

Impact des indices climatiques (A.A.O, I.O.D et ENSO) sur la prévision de la sécheresse dans toute la partie sud de Madagascar

Jacques Chrysologue RATSIMAVO^{1*}, Adolphe Andriamanga RATIARISON²,
Olga RAMIARINJANAHARY¹, Jean Eugène RANDRIANANTENAINA¹,
Niry Arinavalona RAKOTOVAO² et Myriame HAJARISOA¹

¹ *Faculté des Sciences, Département de Physiques et Applications, Université de Toliara, Madagascar*

² *Laboratoire de la Dynamique de l'Atmosphère du Climat et des Océans (DyACO), Université d'Antananarivo, Sciences et Technologie, Physique et Applications, Madagascar*

* Correspondance, courriel : rat.chryso@gmail.com

Résumé

Ce travail consiste à étudier la sécheresse la partie Sud de Madagascar dans le but de faire sa prévision. Notre zone d'étude est délimitée entre les latitudes 21° et 26° Sud et les longitudes 43° et 48° Est. Nous avons utilisé des données de pluie et des indices climatiques : Dipôle de l'Océan Indien (I.O.D), Oscillation de l'Antarctique (A.A.O) et El niño Southern Oscillation (ENSO) pendant les trente années (1983 à 2012). L'Analyse en Composante Principale (A.C.P) nous a permis de définir trois zones : Zone A (Ambovombe, Androka, Basibasy, Cap Sainte Marie, Farafangana, Tolagnaro et Toliara), Zone B (Betroka, Manakara, Midongy et Vangaindrano) et la Zone C (Bekily). La modélisation (SARIMA (5, 1, 7) (1, 1, 1)₁₂) pour la Zone A, nous a donné une prévision fiable. L'inter corrélation des indices climatiques (A.A.O, I.O.D et ENSO) avec la sécheresse (Zone A) selon la corrélation de Spearman (non linéaire) nous a montré qu'il y a une faible corrélation entre les variables sauf pour la Nino 3.

Mots-clés : *sécheresse, SARIMA, test, Ljung-Box, Jarque-Bera, corrélation.*

Abstract

Impact of the climates indices (A.A.O, I.O.D and ENSO) on the forecasting of the drought in the whole left south of Madagascar

This work consists of studying drought in order to make its prediction. Our study area is the southern part of Madagascar (delineated between latitudes 21 ° and 26 ° South and longitudes 43 ° and 48 ° East). We used rain data and climatic indices: Indian Ocean Dipole (I.O.D.), Antarctic Oscillation (A.A.O) and El niño Southern Oscillation (ENSO) during the thirty years (1983 to 2012). The Principal Component Analysis (P.C.A) allowed us to define three zones: Zone A (Ambovombe, Androka, Basibasy, Cape Santa Maria, Farafangana, Tolagnaro and Toliara), Zone B (Betroka, Manakara, Midongy and Vangaindrano) Zone C (Bekily). The modeling (SARIMA (5, 1, 7) (1, 1, 1) 12) for Zone A gave us a reliable forecast. The inter-correlation of climatic indices (A.A.O, I.O.D and ENSO) with drought (Zone A) according to Spearman's correlation (nonlinear) showed that there is a weak correlation between variables except for Nino 3.

Keywords : *drought, SARIMA, test, Ljung-Box, Jarque-Bera, correlation.*

1. Introduction

La pluviométrie est un facteur déterminant pour le développement d'un pays. L'île de Madagascar est située au Sud-Est de l'Afrique. Elle est localisée entre les longitudes 43°Est et 51°Est, et les latitudes 11°Sud et 26°Sud [1]. Le changement climatique a un impact majeur au niveau mondial et Madagascar est l'un des pays touchés par ce phénomène en particulier la sécheresse de la partie Sud [2]. Nous avons étudié la sécheresse en vue de pouvoir faire une prévision à long terme. Le but de ce travail est de prévoir la sécheresse en appliquant une modélisation ARIMA (p, d, q), et de voir son inter corrélation avec l'A.A.O, I.O.D et ENSO (Nino1_2, Nino3, Nino3_4 et Nino4) dans la partie Sud de Madagascar.

2. Méthodologie

2-1. Données

Nous avons utilisé des données de pluie provenant du Météo France/La Réunion et du site web de la NOAA pour les indices climatiques [A.A.O, I.O.D et ENSO (Nino1_2, Nino3, Nino3_4 et Nino4)] sur une durée de trente ans (de 1983 à 2012).

2-2. Analyse en Composante Principal (A.C.P) [3 - 6]

But : Nous avons fait appel à cette méthode pour faire le zonage de notre Zone d'étude. Une méthode de statistique exploratoire permettant de décrire un grand tableau de données de type individus / variables. L'A.C.P propose une représentation dans un espace de dimension réduite, permettant ainsi de mettre en évidence d'éventuelles structures au sein des données. Pour cela, nous recherchons les sous-espaces dans lesquels la projection du nuage déforme le moins possible le nuage initial.

2-2-1. Données de départ

Les données sont représentées sous la forme d'un grand **Tableau** à n lignes et p colonnes, que l'on stocke sous la forme d'une matrice X de taille $n \times p$:

- chaque ligne (x_i^1, \dots, x_i^p) de X représente les valeurs prises par l'individu i sur les p variables ;
- de même chaque colonne $(x_1^j, \dots, x_n^j)^t$ de X représente les valeurs de la variable j pour les n individus.

Par simplicité de langage, l'individu i sera identifié au vecteur $x_i = (x_i^1, \dots, x_i^p)^t$ de \mathbb{R}^p tandis que la variable j sera identifié au vecteur $x^j = (x_1^j, \dots, x_n^j)^t$ de \mathbb{R}^n .

2-2-1-1. Données centrées

En retranchant \bar{x}^j aux valeurs de la variable j prises par les individus, on construit le tableau Y des données centrées

$$y_i^j = x_i^j - \bar{x}^j \quad (1)$$

Le terme *centré* signifie que les moyennes des variables y^j sont nulles.

2-2-1-2. Données centrées réduites (normées)

En divisant chaque colonne j du tableau centré Y par l'écart-type s^j de la variable j , on construit le tableau Z des données centrées réduites :

$$z_i^j = \frac{x_i^j - \bar{x}^j}{s^j} \tag{2}$$

2-2-2. Point moyen du nuage des individus

Le vecteur g des moyennes arithmétiques de chacune des p variables définit le point moyen du nuage des individus :

$$g = (\bar{x}^1, \dots, \bar{x}^p)^t \tag{3}$$

où, $\bar{x}^j = \frac{1}{n} \sum_{i=1}^n x_i^j$ est la moyenne des valeurs prises par la j -ème variable.

2-2-3. Matrice de variance

On appelle matrice de variance la matrice symétrique V contenant les variances s_j^2 sur la diagonale et les covariances v_{kl} en dehors de la diagonale (ligne k colonnes l pour v_{kl}). Cette matrice s'écrit :

$$V = \frac{1}{n} X' X - g g' = \frac{1}{n} Y' Y \tag{4}$$

On note $s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_i^j - \bar{x}^j)^2$ la variance de la variable j , sa racine carré s_j est l'écart-type et

$v_{kl} = \frac{1}{n} \sum_{i=1}^n (x_i^k - \bar{x}^k)(x_i^l - \bar{x}^l)$ est la covariance des variables k et l .

De même, on définit le coefficient de corrélation linéaire entre les variables k et l par :

$$r_{kl} = \frac{v_{kl}}{s_k s_l} \tag{5}$$

Ce coefficient exprime le niveau de corrélation (linéaire) entre les variables k et l : plus il est proche de 1, plus les variables sont corrélées positivement, plus il est proche de -1, plus elles sont corrélées négativement. Un coefficient de corrélation nul indique l'absence de corrélation linéaire.

2-2-4. Les espaces des individus et des variables

2-2-4-1. L'espace des individus

2-2-4-1-1. La métrique

L'objectif de l'ACP est de visualiser ce nuage de point dans un espace de faible dimension le plus fidèlement possible. L'analyse repose donc sur les distances entre individus dans \square^p , et le choix de la métrique aura donc

une influence sur le résultat de l'ACP. La distance $d(i, j)$ entre deux individus i et j peut s'écrire de façon générale :

$$d^2(i, j) = d^2(x_i, x_j) = \|x_i - x_j\|_M^2 = \langle x_i - x_j, x_i - x_j \rangle_M = (x_i - x_j)' M (x_i - x_j) \quad (6)$$

où, M est une matrice symétrique définie positive de taille p spécifiant la distance choisie, $\langle \cdot, \cdot \rangle_M$ et $\| \cdot \|_M$ étant le produit scalaire et la norme associés à la métrique M .

2-2-4-1-2. L'inertie

L'inertie est une notion fondamentale en A.C.P, puisqu'elle est une mesure de la dispersion du nuage de points autour de son centre de gravité g . L'inertie totale du nuage autour de son centre de gravité g est :

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(g, x_i) = \frac{1}{n} \sum_{i=1}^n (x_i - g)' M (x_i - g) \quad (7)$$

Plus l'inertie est grande, plus le nuage est dispersé, et au contraire plus elle est petite, plus le nuage est concentré sur son centre de gravité. Lorsque $M = I$, l'inertie totale peut s'écrire

$$I_g = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_i^j - \bar{x}^j)^2 = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (x_i^j - \bar{x}^j)^2 = \sum_{j=1}^p s_j^2 \quad (8)$$

qui n'est autre que la trace de la matrice de variance V

$$I_g = \text{tr}(V) \quad (9)$$

Ce résultat se généralise au cas d'une métrique M quelconque en $I_g = \text{tr}(VM)$

Dans le cas de l'ACP normée, la matrice de variance du tableau Z est la matrice de corrélation R et sa trace est donc égale au nombre p de variables $I_g = p$. L'inertie du nuage par rapport à un axe Δ passant par le centre de gravité g est

$$I_\Delta = \frac{1}{n} \sum_{i=1}^n d^2(\Delta, x_i) = \frac{1}{n} \sum_{i=1}^n d^2(h_{\Delta i}, x_i) \quad (10)$$

où, $h_{\Delta i}$ est la projection orthogonale de x_i sur l'axe Δ . Cette inertie mesure la proximité du nuage des individus à l'axe.

De même, l'inertie du nuage par rapport à un sous-espace vectoriel F passant par le centre de gravité g est

$$I_F = \frac{1}{n} \sum_{i=1}^n d^2(h_{Fi}, x_i) \quad (11)$$

où h_{Fi} est la projection orthogonale de x_i sur le sous-espace F .

Soit F^\perp le supplémentaire orthogonal de F dans \mathbb{R}^p , alors d'après le théorème de Pythagore (cf. **Figure 14** en dimension 3) on a

$$d^2(h_{F_i}, x_i) + d^2(h_{F^{\perp i}}, x_i) = d^2(g, x_i) = d^2(g, h_{F_i}) + d^2(g, h_{F^{\perp i}}) \tag{12}$$

2-2-4-1-3. L'espace des variables

Chaque variable est considérée comme un vecteur d'un espace de dimension n, l'espace des variables. La métrique utilisée pour le calcul des distances entre variables est la métrique identité, car on suppose que tous les individus ont le même poids. Si ce n'est pas le cas une métrique spécifique doit alors être utilisée. Soit y^1, \dots, y^p les variables x^1, \dots, x^p centrées. On a les propriétés suivantes :

– le produit scalaire entre deux variables y^k et y^l (sous la métrique identité) est

$$\langle y^k, y^l \rangle = \sum_{i=1}^n y_i^k y_i^l = v_{kl} \tag{13}$$

– le carré de la norme d'une variable est égal à sa variance

$$\|y^k\|^2 = s_k^2 \tag{14}$$

et l'écart-type de la variance représente donc sa longueur,

– le cosinus de l'angle θ_{kl} entre deux variables y^k et y^l est leur coefficient de corrélation linéaire :

$$\cos(\theta_{kl}) = \frac{\langle y^k, y^l \rangle}{\|y^k\| \|y^l\|} = \frac{v_{kl}}{s_k s_l} = r_{kl} \tag{15}$$

Nous nous intéressons donc, dans l'espace des variables, aux angles entre variables plutôt qu'aux distances, et on représente les variables comme des vecteurs et non des points.

2-2-5. L'analyse en composantes principales

L'ACP consiste à chercher un sous-espace Fk de dimension k inférieure à celle de l'espace de départ, tel que le nuage, une fois projeté dans ce sous-espace, soit au minimum déformé. Comme la projection diminue nécessairement les distances, on cherche le sous-espace Fk qui maximise la moyenne des carrés des distances entre individus :

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n d^2(i, j) = 2 I_g \tag{16}$$

Il faut donc que l'inertie du nuage projeté soit maximale.

2-2-5-1. Recherche de l'axe portant l'inertie maximale

On cherche l'axe Δ_u tel que l'inertie $I_{\Delta_u^\perp}$ expliquée par cet axe soit maximale. Cela revient à chercher Δ_u tel que I_{Δ_u} soit minimale d'après le théorème de Huygens (12). Puisqu'en projetant sur l'axe Δ_u on perd l'inertie I_{Δ_u} , on aura bien une inertie restante $I_{\Delta_u^\perp}$ maximale, ce qui revient à déformer le moins possible le nuage des individus. Comme on est en ACP normée la métrique M est l'identité et la matrice de variance V est égale

à la matrice de corrélation R . L'inertie expliquée par l'axe Δ_u est alors $I_{\Delta_u} = u' R u$. Il faut donc trouver le vecteur unitaire u solutions du problème d'optimisation sous contrainte suivant :

$$\begin{cases} \max_u u' R u \\ u' u = 1 \end{cases} \quad (17)$$

La matrice de corrélation R étant symétrique, elle est diagonalisable : $R = P \Delta P'$ où Δ est la matrice diagonale composée des valeurs propres $\lambda_1 \geq \dots \geq \lambda_j \geq \dots \geq \lambda_p$, et où la matrice de passage P est la matrice orthogonale dont les colonnes sont les vecteurs propres v_j de R . Il vient alors que

$$u' R u = \sum_{j=1}^p \lambda_j \langle u, v_j \rangle^2 \quad (18)$$

En nommant u_j les composantes du vecteur u (normé) dans la base des vecteurs propres, on a

$$u' R u = \sum_{j=1}^p \lambda_j u_j^2 \leq \underbrace{(u_1^2 + \dots + u_p^2)}_{= u' u = 1} \leq \lambda_1 \quad (19)$$

Le vecteur u maximisant cette quantité n'est autre que v_1 , le vecteur propre associé à la plus grande valeur propre λ_1 de R , pour lequel on a donc $I_{\Delta_{v_1}} = \lambda_1$.

2-2-5-2. Recherche des axes suivants

Nous cherchons cette fois un vecteur unitaire u , orthogonal au précédent (v_1), et maximisant la quantité $u' R u$. En nommant encore u_j les composantes de ce vecteur u dans la base des vecteurs propres, on a :

$$u' R u = \sum_{j=2}^p \lambda_j u_j^2 \leq (u_2^2 + \dots + u_p^2) \leq \lambda_2 \quad (20)$$

Le vecteur u maximisant cette quantité n'est autre que v_2 , le vecteur propre associé à la seconde plus grande valeur propre λ_2 de R , pour lequel on a donc $I_{\Delta_{v_2}} = \lambda_2$. Et ainsi de suite. On en déduit ainsi le théorème suivant.

Notations

- Les axes Δ_{v_k} sont appelés axes factoriels ou axes principaux ;
- L'inertie expliquée par l'axe Δ_{v_k} est la valeur propre λ_k ;
- L'inertie expliquée par le sous-espace factoriel FK engendré par les axes factoriels est

$$I_{F_K} = \lambda_1 + \dots + \lambda_K \quad (21)$$

et le pourcentage d'inertie expliqué par ce sous-espace est donc $\frac{\lambda_1 + \dots + \lambda_k}{p}$.

- On appelle k-ième composante principale les coordonnées $c^k \in \mathbb{R}^n$ des n individus sur l'axe factoriel Δ_{v_k} , qui sont les projections des individus sur ces axes :

$$c^k = Z v_k \tag{22}$$

Les composantes principales peuvent être vues comme les nouvelles variables dans le sous espace factoriel.

2-2-6. Interprétation des résultats d'une ACP

2-2-6-1. Qualité de représentation sur les axes factoriels

Un des points les plus délicats de l'ACP est d'apprécier la perte d'information engendrée par la réduction de la dimension. On a vu précédemment que la qualité de représentation du nuage dans le sous-espace factoriel est exprimée par le pourcentage d'inertie expliquée $\frac{\lambda_1 + \dots + \lambda_k}{p}$. La qualité de représentation d'un individu

i sur un axe factoriel est mesurée par le cosinus carré de l'angle entre l'axe factoriel et le vecteur z_i . Plus le cosinus est grand, plus z_i sera proche de l'axe factoriel et donc sera bien représenté sur cet axe. La qualité de représentation d'une variable j sur le k-ème axe factoriel est exprimée par le coefficient de corrélation linéaire $r(c^k, z^j)$ entre la variable initiale z^j et la nouvelle variable, composante principale, c^j . La valeur de cette corrélation sera également très importante pour interpréter les nouveaux axes factoriels en fonction des variables initiales. Le calcul de cette corrélation est :

$$r(c^k, z^j) = \sqrt{\lambda_k v_k^j} \quad \text{où } v_k^j \text{ est la } j\text{-ème coordonnée du vecteur } v_k \tag{23}$$

2-2-6-2. Sélection du nombre k d'axes factoriels à retenir

La sélection du nombre d'axes à retenir, abordée précédemment, est une étape importante d'une ACP. Les critères utilisés sont en général empiriques : la méthode du coude qui consiste à détecter un coude sur le diagramme des valeurs propres, ou le critère de Kaiser. Ce dernier consiste à ne retenir que les valeurs propres supérieures à la moyenne : puisqu'on est en ACP normée la somme des valeurs propres est égale au nombre p de ces valeurs propres, et la moyenne des valeurs propres est donc égale à 1. On ne retiendra donc par le critère de Kaiser que les valeurs propres supérieures à 1.

2-3. Modélisation ARIMA [7]

But : Nous avons fait cette modélisation pour prévoir la sécheresse dans la partie Sud de Madagascar.

$$\text{ARIMA (p, d, q)} \begin{cases} \text{AR(p)} & \text{Auto-régressif d'ordre p} \\ \text{I(d)} & \text{Intégré d'ordre d} \\ \text{MA(q)} & \text{Moyenne Mobile d'ordre q} \end{cases}$$

On utilise trois types de processus pour construire un modèle représentant le mieux possible le comportement d'une série temporelle : identification, estimation et diagnostic. Et il convient de réitérer jusqu'à ce que le résultat soit jugé satisfaisant.

2-3-1. Bruit blanc

Un bruit blanc est un processus stationnaire dont les accroissements sont indépendants et stationnaires. Le modèle du « bruit blanc » constitue la référence pour les résidus d'un modèle correctement spécifié. De ce fait :

$$\left. \begin{array}{l} \text{La moyenne : } E[Y_t] \\ \text{La variance : } V[Y_t] \\ \text{Le covariances : } \text{cov}[Y_t, Y_{t+\theta}] \end{array} \right\} \text{ sont indépendants du temps } t \quad (24)$$

2-3-2. Rendre stationnaire une série

Si la série n'est pas stationnaire, il faut d'abord la rendre stationnaire. Pour la rendre stationnaire, nous avons deux méthodes de transformations différentes : par différenciation et par racine carrée (ou logarithmique). Une fois obtenue la stationnarité de la série, l'étape suivante consiste à analyser le graphe de la fonction d'autocorrélation (FAC) et celui de la fonction d'autocorrélation partielle (FAP) afin de déterminer les paramètres (p, d, q) du modèle.

2-3-3. Exposant de Hurst [8]

L'exposant de Hurst permet de tester la robustesse de nos résultats. Il s'agit des statistiques R/S, introduite par Hurst en 1951. Cet exposant permet de classer les séries temporelles en fonction de la nature de leur mémoire, autrement dit de leur structure de dépendance [9]. C'est une mesure de persistance absolue [10]

Soit une série temporelle de rendements X_t , $t = 1, 2, \dots, T$, de moyenne \bar{X}_T , la statistique R/S, notée, Q_t s'écrit :

$$Q_t = \frac{R}{S_T} \quad (25)$$

$$\text{où, } R = \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T) \quad \text{et} \quad S_T = \left[\frac{1}{T} \sum_{j=1}^k (X_j - \bar{X}_T)^2 \right]^{\frac{1}{2}}$$

Cette statistique est asymptotiquement proportionnelle à T^H , où la constante H comprise entre 0 et 1, est appelée exposant de Hurst. L'exposant de Hurst est ainsi donné par :

$$H = \frac{\log Q_t}{\log T}$$

Si : $\begin{cases} 0,5 < H < 1 : \text{Le phénomène présente une dépendance à long terme (présence) de persistance.} \\ 0 < H < 0,5 : \text{Le phénomène est anti-persistant.} \\ H = 0,5 : \text{Le phénomène n'a pas de mémoire.} \end{cases}$

2-3-4. Coefficient d'autocorrélation

Le coefficient d'ordre k, ρ_k , peut être estimé par :

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y}_1)(y_{t-k} - \bar{y}_2)}{\sqrt{\sum_{t=k+1}^n (y_t - \bar{y}_1)^2 \sum_{t=k+1}^n (y_{t-k} - \bar{y}_2)^2}} \text{ avec } \bar{y}_1 = \frac{1}{n-k} \sum_{t=k+1}^n y_t \text{ et } \bar{y}_2 = \frac{1}{n-k} \sum_{t=k+1}^n y_{t-k} \quad (26)$$

2-3-5. Cas particuliers : Les processus SARIMA [11]

Un processus SARIMA (Seasonnal AutoRegressive Integreted Moving Average) d'ordre $(p, d, q)(P, D, Q)_s$, a une saisonnalité de période s et qu'on peut écrire :

$$\Phi_1(B)\Phi_2(B^S)(1-B)^d(1-B^S)^D X_t = \Theta_1(B)\Theta_2(B^S)a_t \quad (27)$$

où, Φ_1 : est un polynome de degré p ; Φ_2 : est un polynome de degré P ; Θ_1 : est un polynome de degré q ; Θ_2 : est un polynome de degré Q .

2-4. Corrélation de Spearman [12, 13]

But : Nous voulons voir la relation des indices climatiques [A.A.O, I.O.D et ENSO (Nino1-2, Nino3, Nino3-4, Nino 4)] avec la sécheresse. La meilleure **Formule** pour calculer le coefficient r_s de Spearman est :

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N} \quad (28)$$

2-4-1. Observations ex-aequo

Lorsqu'il y a un nombre important d'ex-aequo pour une donnée (ou les deux), il faut utiliser la **Formule** suivante pour calculer r_s :

$$r_s = \frac{\sum X^2 + \sum Y^2 - \sum d}{2\sqrt{\sum X^2 \sum Y^2}} \quad (29)$$

2-4-2. Signification de r_s [14]

Pour savoir si les deux variables étudiées ne sont pas liées entre elles, nous appliquons la **Formule** suivante :

$$t_{Calculé} = r_s \sqrt{\frac{N-2}{1-r_s^2}} \quad (30)$$

⇒ Il faut comparer $|t_{Calculé}|$ et $t_{S(\text{Théorique})}$ $\begin{cases} \text{si } |t_{Calculé}| > t_{S(\text{Théorique})} \Rightarrow \text{Il y a significativité} \\ \text{si } |t_{Calculé}| < t_{S(\text{Théorique})} \Rightarrow \text{Il n'y a pas significativité} \end{cases}$

3. Résultats et discussion

Etant donné que la partie Sud de Madagascar est très vaste, nous avons pris quelques lieux remarquables comme : Ambovombe, Androka, Basibasy, Bekily, Betroka, Cap Sainte Marie, Farafangana, Manakara, Midongy, Tolagnaro, Toliara et Vangaindrano. Nous avons pris ces lieux comme individus. Ainsi, nous obtenons 12 individus.

3-1. L'Analyse en Composante Principale (A.C.P)

Le **Tableau 1** donne le nombre d'évènements de « jours de non pluie » annuel des douze (12) individus (lieux). Les années 1983 à 2012 sont prises comme des variables. D'après l'OMM (Organisation Mondiale de la Météorologie), on compte comme « jour de pluie », si la hauteur de pluie durant cette journée est supérieure ou égale à 1mm ($RR \geq 1mm$), dans le cas contraire ($RR < 1mm$) on parle de « jour de non pluie » [15].

Tableau 1 : Nombre de « jours de non pluie » par individu (annuel)

	A_83	A_84	A_85	A_86	A_87	A_88	A_89	A_90	A_91	A_92	A_93	A_94	A_95	A_96	A_97	A_98	A_99	A_00	A_01	A_02	A_03	A_04	A_05	A_06	A_07	A_08	A_09	A_10	A_11	A_12
Ambovombe	345	327	337	327	335	346	314	340	341	341	333	326	331	318	327	322	313	328	319	328	319	328	319	329	318	329	332	326	319	324
Androka	351	347	337	345	355	348	335	353	347	358	349	347	347	339	338	339	329	336	330	345	332	337	337	338	336	342	344	346	338	340
Basibasy	351	334	333	339	342	330	327	347	328	337	331	337	330	324	322	319	307	315	315	334	328	331	335	321	320	325	326	335	336	335
Bekily	330	306	317	307	324	325	297	329	317	323	311	316	320	296	314	307	297	306	282	316	302	306	301	302	300	312	317	303	314	312
Betroka	306	282	303	289	298	309	288	320	308	313	306	303	297	290	298	289	292	296	272	306	290	290	287	295	278	300	301	293	298	287
Cap Ste Marie	351	327	346	329	342	346	321	344	344	341	339	337	335	330	327	334	326	333	319	330	325	323	322	337	329	332	328	331	324	315
Farafangana	312	240	270	247	266	254	226	267	265	244	272	214	245	231	255	231	223	223	234	226	196	231	218	301	284	296	291	284	285	285
Manakara	318	297	309	285	308	314	306	320	320	316	304	297	301	293	297	296	308	300	299	302	291	294	288	306	290	301	301	293	294	304
Midongy	315	292	309	289	304	318	289	320	312	320	301	296	303	291	292	289	287	296	285	303	283	282	286	299	280	285	290	284	283	283
Tolagnaro	305	272	301	283	287	311	282	291	305	298	276	268	284	286	282	264	254	270	270	257	255	256	249	282	269	284	272	278	268	282
Toliara	354	333	341	340	344	337	331	343	330	346	340	338	338	324	327	331	317	319	318	340	327	323	335	337	328	328	334	329	341	329
Vangaindrano	311	293	312	294	304	319	291	324	317	319	305	296	305	287	295	293	291	300	292	301	283	288	284	307	284	295	290	289	288	284

3-1-1. Valeurs propres

Le **Tableau 2** et la **Figure 1** montrent qu'on doit garder seulement les deux premiers facteurs F_1 et F_2 (selon le critère de Kaiser ou la méthode de coude).

Tableau 2 : Valeurs propres

	F1	F2	F3
Valeur propre	27,010	1,973	0,326
Variance (%)	90,032	6,575	1,088
% cumulé	90,032	96,608	97,696

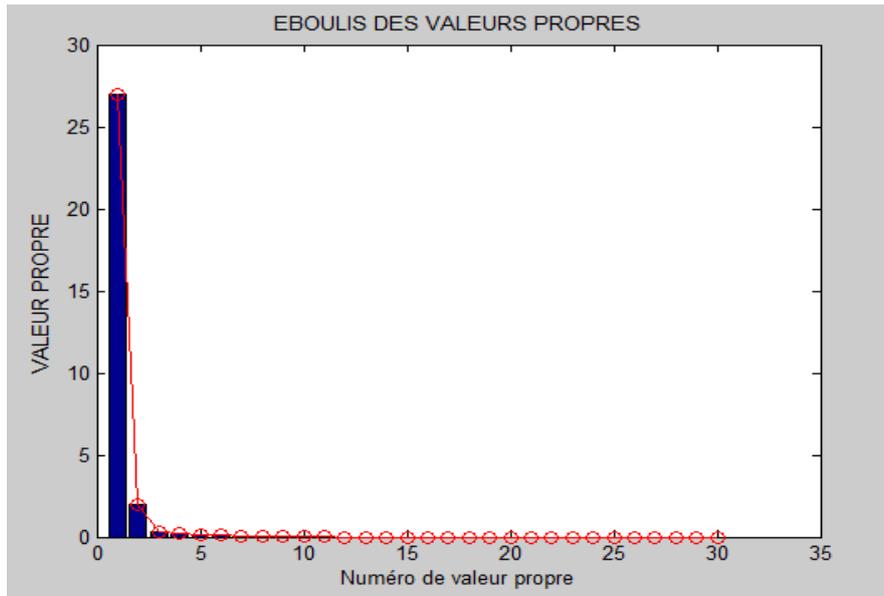


Figure 1 : Eboulis des valeurs propres

3-1-2. Projection des variables

On remarque que toutes les variables (années) sans exception sont fortement corrélées avec l'axe F_1 et faiblement corrélées avec F_2 , voir **Figure 2**.

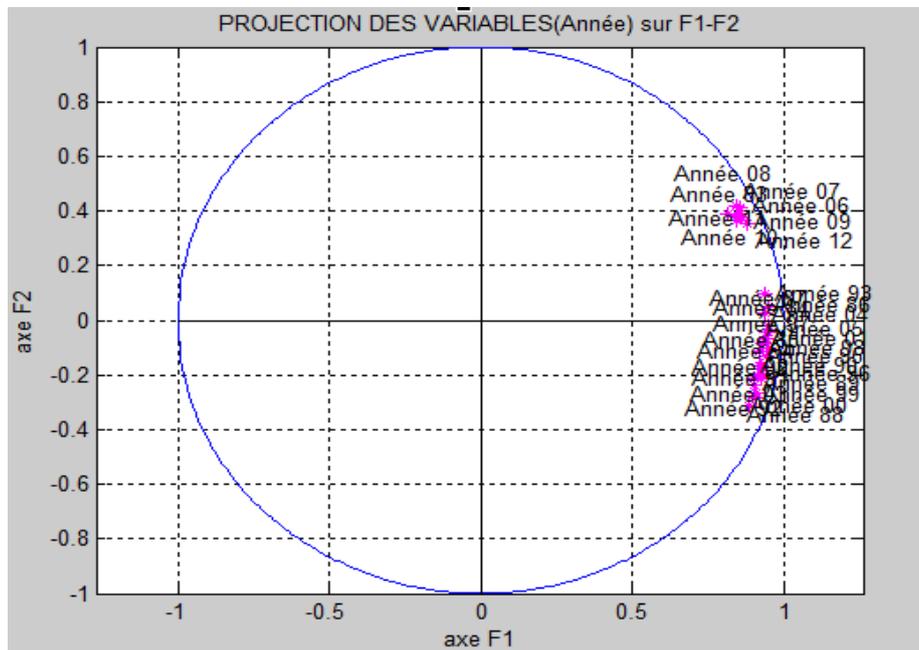


Figure 2 : Représentation des variables sur F1-F2

3-1-3. Projection des individus

A partir de cette **Figure 3**, on peut voir qu'il y a des individus (lieux) qui ont à peu près les mêmes coordonnées (en valeur absolue suivant l'axe F_1), donc on peut les regrouper et nous avons trois zones.

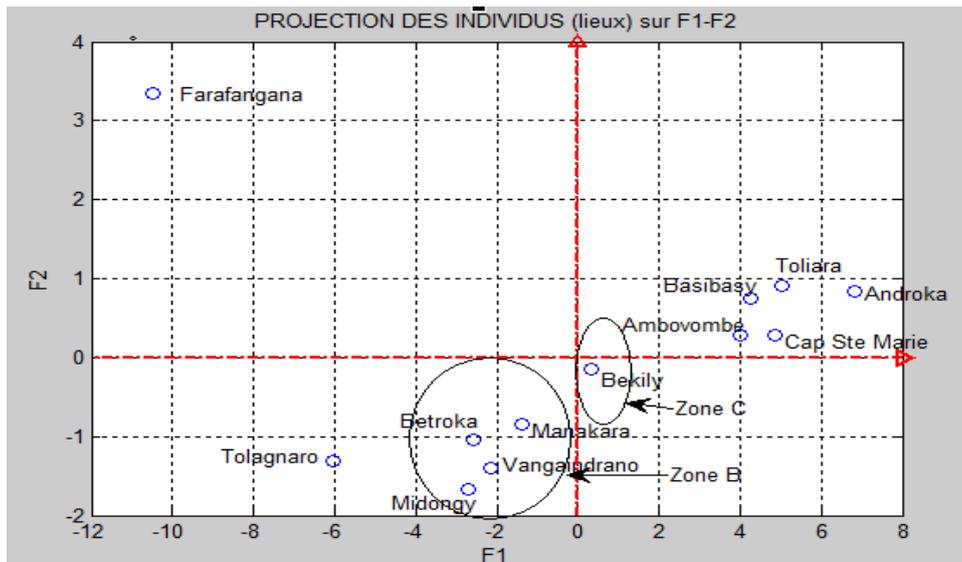


Figure 3 : Projection des individus sur F_1 - F_2

Tableau 3 : Répartition par Zone

Zone A	Zone B	Zone C
➤ Ambovombe	➤ Betroka	➤ Bekily
➤ Androka	➤ Manakara	
➤ Basibasy	➤ Midongy	
➤ Cap Sainte Marie	➤ Vangaindrano	
➤ Farafangana		
➤ Tolagnaro		
➤ Toliara		

Voici la représentation de ces individus (lieux) sur la carte de Madagascar dans la partie Sud de Madagascar.

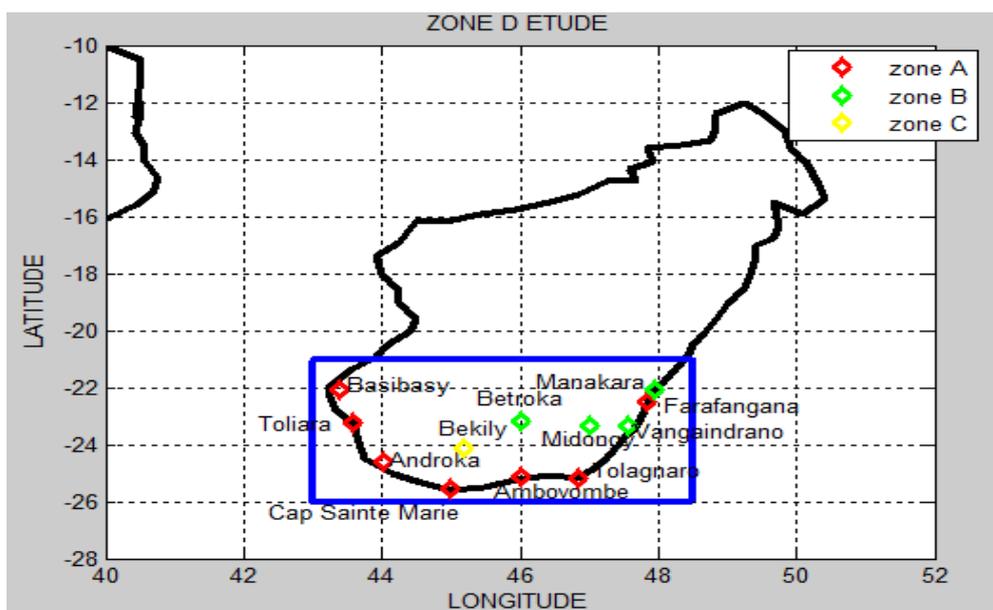


Figure 4 : Représentation des individus dans notre Zone d'étude

3-2. La Modélisation ARIMA (Zone A)

3-2-1. Données

La Zone A comprend sept (07) individus (*Ambovombe, Androka, Basibasy, Cap Sainte Marie, Farafangana, Tolagnaro et Toliara*). Pour la modélisation ARIMA de cette zone, nous avons fait la moyenne arithmétique de ces sept (07) individus et c'est notre donnée de départ pour le modèle.

3-2-2. Exposant de Hurst

L'exposant de Hurst est de $H = 0,5384$ ($H > 0,5$: la mémoire est à long terme). On peut donc faire une prévision des données mensuelles à long terme.

3-2-3. Vérification de la stationnarité

Nous avons remarqué que notre série n'est pas stationnaire, on a effectués trois différents tests [Dickey-Fuller Augmenté (ADF), KPSS et Philip-Pearson (PP)], dont les résultats sont affichés à la première ligne du **Tableau 4**.

3-2-3-1. Différenciation stationnaire

Quand on a une série non stationnaire, nous devons appliquer une première différenciation stationnaire, pour rendre la série stationnaire, et après on refait les trois tests (ADF, KPSS et PP), dans le but de voir si la série différenciée est stationnaire. Le **Tableau 4**, deuxième ligne nous montre qu'après différenciation, la série est devenue stationnaire.

Tableau 4 : Test de stationnarité

	Dickey-Fuller augmenté(ADF)	KPSS	Philip Pearson(PP)	Conclusion
Zone A	H = 0	H = 1	H = 0	Non stationnaire
Diff (Zone A)	H = 1	H = 0	H = 1	Stationnaire

Etant donné que la différenciation première de la série (Zone A) est stationnaire, donc $d = 1$. Les pics importantes aux décalages 12, 24, 36, 48, 60, etc. mois signifient la présence d'une saisonnalité de 12 mois dans notre série ($s = 12$) (voir **Figure 5** à droite).

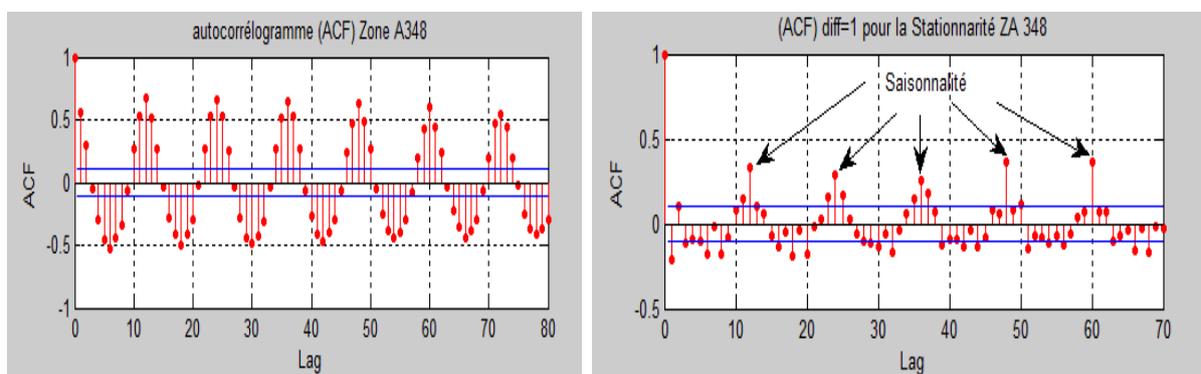


Figure 5 : Autocorrélation Zone A (à gauche) et diff Zone A (à droite)

3-2-3-2. Différenciation saisonnière

Nous devons travailler sur une série stationnaire et non saisonnière, dans ce cas, nous avons effectué une première différence saisonnière pour enlever la saisonnalité (12 mois) car à ce stade elle est déjà stationnaire. Et après les deux premières différenciations [stationnaire ($d = 1$) et saisonnière ($S = 12$)], on voit bien que notre série est à la fois stationnaire et non saisonnière. Une fois que d et S sont fixés ($d = 1$ et $S = 12$), il faut que nous cherchons les valeurs de p et q . Pour cela, il faut prendre le dernier pic qui sort de l'intervalle de confiance. La **Figure 6** suivante nous impose à prendre $p \leq 18$ (d'après la PACF) et que $q \leq 13$ (d'après la ACF).

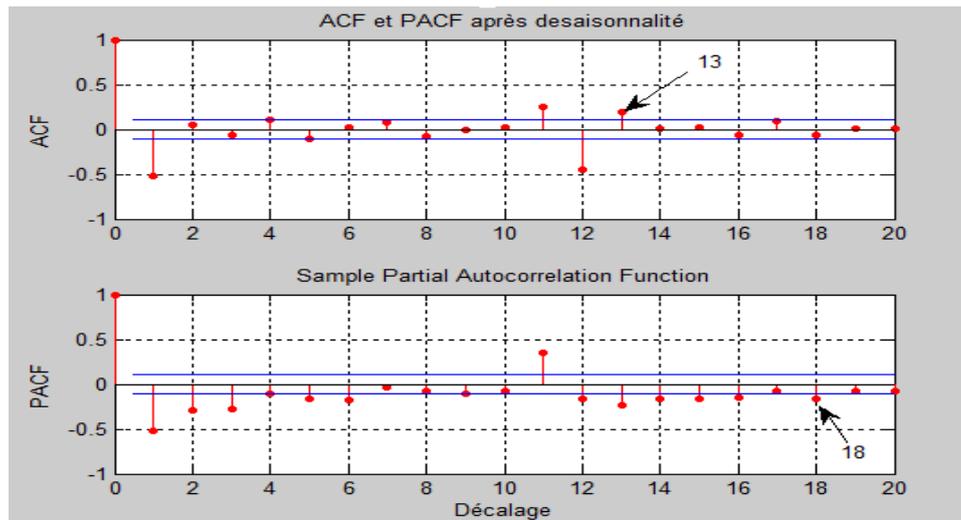


Figure 6 : Choix des paramètres p et q

Avec d et S fixes, il faut combiner les valeurs des p et q , comme l'indique le **Tableau 5** suivant.

Tableau 5 : Choix du modèle le mieux adapté

$q \backslash p$	0	1	2	...	p_{\max}
0	ARMA(0,0)	AR(1)	AR(2)	...	AR(p_{\max})
1	MA(1)	ARMA(1,1)	ARMA(2,1)	...	ARMA($p_{\max}, 1$)
2	MA(2)	ARMA(1,2)	ARMA(2,2)	...	ARMA($p_{\max}, 2$)
⋮	⋮	⋮	⋮	⋮	⋮
q_{\max}	MA(q_{\max})	ARMA(1, q_{\max})	ARMA(2, q_{\max})	...	ARMA(p_{\max}, q_{\max})

A chaque combinaison des paramètres (p , d , q), il faut calculer la valeur de Bayesian Information Criterion (B.I.C) et ne garder que le modèle qui a la plus petite valeur de B.I.C dont les résidus satisfont deux conditions : non corrélés et suivent la loi. Pour notre cas, la valeur de B.I.C est de 1703 ce qui correspond à SARIMA(5, 1, 7)(1, 1, 1)₁₂.

3-2-4. Les Résidus

La vérification des résidus (non corrélés et normales) est une des étapes très importantes pour l'ARIMA.

3-2-4-1. Test de non autocorrélation

3-2-4-1-1. Test de Ljung Box

Après avoir appliqué le test de Ljung-Box aux résidus, nous avons trouvé que $p_{\text{value}} = 0,7991$, et nous devons comparer ce p_{value} au taux de signification de 5 % (0,05).

⇒ Comme $p_{\text{value}} = 0,7991$ est largement supérieure à 0,05 donc les résidus sont non corrélés

3-2-4-1-2. Vérification des autocorrélogrammes

L'A.C.F et le P.A.C.F des résidus nous montrent qu'aucun pic ne sort de l'intervalle de confiance, ce qui veut dire que les valeurs des coefficients de corrélation des résidus sont significativement nulles. Ceci nous prouve que les résidus sont non auto corrélés. Le test de Ljung-Box est confirmé par des résultats de l'ACF et PACF de la **Figure 7** suivante :

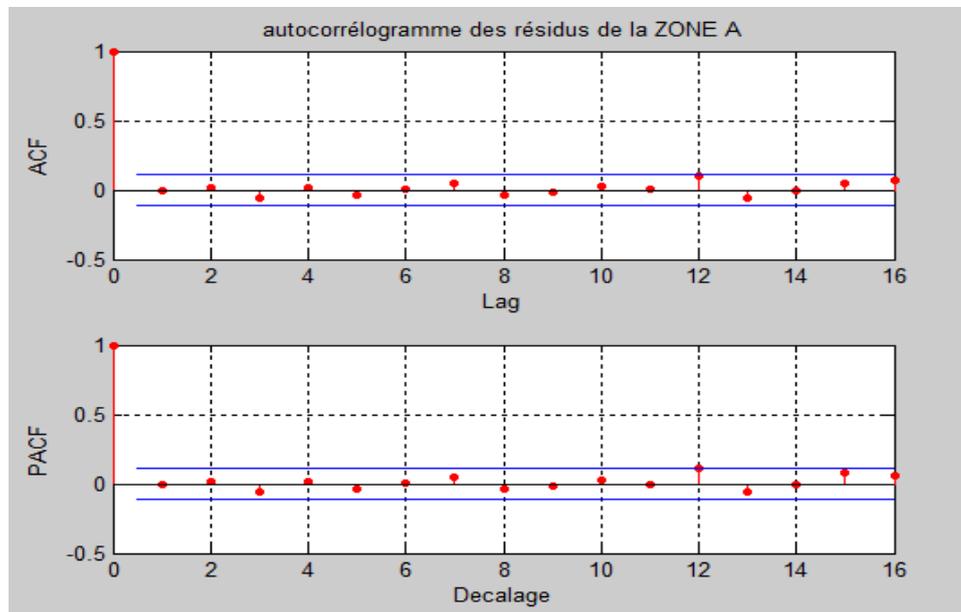


Figure 7 : Autocorrélogramme des résidus

3-2-4-2. Test de normalité

3-2-4-2-1. Test de Jarque-Bera

Ce test est utilisé pour vérifier la normalité ou non des résidus. La $p_{\text{value(Jarque-Bera)}} = 0,167$ est plus grande comparé au taux de signification de 0,05 (5 %), ce qui veut dire que les résidus suivent une loi normale.

3-2-4-2-2. Vérification par des figures

La parcelle quantile-quantile (QQ-plot) et l'estimation de la densité du noyau ne montrent aucune violation évidente de l'hypothèse de normalité. Ceci nous confirme que les résidus suivent une loi de normalité.

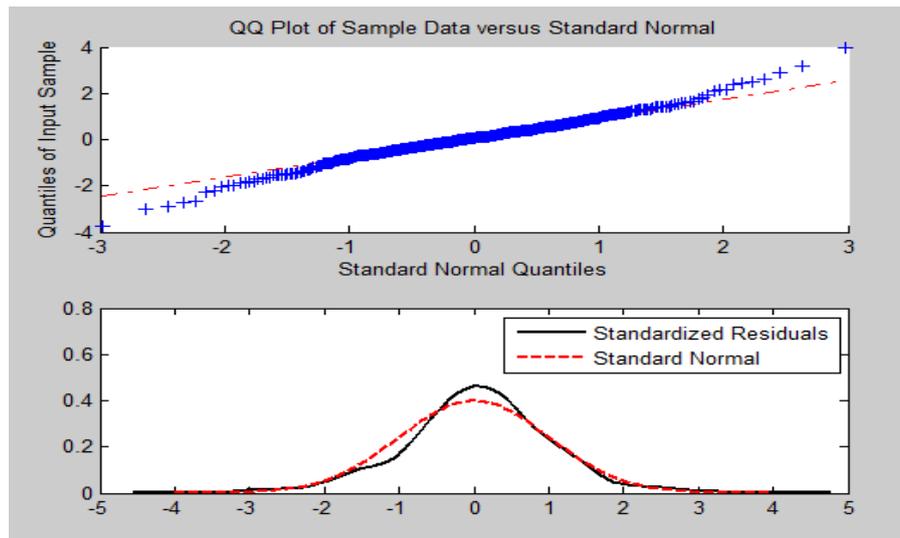


Figure 8 : Vérification de la normalité des résidus

Nous avons vu que nos résidus sont à la fois non auto corrélés et suivent une loi de normalité ce qui implique que notre résidus se comportent comme un bruit blanc, donc nous allons faire la prévision.

3-2-4-3. Validation et prévision

Pour notre cas, nous avons $N = 360$ mois et nous avons travaillé sur $N' = 348$ mois et nous avons laissé les 12 dernier mois (soit 1 an) pour la validation.

3-2-4-3-1. Validation

D'après la **Figure 9** suivante, la validation (en rouge) arrive bien à suivre notre donnée observée (en bleue) et ceci est confirmé par le **Tableau 6** suivant (l'erreur est plus ou moins faible).

Tableau 6 : Validation de notre modèle

Donnée Observée	18	22	24	25	30	30	31	31	29	28	24	23
Validation	11	21	28	24	29	30	29	29	30	28	27	25
Erreur (en valeur absolue)	7	1	4	1	1	0	2	2	1	0	3	2

3-2-4-3-2. Prévision

Après avoir calculé le pourcentage d'erreur absolue moyenne (en anglais : Mean Absolute Percentage Error ou M.A.P.E), on a pu trouver la valeur de $M.A.P.E = 0,02\%$ ($M.A.P.E \ll 10\%$) donc nous aurons une prévision excellente en utilisant notre modèle $SARIMA(5, 1, 7)(1, 1, 1)_{12}$.

Tableau 7 : Prévision pour les futurs mois

Année 2013	Janv	Fév	Mar	Avr	Mai	Juin	Jull	Aou	Sept	Oct	Nov	Déc
Prévision	10	21	28	23	29	30	28	29	29	28	27	25
Année 2014	Janv	Fév	Mar	Avr	Mai	Juin	Jull	Aou	Sept	Oct	Nov	Déc
Prévision	9	21	27	23	29	30	28	29	29	28	27	25

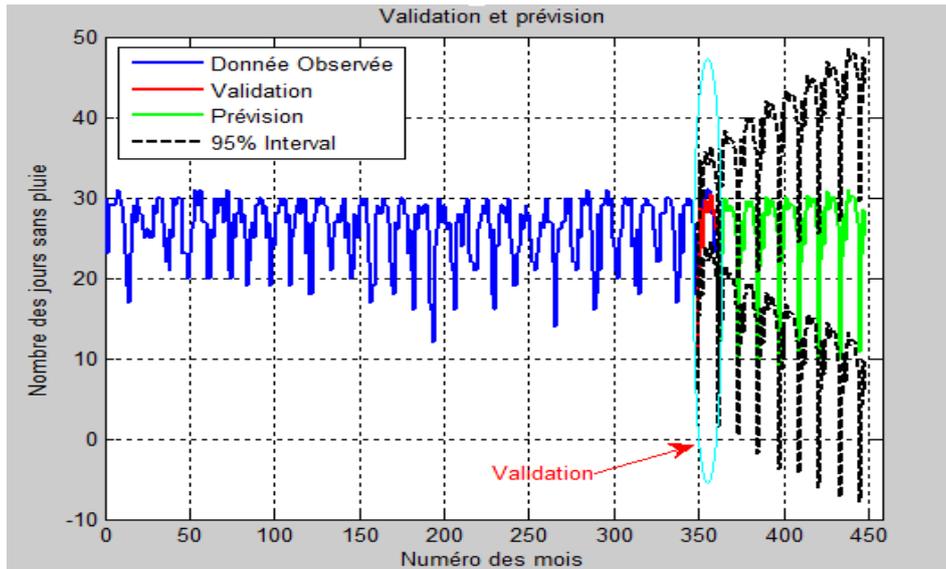


Figure 9 : Validation et Prédiction de la sécheresse (Zone A)

3-3. La corrélation de Spearman

Après avoir appliqué l'Équation 5 pour une corrélation non linéaire pour $N = 360$ (degré de liberté) et pour la valeur théorique du coefficient de Spearman ($t_{(Théorique)} = 1,96$) pour un degré de liberté supérieur à 120 ou infini (d'après Table de Student) et l'Équation 6 pour voir la significativité de $r_{Spearman(Observée)}$, nous avons eu les résultats regroupés dans le Tableau 8 suivants :

Tableau 8 : Corrélation des indices climatiques avec la sécheresse (Zone A)

	Corrélation de Spearman entre la Zone A et ...					
	Nino1_2	Nino3	Nino3_4	Nino4	A.A.O	I.O.D
ρ (Rho)	0.0620	0.1091	0.0815	-0.0101	-0.0254	-0.0083
$t_{Calculé}$	1.1758	2.0772	1.5465	-0.1912	-0.4812	-0.1578
$t_{Théorique}$	1,96	1,96	1,96	1,96	1,96	1,96
Conclusion	Non Significative	SIGNIFICATIVE	Non Significative	Non Significative	Non Significative	Non Significative

- L'A.A.O, l'I.O.D, Nino1_2, Nino3_4, Nino4 n'ont pas d'intercorrélations significatives sur la sécheresse (Zone A);
- La Nino3, est le seul indice climatique qui a une intercorrélations significative avec la sécheresse (Zone A) :

$$\left. \begin{array}{l} r_{Spearman(Observé)} = 0,1091 \\ t = r_s \sqrt{\frac{N - 2}{1 - r_s^2}} = 2,0772 \\ t_{(Théorique)} = 1,96 \end{array} \right\} \Rightarrow 2,0772 > 1,96 \Rightarrow r_{Spearman}(Nino\ 3\ et\ ZONE\ A)\ est\ significative$$

⇒ La Nino 3 et la sécheresse sont significativement liées, autrement dit, la Nino 3 a un impact sur la sécheresse (Zone A).

4. Conclusion

En conclusion, l'Analyse en Composante Principale (A.C.P) de la sécheresse dans notre zone d'étude nous a révélée qu'on peut subdiviser notre Zone d'étude en trois zones : Zone A (*Ambovombe, Androka, Basibasy, Cap Sainte Marie, Tolagnaro, Toliara et Farafangana*), Zone B (*Betroka, Manakara, Midongy et Vangaindrano*) et la Zone C (*Bekily*). Nous avons fait une excellente prévision de la sécheresse en appliquant la modélisation SARIMA (5, 1, 7) (1, 1, 1)₁₂. Parmi les indices climatiques [A.A.O, I.O.D et ENSO (Nino1_2, Nino3, Nino3_4 et Nino4)], seul la Nino 3 qui a un impact sur la sécheresse (Zone A), selon la corrélation de Spearman [16].

Références

- [1] - JEAN EUGENE RANDRIANANTENAINA, "La pluviométrie et la circulation du vent zonal dans la région de l'extrême sud-est de Madagascar", Université de Toliara, (Novembre 2011)
- [2] - RABEFITIA ZOHARIMALALA, RAHOLIJAQ NIRIVOLOLONA, "Climatologie" Antananarivo Madagascar, (Novembre 1992)
- [3] - ARNAUD MARTIN, "L'analyse de données", (Sept 2004)
- [4] - MICHEL RAYNAUD, Statistiques Appliquées, "Cours d'Analyse en Composante Principale"
- [5] - G. SAPORTA, "Probabilités, analyses des données et statistiques". Dunod, Paris (1990) et "Probabilités, analyse des données et analyse des données et statistique". 2^{ème} édition, Technip, (2006)
- [6] - TODIHASINA ROSELIN RANDRIANANTENAINA, "Impact du réchauffement global sur les risques de canicule et de sécheresse dans la région du BOINA", Université d'Antananarivo, (Juin 2017)
- [7] - Dominique DESBOIS, "Une introduction à la méthodologie de Box et Jenkins", *Revue MODULAD*, (2005)
- [8] - NIRY HAVANA RAZANATOMPOHARIMANGA, "Prévision des périodes de pluies (début et fin) dans la partie Centre-Ouest de Madagascar", (Décembre 2015)
- [9] - F. MOUNA, "Persistance de la performance des fonds de couverture: Approche non paramétrique", (2008)
- [10] - W. C. NEWBY & K. WEST, "A simple Positive-definite: Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Vol. 55, (1987) 703 - 708 p.
- [11] - BERNARD RAPACCHI, "Les séries chronologiques", (Aout 1993)
- [12] - <http://www.cons-dev.org/elearning/ando/06/61/61.html>, (Avril 2017)
- [13] - RICO-RAKOTOMALALA, "Analyse de corrélation : étude des dépendances variables quantitatives", 4^{ème} Edition, (2015)
- [14] - HAJANIRINA RICHARD RAMIANDRISOA, Mémoire de D.E.A : "la description de l'indice climatique ENSO, et son influence sur la précipitation dans la partie sud-ouest de Madagascar", Université de Toliara, (Mars 2013)
- [15] - JOURNAL OF GEOPHYSICAL RESEARCH, Vol. 111, D14102, doi :10.1029/2005 JD006289, (2006)
- [16] - http://www.nature.com/articles/srep00269?message-global=remove&WT.mc_id=FBK_SciReports, (Mars 2017)