :** les moyennes journalières mensuelles des irradiations globales calculées pour les différentes années : 2005, 2006, 2007, 2008, 2009, et 2010.

**Conclusion 1 :**

Concernant les prévisions des moyennes journalières de l'irradiation globale, on peut dire qu'avec un MPE de 1.03% et un MBE de -0.08 KWh/m<sup>2</sup>, notre modèle est assez satisfaisant et valable pour la prévision

Pour ce qui est des prévisions des moyennes journalières mensuelles de l'irradiation globale calculées à partir des moyennes journalières (Fig.15), on peut faire quelques remarques concernant les écarts trouvés entre les données mesurées et celles calculées qui sont traduites par le fait que le site de Ghardaïa connaît des changements climatiques notamment des vents de sable fréquents durant les mois d'Avril, mai, Juin et Octobre ce qui perturbe le rayonnement au niveau du sol ainsi l'irradiation solaire. Et ces perturbations changent d'une année à une autre (Fig.16 et 17). L'important écart observé au mois d'octobre 2009 n'est pas un problème de prévision mais parce qu'il est caractérisé par des journées bien ensoleillées tout au long de ce mois-ci contrairement aux années précédentes (2005, 2006, 2007 et 2008) car en général la moyenne journalière mensuelle de l'irradiation globale pour le mois d'Octobre se situe entre 4500 et 5500 Wh/m<sup>2</sup>/jour alors qu'en 2009 elle dépasse les 5900 Wh/m<sup>2</sup>/jour.

Comparant maintenant nos résultats trouvés et ceux de nos collègues M.Koussa et al (Tableau.6), on conclut que nos tests statistiques concurrencent largement ceux des différents modèles empiriques employés, par ailleurs notre modèle SARIMA peut être validé et appliqué pour la prévision des irradiations globale journalière.

**VI. Autres techniques de modélisations ;**

Précédemment, nous avons considéré la série tel qu'elle est en lui appliquant un simple changement de variable (Logarithmique), maintenant nous allons essayer d'apporter d'autres changements à notre série afin d'apporter de meilleurs résultats si possible, pour cela nous avons choisi deux manières de procéder :

**➤ La première procédure est la suivante :**

En traçant la courbe de notre variable  $E_t$  en fonction du temps (Fig.8), on a observé qu'il y'avait déjà une saisonnalité apparente avec une tendance sinusoïdale dont la périodicité est de 365 jours, par ailleurs, on pourra décomposer notre signal  $E_t$  en deux composantes :

$$\text{Soit ; } E_t = U_t + Y_t$$

Tel que :  $U_t$  est la composante périodique que nous décomposerons en série de Fourier

Et  $Y_t$  est la composante stochastique que nous modéliserons avec le modèle stochastique.

### *Analyse harmonique :*

L'expression générale de la décomposition en série de Fourier est :

$$U_t = \beta_0 + \sum_{i=1}^k (A_i \cos \omega_i t + B_i \sin \omega_i t) \quad (156)$$

Avec  $t$  est le jour de l'année,  $k$  est le nombre d'harmoniques,  $A_i$  et  $B_i$  sont les coefficients de Fourier.

$$\hat{\beta}_0 = \frac{1}{n} \sum_{t=1}^n Y_t \quad (157)$$

$$\hat{A}_i = \frac{2}{n} \sum_{t=1}^n Y_t \cos \omega_i t \quad (158)$$

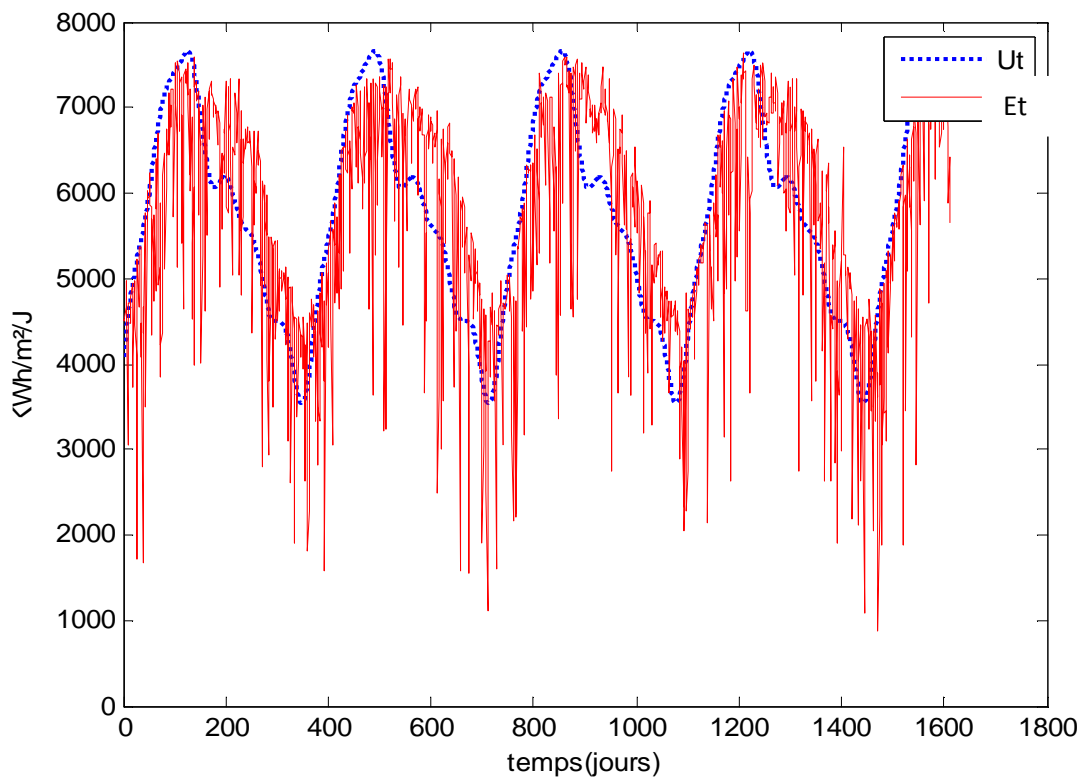
$$\hat{B}_i = \frac{2}{n} \sum_{t=1}^n Y_t \sin \omega_i t \quad (159)$$

Après avoir estimé les coefficients  $A_i$  et  $B_i$  de Fourier, l'estimation de la composante tendancielle de l'irradiation globale journalière est comme suit :

$$\begin{aligned} \hat{U}_t = & 5.711 - 1.068 \cos \frac{2\pi t}{365} - 0.532 \cos \frac{4\pi t}{365} - 0.125 \cos \frac{8\pi t}{365} \\ & + 1.179 \sin \frac{2\pi t}{365} + 0.118 \sin \frac{8\pi t}{365} + 0.165 \sin \frac{12\pi t}{365} \end{aligned} \quad (160)$$

Avec un coefficient de corrélation  $R^2 = 65\%$ .

La Fig.18 montre le tracé des données de l'irradiation globale ainsi que la courbe de tendance sinusoidale  $U_t$



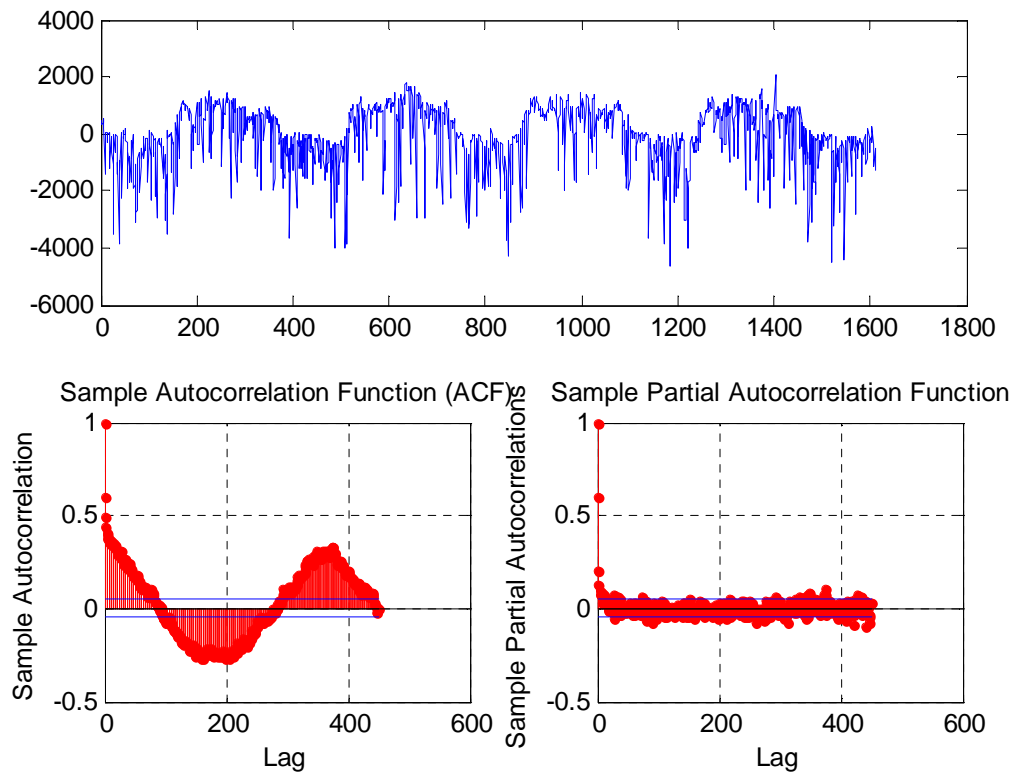
**Fig.18** : données de l'irradiation globale ainsi que la courbe de tendance sinusoïdale  $U_t$

### **Modélisation stochastique de $Y_t$ :**

Après avoir retranché la composante tendancielle de notre signal  $E_t$ , on obtient la composante stochastique  $Y_t$  que nous modéliserons avec le modèle SARIMA ( Saisonal Autoregressive Integrated Moving Average)

La Fig.19 présente la série  $Y_t$  ainsi que ses autocorrélogramme et autocorrélogramme partiel, qui présente des autocorrélations significatives avec encor une saisonnalité très apparente de 365 jours, ce qui nous amène à différencier la série  $Y_t$  pour la rendre stationnaire.



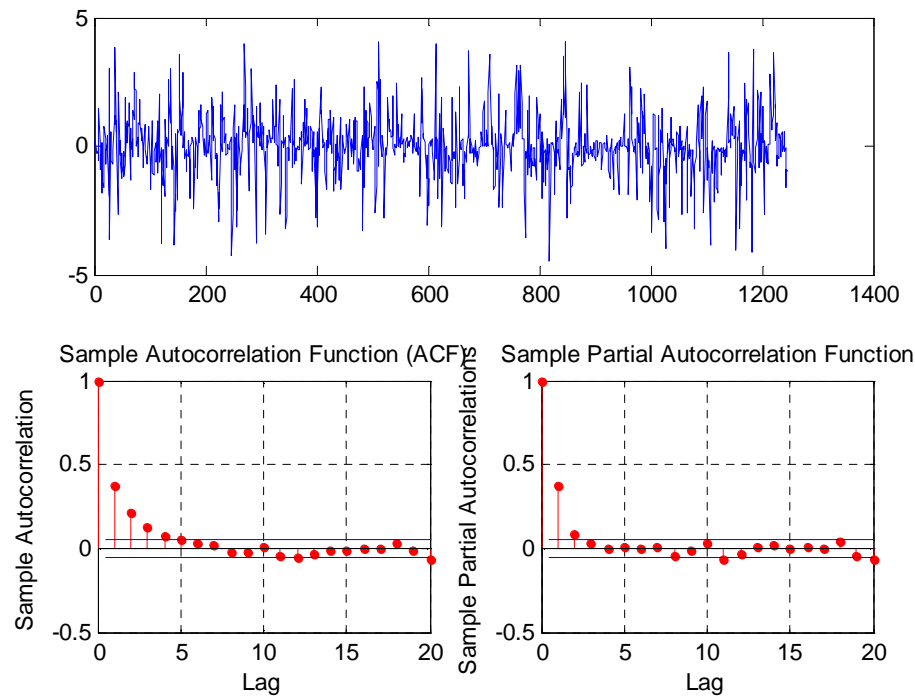


**Fig. 19.** l'autocorrélogramme et l'autocorrélogramme partiel de  $Y_t$

On posera :

$$W_t = (1 - B^{365})Y_t \tag{161}$$

Dont les autocorrélogrammes sont présentés dans la Fig.20.



**Fig..20 :**  $W_t$  et ses autocorrélogrammes

On voit bien que la différentiation  $\Delta_{365}Y_t$  réduit nettement les autocorrélations et élimine la saisonnalité, ainsi notre série temporelle  $W_t$  devient stationnaire.

Et c'est à partir de ces autocorrélogrammes qu'on déduira empiriquement le degré maximale des polynômes  $\varphi(B)$   $\vartheta(B^s)$   $\theta(B)$   $\vartheta(B^s)$ , mais reste à tester tous les cas possibles lors de l'estimation des paramètres en se basant sur quelques tests statistiques afin d'arriver au modèle approprié comme nous l'avons déjà mentionné ultérieurement.

### *L'estimation :*

La forme général de notre modèle en tenant compte de tous les paramètres possibles a estimer est comme suit : SARIMA(2,0,3)(0,1,1) ;

$$(1 - B^{365})\varphi(B)Y_t = \theta(B)(1 - \vartheta B^{365})\varepsilon_t \quad (162)$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3$$

Pour estimer le vecteur des paramètres :  $\theta = (\varphi_1, \varphi_2, \theta_1, \theta_2, \theta_3, \vartheta, \text{ et } \sigma_\varepsilon^2)$ , On pose :  $W_t = (1 - B^{365})Y_t$

D'où ;

$$W_t(1 - \varphi_1 B - \varphi_2 B^2) = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)(1 - \vartheta B^{365})\varepsilon_t \quad (163)$$

Pour les étapes d'estimation et de prévision nous suivrons les mêmes que celles déjà vu avec le premier modèle, et nous résumerons les résultats de l'estimations des différent paramètres dans le tableau qui suit :

	MVS	t-statistic	AIC	BIC
$\varphi_1 \theta_1 \vartheta$	-1724.45	0.802	54.89	70.28
$\varphi_1 \theta_1 \varnothing \vartheta$	-1723.57	0.931	55.15	75.66
$\varphi_1 \theta_1 \varnothing$	-1766.12	0.011	138.24	153.62
$\varphi_1 \varnothing$	-1771.22	0.013	146.45	156.70
$\varphi_1 \varnothing \vartheta$	-1728.57	0.911	63.14	78.53
$\varphi_1 \vartheta$	-1729.50	0.778	63.00	73.25
$\varphi_1 \theta_1 \theta_2 \varnothing$	-1766.11	0.013	140.22	160.73
$\varphi_1 \varphi_2 \varnothing$	-1766.50	0.001	139.00	154.38
$\varphi_1 \varphi_2 \theta_1 \varnothing$	-1766.11	0.012	140.22	160.73
$\varphi_1 \varphi_2 \theta_1 \theta_2 \vartheta$	-1724.36	0.814	58.73	84.37
$\varphi_1 \theta_1 \theta_2 \vartheta$	<b>-1724.44</b>	<b>0.805</b>	56.88	77.39
$\varphi_1 \varphi_2 \theta_1 \vartheta$	-1724.44	0.804	56.88	77.39
$\varphi_1 \theta_1 \theta_2 \theta_3 \vartheta$	-1724.42	0.800	58.84	84.48
$\varphi_1 \theta_2 \theta_3 \vartheta$	-1724.86	0.776	57.71	78.22
$\varphi_1 \theta_1 \theta_2 \varnothing \vartheta$	-1723.57	0.933	57.13	82.77
$\varphi_2 \theta_1 \theta_2 \vartheta$	-1724.38	0.814	56.75	77.26
$\varphi_1 \theta_1 \theta_3 \vartheta$	-1724.42	0.801	56.84	77.35
$\varphi_1 \theta_1 \theta_2 \vartheta$	-1726.47	0.774	58.94	74.32
$\theta_1 \theta_2 \vartheta$	-1732.43	0.755	70.85	86.24

**Tableau.7** : critères de sélection du modèle

Vu que le minimum de AIC et BIC est observée pour le modèle SARIMA(1,0,1)(0,1,1) (la première ligne) ou la valeur du maximum de vraisemblance est plus élevée comparé au modèle SARIMA(1,0,3)(0,1,1) ou la t-statistique est inférieure, aussi, la probabilité d'avoir des résiduels Bruit Blanc est plus importante dans ce modèle. Ceci nous mène à conclure que le modèle SARIMA(1,0,1)(0,1,1) est le meilleur modèle que nous adopterons pour la prévision et qui s'écrit comme suit avec les paramètres estimés :

$$W_t(1 - 0.575L) = (1 - 0.241L)(1 - 0.607L^{365})\varepsilon_t \quad (164)$$

Avec :  $\varepsilon_t \sim N(0, 0.93)$

**La prévision (selon Box Jenkins) :**

Pour la prévision, on revient à notre série  $Y_t$  Tel que :

$$\left\{ \begin{array}{l} W_t = \Delta_{365} Y_t = Y_t - Y_{t-365} \\ W_{t-1} = \Delta_{365} Y_{t-1} = Y_{t-1} - Y_{t-366} \end{array} \right. \quad (165)$$

On remplace (164) dans (163) :

$$Y_t = \varphi_1 Y_{t-1} + Y_{t-365} - \varphi_1 Y_{t-366} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \vartheta \varepsilon_{t-365} + \vartheta \theta_1 \varepsilon_{t-366} \quad (166)$$

La prévision à l'horizon « h » selon Box Jenkins s'écrit :

$$\begin{aligned} Y_{t+h} &= \varphi_1 Y_{t+h-1} + Y_{t+h-365} - \varphi_1 Y_{t+h-366} \\ &+ \varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1} - \vartheta \varepsilon_{t+h-365} + \vartheta \theta_1 \varepsilon_{t+h-366} \end{aligned} \quad (167)$$

On procédera de la manière que précédemment.

Par exemple, pour prévoir la valeur de  $\hat{Y}_t(5) = Y_{t+5}$ , c.à.d. la cinquième valeur à partir de l'origine t, on fait comme suit :

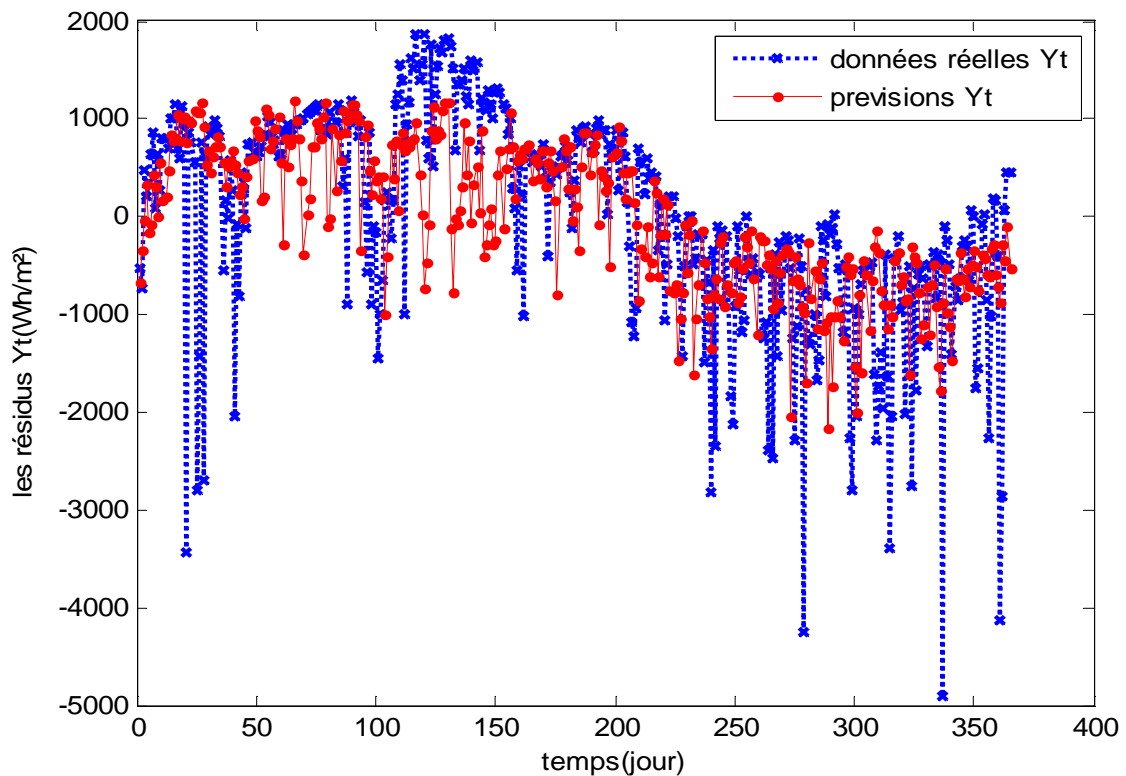
$$\begin{aligned}
 Y_{t+5} &= \varphi_1 Y_{t+4} + Y_{t-360} - \varphi_1 Y_{t-361} + \varepsilon_{t+5} \\
 &\quad - \theta_1 \varepsilon_{t+4} - \vartheta \varepsilon_{t-360} + \vartheta \theta_1 \varepsilon_{t-361}
 \end{aligned}
 \tag{168}$$

D'ou:

$$\begin{aligned}
 \hat{Y}_t(5) &= \varphi_1 \hat{Y}_t(4) + Y_{t-360} - \varphi_1 Y_{t-361} - \vartheta (Y_{t-360} - \hat{Y}_{t-361}) \\
 &\quad + \vartheta \theta_1 (Y_{t-361} - \hat{Y}_{t-362})
 \end{aligned}
 \tag{169}$$

Pour le calcul de  $\hat{Y}_{t-361}$  et  $\hat{Y}_{t-362}$ , on se réfère toujours à l'équation (166).

Dont les résultats de la prévision de  $Y_t$  sont montrés sur la Fig.21 : du 1 Juin 2009 au 31 Mai 2010.

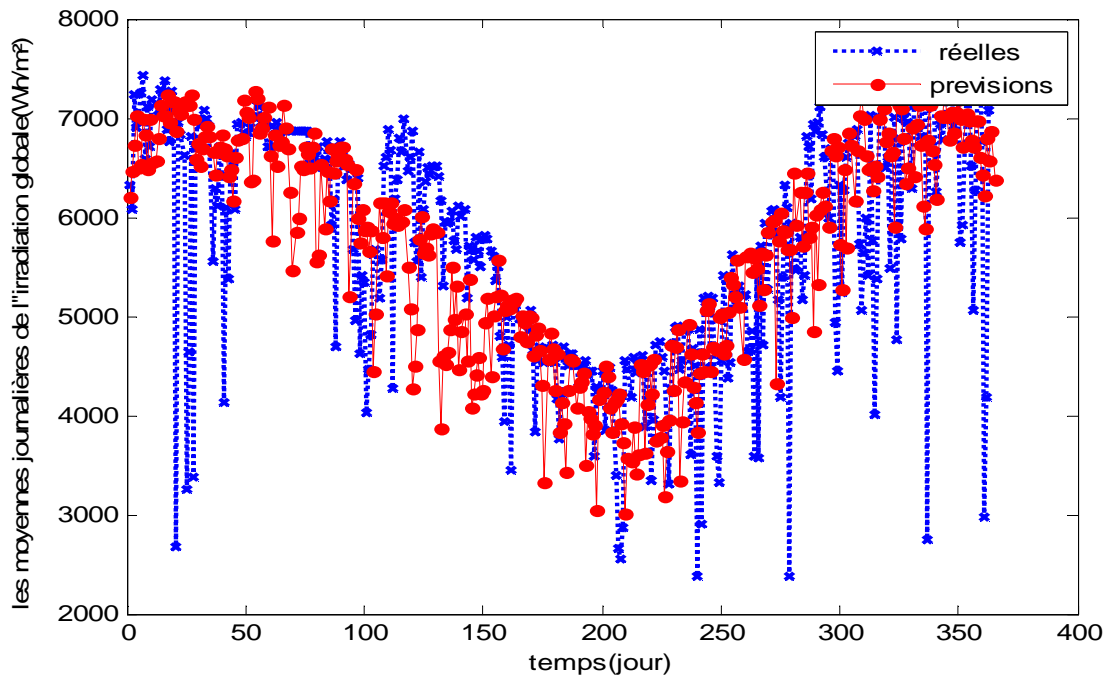


**Fig.21** : Données des résidus  $Y_t$  et ses prévisions

Avec un coefficient de corrélation de 0.54

Pour calculer les prévisions de notre série  $E_t$ , on effectuera une somme de  $Y_t$  et de  $U_t$  ie, la composante en série de Fourier et les résidus stochastiques.

Nous aboutirons aux résultats suivant : (Fig.22) avec un coefficient de corrélation de 66%.



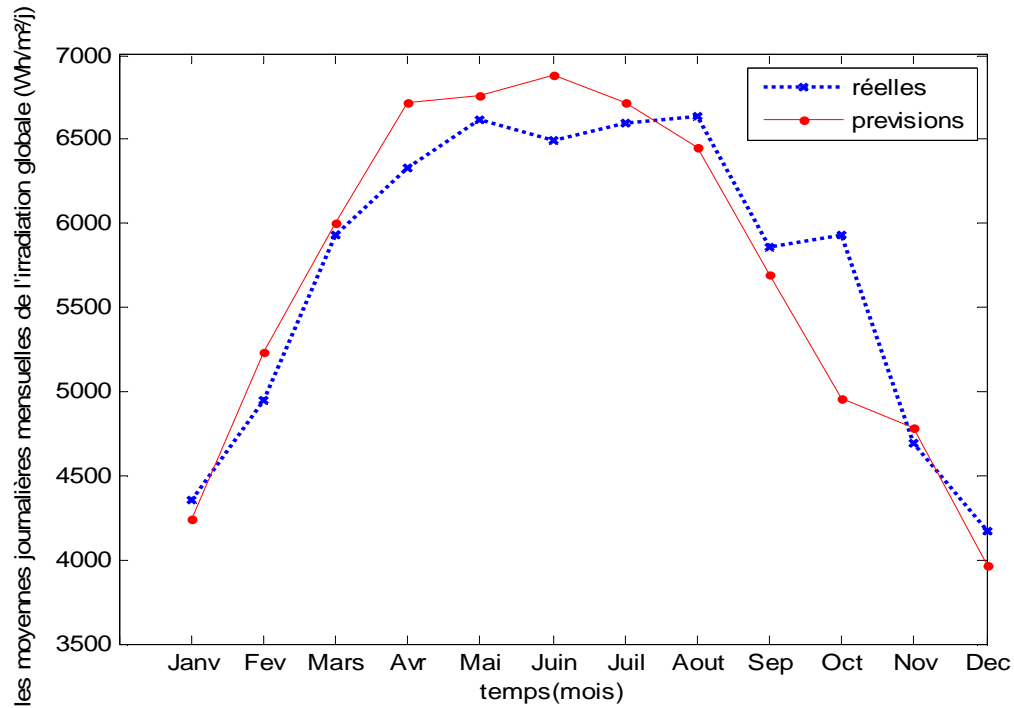
**Fig. 22** : les données de l'irradiation globale et ses prévisions

Avec les différents tests statistiques que nous résumerons dans le tableau suivant :

RMSE KWh/m <sup>2</sup>	NRMSE %	MBE KWh/m <sup>2</sup>	NMBE %	MPE %	R <sup>2</sup>
0.95	16.64	-0.02	-0.32	2.29	0.66

**Tableau.8** : tests statistiques pour les moyennes journalières de l'irradiation globale

Pour ce qui est des calculs des moyennes journalières mensuelles de l'irradiation globale et leurs prévisions du mois de Juin 2009 au mois Mai 2010 calculées à partir des données et prévisions journalières : voir Fig.23.



**Fig.23 :** Les moyennes journalières mensuelles de l’irradiation globale reçus sur un plan horizontale à Ghardaïa et leurs prévisions.

Nous présentons dans le tableau suivant quelques tests statistique calculées pour les moyennes journalières mensuelles :

<b>RMSE KWh/m<sup>2</sup></b>	<b>NRMSE %</b>	<b>MBE KWh/m<sup>2</sup></b>	<b>NMBE %</b>	<b>MPE %</b>	<b>R<sup>2</sup></b>
<b>0.91</b>	<b>16.07</b>	<b>-0.01</b>	<b>-0.40</b>	<b>2.38</b>	<b>0.94</b>

**Tableau.9 :** tests statistiques calculés pour les moyennes journalières mensuelles de l’irradiation globale

➤ *La seconde procédure envisagée est la suivante :*

Soit  $G_0$  l’irradiation globale hors atmosphère reçu sur un plan horizontale par rapport au site donné. (vu au chapitre 1)

$$G_0 = \left( \frac{24I_0}{\pi} \right) \left( 1 + 0.033 \cos \left( \frac{360N}{365} \right) \right) \cdot (\cos \lambda \cos \delta \sin \omega + 2\pi \omega \sin \lambda \sin \delta / 360) \quad (170)$$

Tel que :

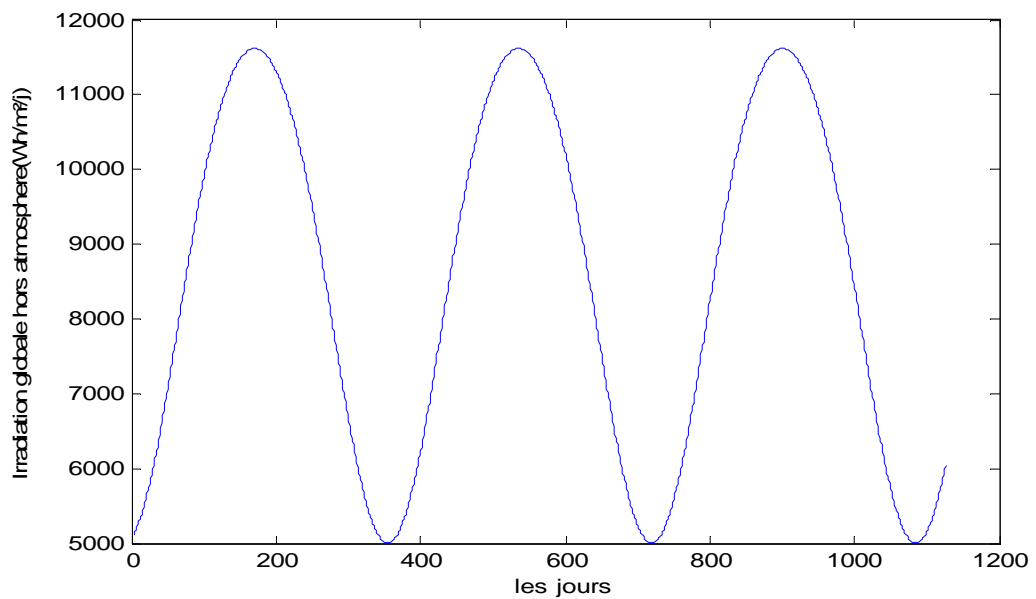
$I_0$  : est la constante solaire,  $I_0 = 1367 \frac{W}{m^2}$

$N$  : est le numéro du jour à partir du 1<sup>er</sup> Janvier (on considère 365 jours dans l'année)

$\lambda$  : est la latitude du lieu, pour Ghardaïa :  $\lambda = 32.23'N$

$\delta$  : est la déclinaison du soleil :

Tous les angles sont en radians.



**Fig.24** : l'irradiation globale hors atmosphère  $G_0$  par rapport au site de Ghardaïa

Cette procédure consiste à retrancher les données de l'irradiations globale  $E_t$  que nous avons mesuré de celles hors atmosphère  $G_0$  pour obtenir un résidu que nous appelons  $G_r$  (**Fig.25**) et que nous essayerons de modéliser avec la méthode de Box Jenkins.

$$\text{On a :} \quad G_0 = E_t + G_r \quad (171)$$



D'ou :  $G_r = G_0 - E_t$  (172)

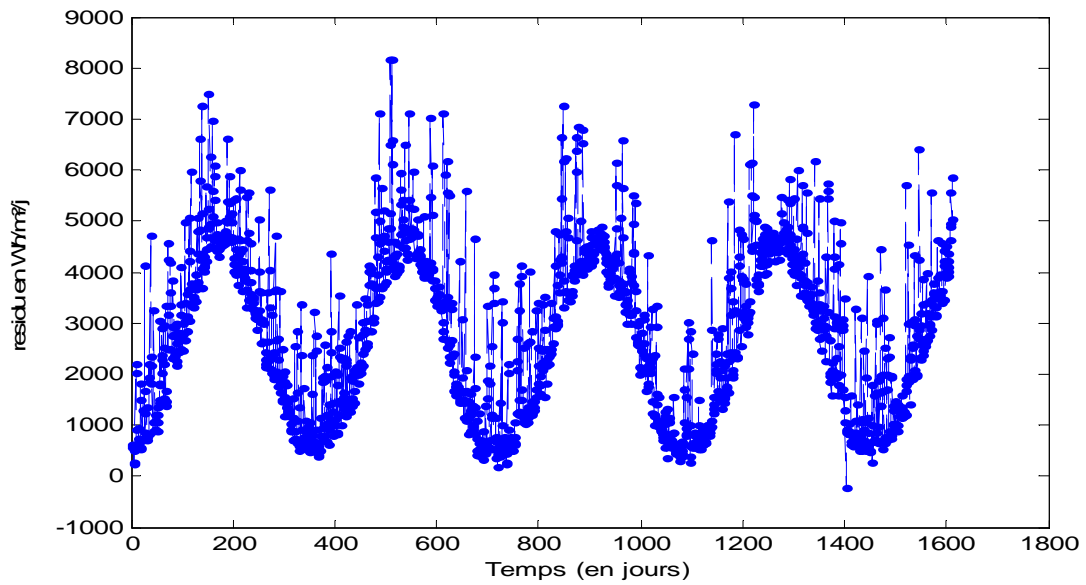


Fig.25 : les résidus  $G_r$  (en  $WH/m^2$ )

Ces résidus  $G_r$  présente aussi une saisonnalité de 365 jours, ce qui fait, nous procéderons à la différenciation afin de stationnariser la série, en multipliant par le filtre linéaire  $(1 - L^{365})$ .

La Fig.26 présente la série ainsi différenciée et ses autocorrélogramme ; simple et partiel.

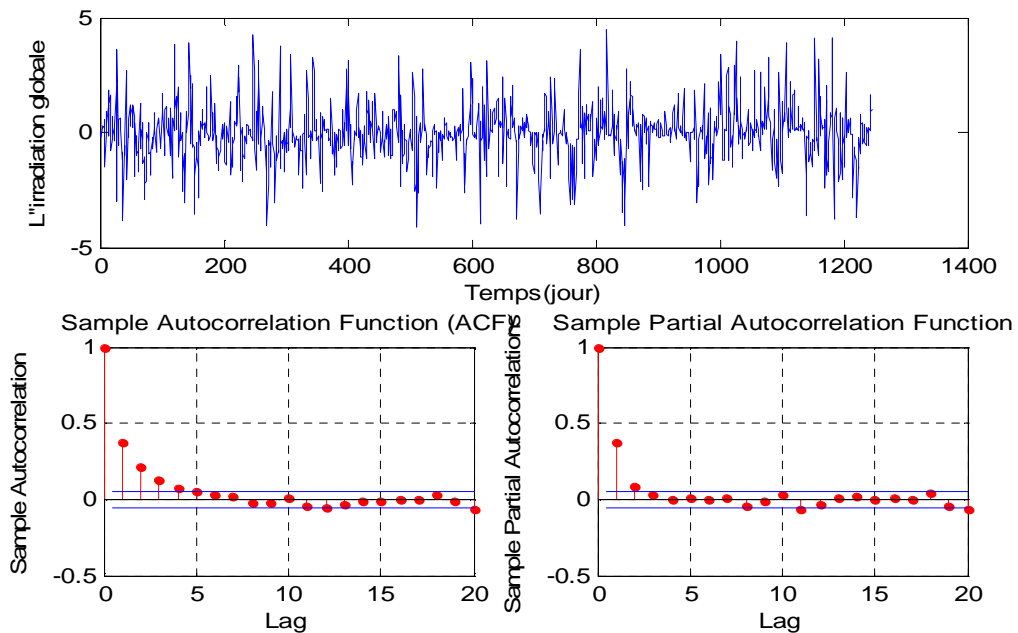


Fig.26 : les résidus  $G_r$  et ses corrélogrammes

Ces autocorrélogramme ressemblent beaucoup à ceux de la série  $Y_r$  étudiée précédemment, ce qui fait on suivra les même étapes pour la modélisation et la prévision.

Le tableau suivant montre les différents tests statistiques calculés lors de l'estimation des paramètres.

Hors atmo	MVS	t-statistic	AIC	BIC
$\varphi_1 \theta_1 \vartheta$	-1724.45	0.80	54.89	70.28
$\varphi_1 \theta_1 \emptyset \vartheta$	-1723.57	0.93	55.15	75.66
$\varphi_1 \emptyset \vartheta$	-1728.57	0.91	63.14	78.53
$\varphi_1 \vartheta$	-1729.50	0.77	63.00	73.25
$\varphi_1 \theta_1 \theta_2 \emptyset$	-1766.11	0.01	140.22	160.73
$\varphi_1 \varphi_2 \emptyset$	-1766.50	0.00	139.00	154.38
$\varphi_1 \varphi_2 \theta_1 \emptyset$	-1766.11	0.01	140.22	160.73
$\varphi_1 \varphi_2 \theta_1 \theta_2 \vartheta$	-1724.44	0.80	58.88	84.52
$\varphi_1 \theta_1 \theta_2 \vartheta$	-1724.44	0.80	56.88	77.39
$\varphi_1 \varphi_2 \theta_1 \vartheta$	-1724.44	0.80	56.88	77.39
$\varphi_1 \theta_1 \theta_2 \theta_3 \vartheta$	-1724.42	0.80	58.84	84.48
$\varphi_1 \theta_2 \theta_3 \vartheta$	-1724.86	0.78	57.71	78.22

**Tableau.10** : critères de sélection du modèle

Ainsi, on peut conclure que le modèle SARIMA(1,0,1)(0,1,1) est choisi pour la prévision.

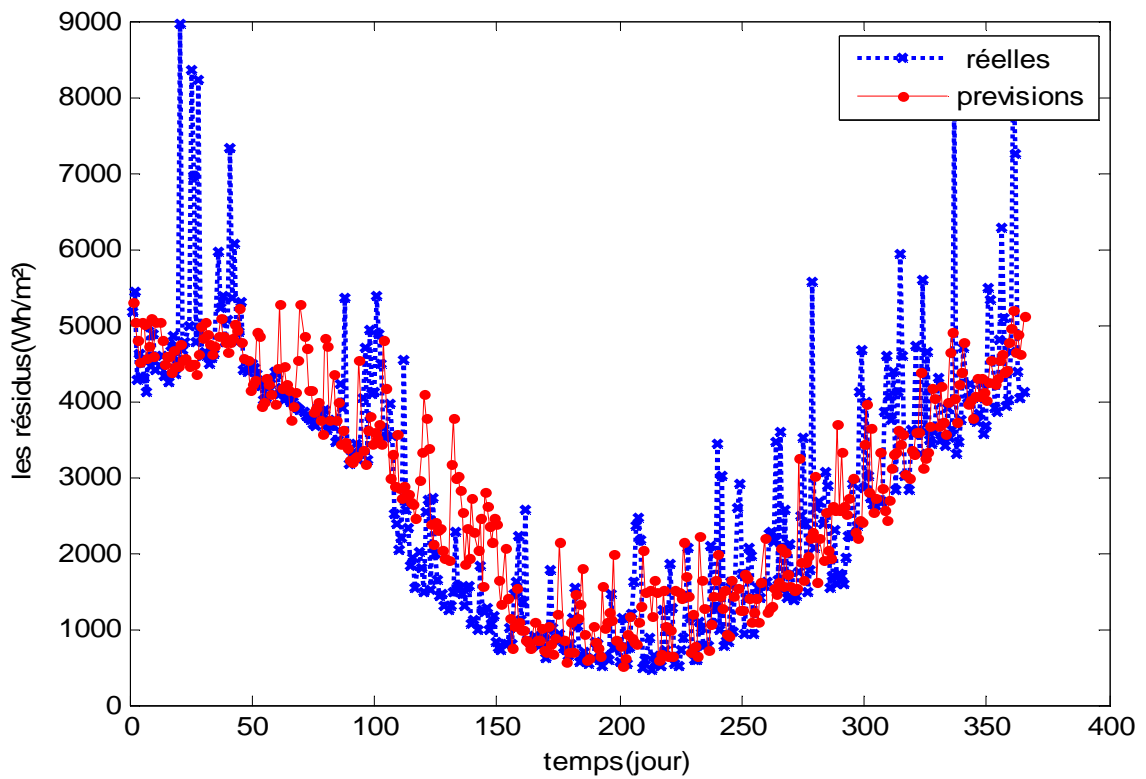
Qui s'écrit :

$$W_t(1 - 0.575L) = (1 - 0.241L)(1 - 0.601L^{365})\varepsilon_t \quad (173)$$

Avec :  $\varepsilon_t \sim N(0, 0.93)$

Et  $W_t = (1 - L^{365})G_r$

Les résultats de la prévision sont montrés sur la Fig.27 :



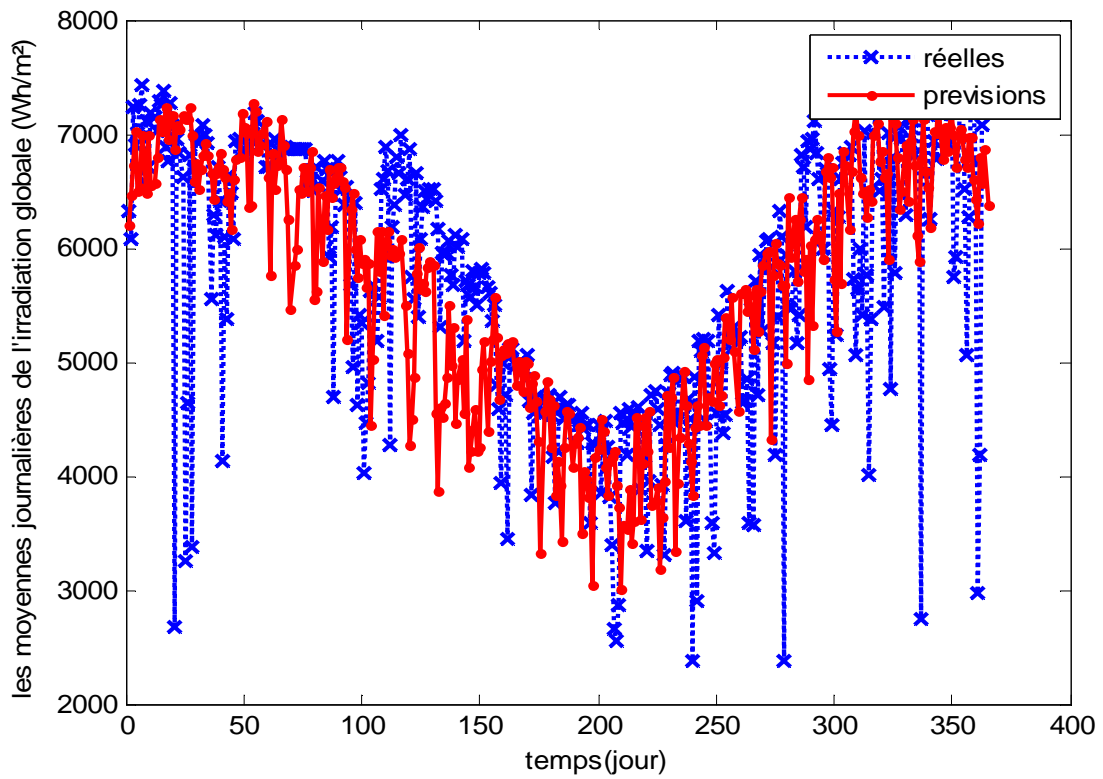
**Fig.27 :** Données des résidus  $G_r$  ainsi que leurs prévisions via le modèle SARIMA (1,0,1)(0,1,1)

Avec un coefficient de corrélation de 0.83

Pour calculer les prévisions de notre série  $E_t$  , on procédera à la différence suivante :

$$E_t = G_0 - G_r \quad (174)$$

Nous aboutirons aux résultats suivant : (Fig.28) avec un coefficient de corrélation de 66%.



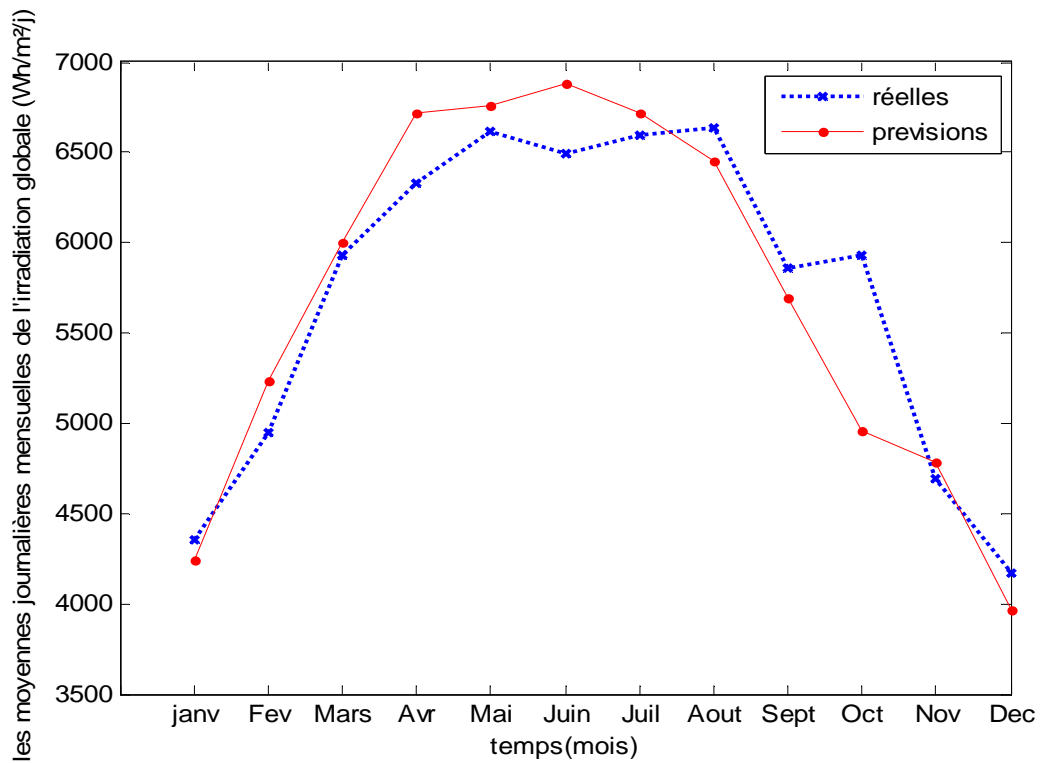
**Fig.28** : les moyennes journalières de l’irradiation globale ainsi que leurs prévisions

Les tests statistiques calculés sur les données moyennes journalières sont montrés sur ce tableau :

<b>RMSE KWh/m<sup>2</sup></b>	<b>NRMSE %</b>	<b>MBE KWh/m<sup>2</sup></b>	<b>NMBE %</b>	<b>MPE %</b>	<b>R<sup>2</sup></b>
<b>0.95</b>	<b>16.63</b>	<b>-0.01</b>	<b>-0.32</b>	<b>1.29</b>	<b>0.66</b>

**Tableau.11** : tests statistiques calculés pour les moyennes journalières de l’irradiation globale

Pour ce qui est des calculs des moyennes journalières mensuelles calculés à partir des prévisions journalières, nous avons le graphe suivant (Fig.29) :



**Fig.29** : les moyennes journalières mensuelles de l’irradiation globale et ses prévisions

<b>RMSE KWh/m<sup>2</sup></b>	<b>NRMSE %</b>	<b>MBE KWh/m<sup>2</sup></b>	<b>NMBE %</b>	<b>MPE %</b>	<b>R<sup>2</sup></b>
<b>0.91</b>	<b>16.07</b>	<b>-0.01</b>	<b>-0.40</b>	<b>2.37</b>	<b>0.94</b>

**Tableau.12** : tests statistiques calculés pour les moyennes journalières mensuelles de l’irradiation globale

**Conclusion 2 :**

Ces deux procédures n’apportent pas un grand changement, mais on peut remarquer qu’il ya une légère amélioration du coefficient de corrélation qui était de 0.64 et qui passe à 0.66 par rapports au moyennes journalières de l’irradiation globale.

Pour ce qui est des moyennes journalières mensuelles, le coefficient de corrélation passe de 93% à 94% mais avec un meilleur MBE de -0.01 alors qu’il était à -0.08.

## Conclusion générale :

Dans ce travail, nous avons fait le tour sur le principe d'analyse des séries temporelles dans le cas général et sur la méthodologie de Box-Jenkins qui consiste à développer des modèles stochastiques de type SARIMA (Seasonal autoregressif Integrated Moving Average) qui peuvent être adaptés à différentes séries temporelles dans plusieurs domaines d'application.

A travers cette techniques de modélisation, on a pu élaborer un modèle de prévision SARIMA(1,0,1)(1,1,1) des moyennes journalières de l'irradiation globale reçu sur un plan horizontal sur le site de Ghardaïa. Des différents tests statistiques comme le RMSE, NRMSE, MBE, NMBE, MPE et la t-statistique ont été calculés afin de valider notre modèle.

Par la suite, on a apporté des changements à notre séries originale dans le but d'améliorer le modèle de prévision ; en premier on a décomposé notre série en coefficients de Fourier et modéliser le résidus stochastique par le modèle SARIMA(1,0,1)(0,1,1). Le second changement consiste à modéliser un autre résidu qui est issus de la différence des irradiations journalières hors atmosphère et celles au sol in situ.

## **Perspectives :**

Développer des modèles hybrides a base de ces modèles SARIMA et des techniques de l'intelligence artificiel afin d'améliorer au mieux le modèle de prévision des données de l'irradiation solaire in situ.

Etendre cette prévision du gisement solaire à travers le territoire national afin d'élaborer une cartographie vu le manque de stations radiométriques existants.

## Références Bibliographiques :

- [1] **Taric de Villers.** (Energie Nouvelle et Environnement) "Guide des Energies Renouvelables". [www.ulb.ac.be/students/desge/cours](http://www.ulb.ac.be/students/desge/cours)
- [2] **Sayigh AAM.** Solar energy engineering. London: Academic Press, INC LTD; 1977.
- [3] **Reddy SJ.** An empirical method for the estimation of total solar radiation. *Solar Energy.* 1971;13:289–94.
- [4] **Garg HP, Garg SN.** Prediction of global solar radiation from bright sunshine hours and other meteorological data. *Energy Convers Manage* 1983;23:113–8.
- [5] **Rietveld MR.** A new method for estimating the regression coefficients in the formula relating solar radiation to sunshine. *Agricult Meteorol* 1978;19:243–52.
- [6] **Sambo AS.** Empirical models for the correlation of global solar radiation with meteorological data for northern Nigeria. *Solar Wind Technol* 1986;3:89–93.
- [7] **Liu BYH, Jordan RC.** The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy* 1960;4:1–19
- [8] **Collares-Pereira Manuel, Rabl Ari.** The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values. *Solar Energy* 1979;22:155–64
- [9] **Bekkouche Sidi Mohammed El Amine.** Modélisation du Comportement Thermique de Quelques Dispositifs Solaires. Thèse de doctorat, l'Université Abou-Bakr Belkaïd – Tlemcen, 2008-2009.
- [10] **Mefti et Bouroubi.** Estimation et Cartographie de la Composante Globale du Rayonnement Solaire. *Rev. Energ. Ren. : Valorisation* (1999) 219-224
- [11] **M. Koussa , A. Malek, M. Haddadi.** Statistical comparison of monthly mean hourly and daily diffuse and global solar irradiation models and a Simulink program development for various Algerian climates, *Energy Conversion and Management* 50 (2009) 1227–1235.
- [12] **S. Benkaciali, K. Gairaa.** Modèles d'estimations du rayonnement global incident sur un plan incliné. International Renewable Energy Congress, November 5-7, 2009 - Sousse Tunisia
- [13] **M.Chegaar, A.Chibani.** Global solar radiation estimation in Algeria, *Energy Conversion and Management* 42 (2001) 967-973.
- [14] **A. Zeroual, M. Ankrim and A. J. Wilkinson.** Stochastic modelling of daily global solar radiation measured in Marrakesh Morocco. *Renewable Energy*, Vol. 6, No. 7, pp. 787 793, 1995.
- [15] **Kadri Yurekli, Huseyin Simsek, Bilal Cemek, Sedat Karaman** Simulating. Climatic variables by using stochastic approach. *Building and Environment* 42 (2007) 3493–3499
- [16] **M. Yusof Sulaiman, W. M. Hlaing, mahdi abd. Wahab and Z. Abidin Sulaiman.** Analysis of residuals in daily solar radiation time series. *Renewable Energy*, Vol. 11, No. 1, pp. 9%105, 1997 Elsevier Science Ltd



- [17] **J.M. Santos, J.M. Pinazo, J. Canada.** Methodology for generating daily clearness index index values  $K_t$  starting from the monthly average daily value  $K^-_t$ . Determining the daily sequence using stochastic models
- [18] **P.K. Jain, E.M. Lungu.** Stochastic models for sunshine duration and solar irradiation. *Renewable Energy* 27 (2002) 197–209 Elsevier Science.
- [19] **C. Benard, E. Boileau et B. Guerrier.** Modélisation de l'irradiation solaire globale à l'aide de processus Arma : application à la prédiction à faible pas de temps (horaire), en vue de l'établissement de commandes optimales dans l'habitat. *Revue Phys. Appl.* 20 (1985). 845-855.
- [20] **LL. Mora-Lopez and M. Sidrach-de-Cardona.** Multiplicative arma models to generate hourly series of global Irradiation. *Solar Energy* Vol. 63, No. 5, pp. 283–291, 1998 Elsevier Science Ltd
- [21] **LL. Mora-Lopez and M. Sidrach-de-Cardona.** Characterization and simulation of hourly exposure series of global radiation. *Solar Energy* Vol. 60, No. 5, pp. 257-270. 1997 Elsevier Science.
- [22] **A.Khireddine et K. Benmahammed.** Méthodologie de Correction de Rayonnement Solaire. *Rev. Energ. Ren.* Vol.4 (2001) 79-87
- [23] ENERGIE SOLAIRE PHOTOVOLTAÏQUE, volume 1, physique et technologie de la conversion photovoltaïque, publié par l'UNESCO 1991.
- [24] **Klaus Haars,** Electricity From Sunlight ,solar energy supply for homes and buildings, edition Vieweg 1997.
- [25] **Bernhard Krineg,** Strom aus der Sonne, Solartechnik in theorie und praxis, Edition 1998.
- [26] **Zekai Sen,** Solar Energy Fundamentals and Modeling Techniques, Atmosphere, Environment, Climate Change and Renewable Energy, Springer 2008 ISBN 978-1-84800-133-6.
- [27] **Arthur Charpentier,** cours de séries temporelles théorie et applications. [www.math.univ-montp2.fr](http://www.math.univ-montp2.fr)
- [28] **Philippe Tassi,** méthodes statistiques, Economica 1989. [www.ebooks-gratuit.org/livres](http://www.ebooks-gratuit.org/livres)
- [29] **Christophe HURLIN,** Econométrie Appliquée Séries Temporelles. UFR Economie Appliquée. Cours de C. Hurlin. [www.univ-orleans.fr](http://www.univ-orleans.fr)
- [30] **François-Eric Racicot et Raymond Théoret,** Le calcul numérique en finance empirique et quantitative. e-book (google)
- [31] **George E.P. Box and Gwilym M.Jenkins.** Time series forecasting and control. Copyright 1976 by Holden day, INC.
- [32] **Akaike H.** A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistics and Mathematics* 1978; 30(A):9–14.
- [33] **Ljung GM, Box GEP.** On a measure of lack of fit in time series models. *Biometrika* 1978;65(2):297–303.
- [34] **Tom Markvart.** Practical handbook of photovoltaic: fundamentals and applications. <http://books.google.fr>



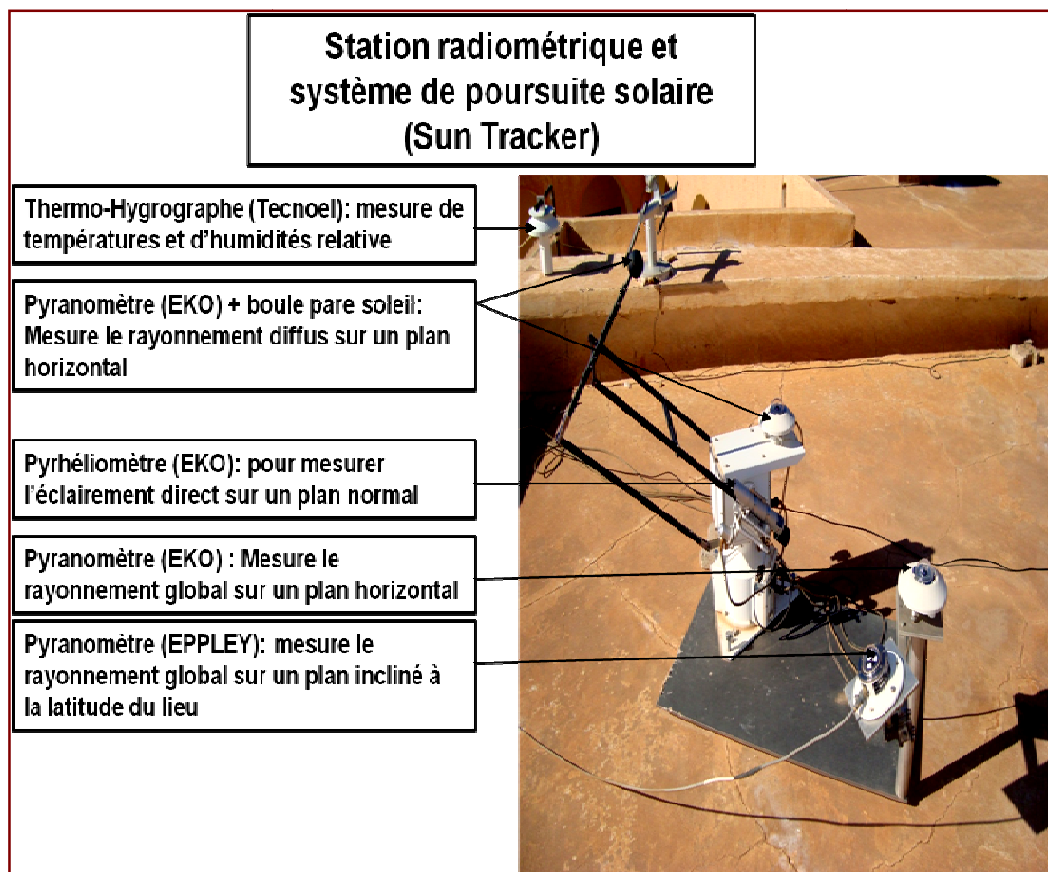
## *Annexe*

---

## 1. Présentation de la station radiométrique :

La station fonctionne par le principe de poursuite bidimensionnelle contenant des capteurs solaires suivants (voir figure. 1) :

- Un Pyranomètre EKO : pour la mesure de l'irradiance globale reçue sur un plan horizontal
- Un pyréliomètre EKO : sert à mesurer l'irradiance directe perpendiculaire au plan du capteur.



**Fig.1** présentation de la station radiométrique installée à l'unité de recherche de Ghardaïa.

## 2. Le prétraitement des mesures effectuées :

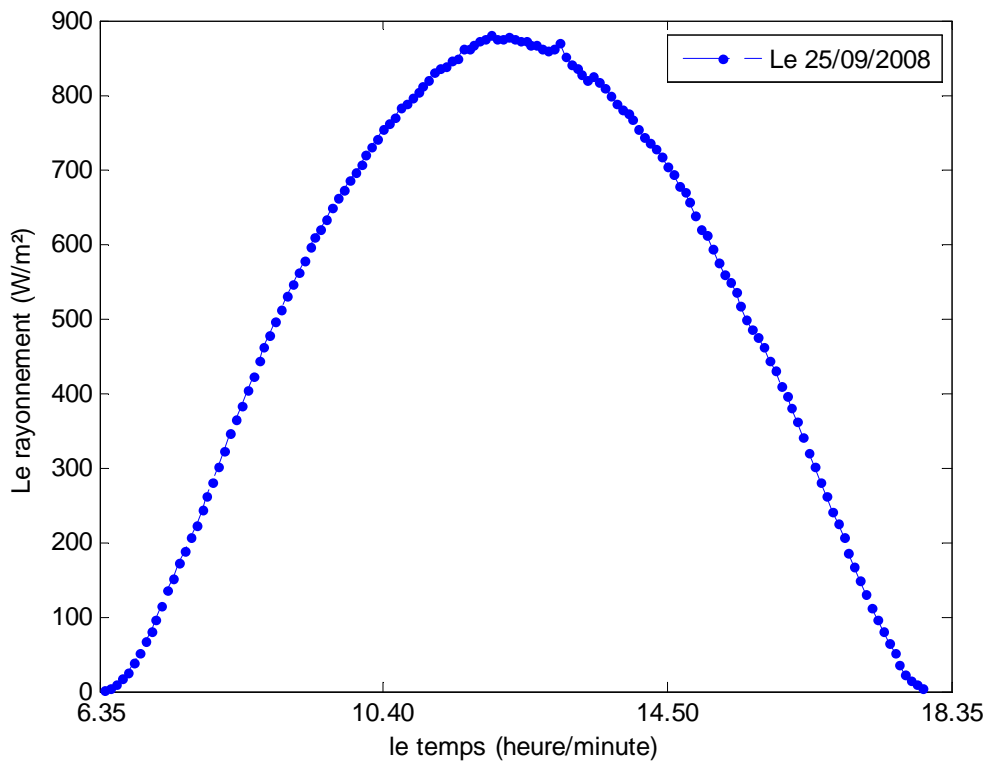
Les mesures du rayonnement solaire global ( $W/m^2$ ) au sol par rapport à une surface horizontale ont été prises à pas de cinq minutes, sans arrêt y compris la nuit.

Cette figure montre l'acquisition des données et le stockage.

	A	B	C	D	E	F	G	H	I	J	K
		Diffus W/m <sup>2</sup>	Direct W/m <sup>2</sup>	Global W/m <sup>2</sup>	G 32° W/m <sup>2</sup>	temperature	humidité	Année	N° du jour	Heure (h-mn)	
1											
2	107	0	0	0	0	6.698	65.32	2005	1	5	
3	107	0	0	0	0	6.527	65.63	2005	1	10	
4	107	0	0	0	0	6.383	66.11	2005	1	15	
5	107	0	0	0	0	6.376	66.49	2005	1	20	
6	107	0	0	0	0	6.36	66.77	2005	1	25	
7	107	0	0	0	0	6.205	67.55	2005	1	100	
8	107	0	0	0	0	6.138	67.74	2005	1	105	
9	107	0	0	0	0	6.115	67.94	2005	1	110	
10	107	0	0	0	0	6.024	68.08	2005	1	115	
11	107	0	0	0	0	6.088	68.09	2005	1	120	
12	107	0	0	0	0	6.101	67.97	2005	1	125	
13										645	
14	107	0	0	0	0	4.522	73.9	2005	1	650	
15	107	0	0	0	0	4.507	74	2005	1	655	
16	107	0	0	0	0	4.444	74	2005	1	700	
17	107	0	0	0	0	4.515	74.1	2005	1	705	
18	107	0	0.025	0	0	4.559	73.9	2005	1	710	
19	107	0	0.422	0	0	4.62	73.6	2005	1	715	
20	107	0	0.422	0	0	4.7	73.4	2005	1	720	
21	107	0	0.496	0	0	4.797	73	2005	1	725	
22	107	0	0.496	0	0	4.895	72.5	2005	1	730	
23	107	0	0.422	0	0	4.945	72.2	2005	1	735	
24	107	0	0.472	0	0	4.989	72	2005	1	740	
25	107	2.165	6.279	2.331	2	5.043	71.8	2005	1	745	
26	107	4.853	53.48	6.328	8	5.111	71.6	2005	1	750	
27	107	8.04	125.3	13.06	21	5.208	71.2	2005	1	755	
28	107	12.32	194.5	23.15	40	5.312	70.8	2005	1	800	

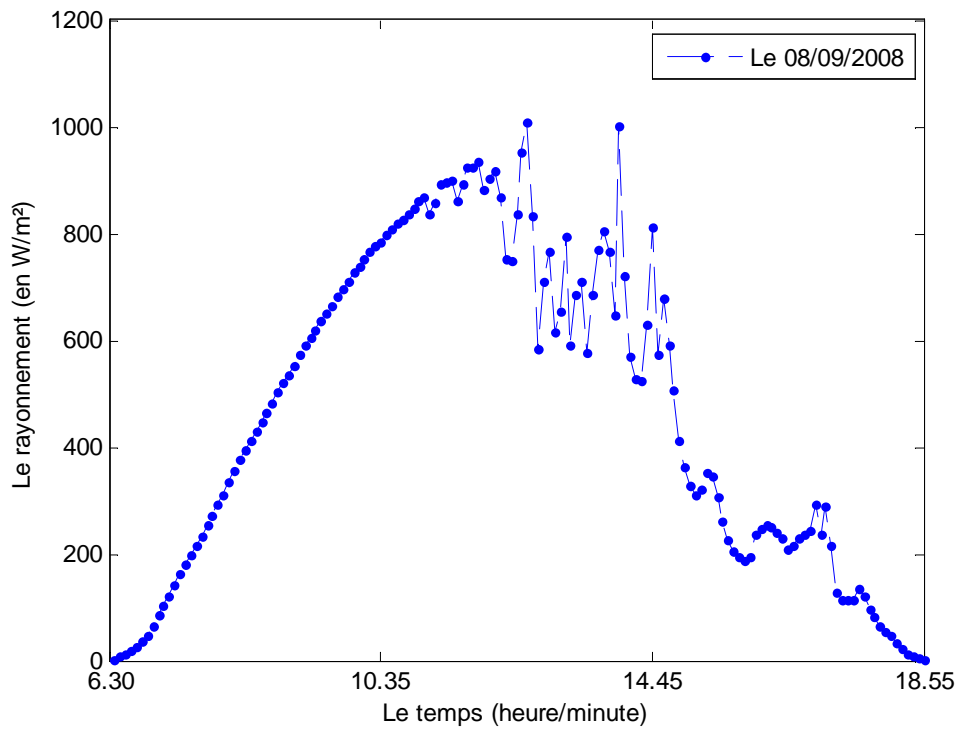
Ensuite, nous avons procédé à une intégration par rapport au temps de ces données et ceci du levé du soleil jusqu'au couché dans le but d'obtenir une énergie qui est l'irradiation solaire globale journalière estimée en  $Wh/m^2/jour$ . Ces traitements sont établis sous le logiciel Excel.

La Fig.2, présente le rayonnement global durant une journée bien ensoleillée exemple le 25 /09/2008 et l'irradiation globale moyenne journalière est estimée à :  $G = 6258.82 Wh/m^2$



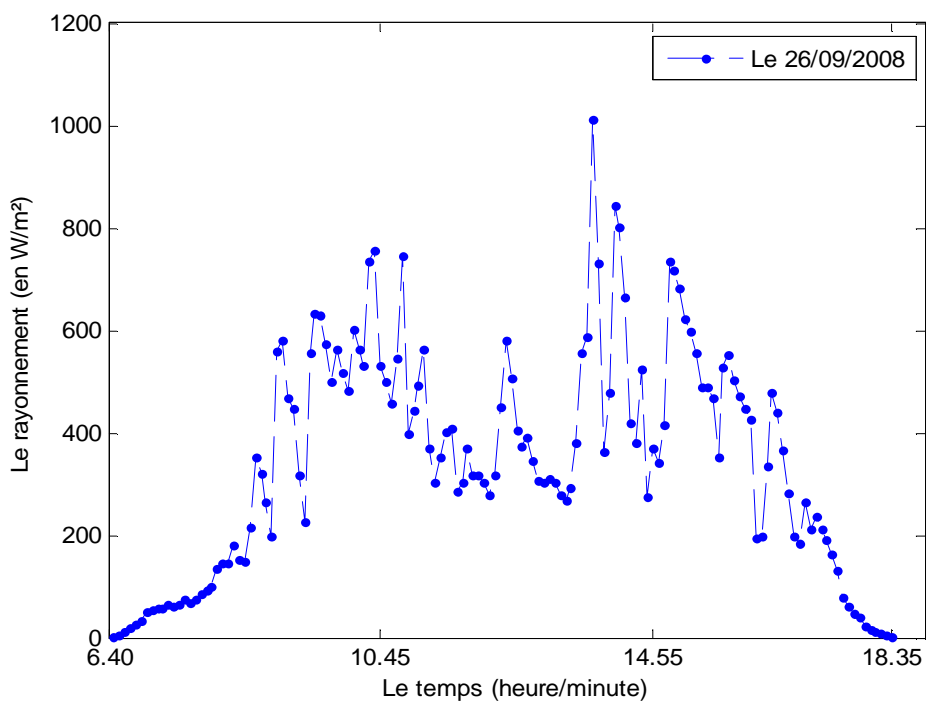
**Fig.2 :** Le rayonnement global durant la journée du 25/09/2008.

La Fig.3, présente le rayonnement global durant une journée perturbée exemple la journée du 08/09/2008. Et l'irradiation globale moyenne journalière estimée à :  $G = 5615.69 Wh/m^2$



**Fig.3** : Le rayonnement global durant la journée du 08/09/2008.

La Fig.4 présente le rayonnement globale durant une journée de mauvais temps exemple celle du 26/09/2008. Et l'irradiation globale moyenne journalière estimée à  $G = 4151.67 Wh/m^2$



**Fig.4** : Le rayonnement global durant la journée du 26/09/2008.

**La liste des figures :**

N° Fig	LE TITRE	Page
<b>Fig.1</b>	L'allure du spectre ; flux de puissance en fonction de la longueur d'onde en microns	<b>12</b>
<b>Fig.2</b>	L'angle et la distance du soleil par rapport à l'atmosphère	14
<b>Fig.3</b>	Parcours du rayonnement solaire. <b>a</b> Direct. <b>b</b> Diffus. <b>c</b> Réfléchis	15
<b>Fig.4</b>	La variation de l'irradiation solaire extraterrestre	16
<b>Fig.5</b>	L'irradiation globale hors atmosphère $G_0$ par rapport au site de Ghardaïa	18
<b>Fig.6</b>	Processus Non Stationnaire : Trend Déterministe	35
<b>Fig.7</b>	La série X et ses prévisions pour les 36 mois à compter de Juillet 1957	57
<b>Fig.8</b>	Données de l'irradiation globale journalières du 1 Janvier 2005 au 31 Mai 2009	58
<b>Fig.9</b>	L'autocorrélogramme et l'autocorrélogramme partiel de $G_t$	59
<b>Fig.10</b>	L'autocorrélogramme et l'autocorrélogramme partiel de $\Delta_{365} G_t$	59
<b>Fig .11</b>	$X_t$ et ses autocorrélogrammes	60
<b>Fig.12</b>	Processus itératif de la modélisation	61
<b>Fig.13</b>	Autocorrélogramme de $\epsilon_t$	64
<b>Fig.14</b>	Données de l'irradiation globale et ses prévisions du 1 Juin 2009 au 31 mai 2010	67



<b>Fig.15</b>	Les moyennes journalières mensuelles de l'irradiation globale reçus sur un plan horizontale à Ghardaïa et leurs prévisions.	69
<b>Fig.16</b>	Les moyennes journalières mensuelles des irradiations globales calculées pour les différentes années : 2005, 2006, 2007, 2008, 2009, et 2010.	71
<b>Fig.17</b>	Les moyennes journalières mensuelles des irradiations globales calculées pour les différentes années : 2005, 2006, 2007, 2008, 2009, et 2010.	72
<b>Fig.18</b>	Données de l'irradiation globale ainsi que la courbe de tendance sinusoïdale $U_t$	75
<b>Fig.19</b>	L'autocorrélogramme et l'autocorrélogramme partiel de $Y_t$	76
<b>Fig.20</b>	$W_t$ et ses autocorrélogrammes	77
<b>Fig.21</b>	Données des résidus $Y_t$ et ses prévisions	80
<b>Fig.22</b>	Les données de l'irradiation globale et ses prévisions	81
<b>Fig.23</b>	Les moyennes journalières mensuelles de l'irradiation globale reçus sur un plan horizontale à Ghardaïa et leurs prévisions.	82
<b>Fig.24</b>	L'irradiation globale hors atmosphère $G_0$ par rapport au site de Ghardaïa	83
<b>Fig.25</b>	Les résidus $G_r$ (en WH/m <sup>2</sup> )	84
<b>Fig.26</b>	les résidus $G_r$ et ses corrélogrammes	84
<b>Fig.27</b>	Données des résidus $G_r$ ainsi que leurs prévisions via le modèle SARIMA (1,0,1)(0,1,1)	86
<b>Fig.28</b>	Les moyennes journalières de l'irradiation globale ainsi que leurs prévisions	87
<b>Fig.29</b>	Les moyennes journalières mensuelles de l'irradiation globale et ses prévisions	88

**La liste des tableaux :**

<b>N° Tableau</b>	<b>LE TITRE</b>	<b>Page</b>
<b>Tableau. 1</b>	Comportement de la FAC et la FACP des processus ARMA (p,q)	<b>43</b>
<b>Tableau.2</b>	Propriétés des processus AR(p), MA(q) et ARMA (p,q)	<b>44</b>
<b>Tableau.3</b>	Critères de sélection du modèle	<b>63</b>
<b>Tableau.4</b>	Tests statistiques pour les moyennes journalières de l'irradiation globale	<b>69</b>
<b>Tableau.5</b>	Tests statistiques calculés pour les moyennes journalières mensuelles de l'irradiation globale	<b>70</b>
<b>Tableau.6</b>	Testes statistiques de quelques modèles empiriques entrepris par M.Koussa [8] et du modèle SARIMA obtenu.	<b>70</b>
<b>Tableau.7</b>	Critères de sélection du modèle	<b>78</b>
<b>Tableau.8</b>	Tests statistiques pour les moyennes journalières de l'irradiation globale	<b>81</b>
<b>Tableau.9</b>	Tests statistiques calculés pour les moyennes journalières mensuelles de l'irradiation globale	<b>82</b>
<b>Tableau.10</b>	Critères de sélection du modèle	<b>85</b>
<b>Tableau.11</b>	Tests statistiques calculés pour les moyennes journalières de l'irradiation globale	<b>87</b>
<b>Tableau.12</b>	Tests statistiques calculés pour les moyennes journalières mensuelles de l'irradiation globale	<b>88</b>

## **Résumé :**

Dans ce travail, nous nous sommes intéressés à la modélisation stochastique et à la prévision de l'irradiation solaire globale journalière dans le site de Ghardaïa en adoptant le modèle SARMA (Seasonal Auto Regressive Moving Average) de Box Jenkins. Le choix de ce modèle par rapport aux modèles empiriques, généralement utilisés, réside dans son aptitude à tenir compte de la majorité des phénomènes aléatoires qui peuvent influencer cette variable météorologique, tel que les nuages, la pluie,.....et surtout le vent de sable qui est un phénomène fréquent dans ce site. Nous proposons dans cette étude un modèle de prévision qui donne une bonne approximation de nos valeurs mesurées de l'irradiation solaire globale moyenne journalière. L'efficacité de ce modèle est appuyée par une série de tests statistiques, et également par une comparaison aux modèles existants dans la littérature.

**Mots clés— Box Jenkins, moyennes journalières de l'irradiation globale, prévision, SARIMA.**

## **Summary:**

In this work, we were interested in stochastic modeling and the forecast of the average day of the total solar irradiation on a horizontal surface in the site of Ghardaïa by adopting model SARMA (Seasonal Auto Regressive Moving Average) of Box Jenkins. The choice of this model compared to the empirical models, generally used, lies in its aptitude to take account of the majority of the random phenomena which can influence this weather variable, such as clouds, rain ...and specially the sand wind which is a frequent phenomenon in this site. We propose in this study a forecasting model which gives a good approximation of our measured values of the total solar irradiation. The effectiveness of this model is approved by a series of statistical tests, and also by a comparison with the existing models in the literature.

**Key words — Box Jenkins, average day of the total solar irradiation, forecast, SARIMA.**

## الخلاصة :

في هذا العمل ، نحن مهتمون بالتمذجة والتنبؤ العشوائي للإشعاع الشمسي اليومي في موقع غرداية وذلك باعتماد نموذج SARIMA لبوكس جينكينز. اختيار هذا النموذج بالمقارنة مع النماذج التجريبية التي يشيع استخدامها ، هو قدرته على حساب الغالبية العظمى من الأحداث العشوائية التي يمكن أن تؤثر على الطقس المتغير ، مثل الغيوم والأمطار... وخصوصا الرياح الرملية التي تعتبر ظاهرة متكررة في مثل هذا الموقع. نقترح في هذه الدراسة نموذج تنبؤي يعطي تقريبا جيد لقيم الإشعاع اليومي المتوسط للطاقة الشمسية المتحصل عليها تجريبيا بجهاز رصد الأشعة الشمسية. تدعم فعالية هذا النموذج سلسلة من الاختبارات الإحصائية و مقارنات بالنماذج المذكورة في المراجع العلمية.

## كلمات مفتاحية

بوكس جينكينز ، المتوسطات اليومية للإشعاع الشمسي ، التنبؤ ، SARIMA.