

TABLE DES MATIERES

TABLE DES MATIERES	i
LISTE DES ACRONYMES	v
NOMENCLATURE.....	vi
LISTE DES FIGURES	vii
LISTE DES TABLEAUX.....	viii
LISTE DES ANNEXES	ix
INTRODUCTION.....	1
PARTIE I : GENERALITES.....	3
I.1. Température	3
I.1.1. Définition	3
I.1.2. Echelle	3
I.2. Homogénéisation	3
I.2.1. Méthode	4
I.2.2. Métadonnée.....	4
I.2.3. Rupture.....	4
I.2.3.1. Saut de moyenne	4
I.2.3.1. Tendance	5
I.2.4. Série de base	5
I.2.5. Séries de référence	5
I.2.6. Série homogène.....	5
I.2.7. Série hétérogène.....	6
I.3. Test statistique	6
I.3.1. Généralité	6
I.3.2. Quelques définitions	6

I.3.3. Etape du test statistique.....	6
I.3.4. Tests paramétriques et tests non paramétriques.....	7
I.3.4.1. Tests paramétriques	7
I.3.4.2. Tests non paramétriques	7
I.4. Exposant de Hurst.....	7
I.5. Modélisation ARIMA des séries chronologique	8
I.5.1. Box & Jenkins.....	8
I.5.2. Les processus ARMA	9
I.5.3. Les processus ARIMA.....	9
I.6. Moyenne des Erreurs Absolues en Pourcentage	10
PARTIE II: MATERIELS ET METHODES.....	11
II.1. Matériels	11
II.1.1. Localisation de la zone d'étude	11
II.1.2. Données utilisées	11
II.1.3. Outils informatique.....	11
II.1.4. Outils statistiques.....	12
II.1.4.1. Test de Buishand	12
II.1.4.2. Test de SNHT	12
II.1.4.2.1. Test pour un saut	13
II.1.4.2.2. Test pour une tendance	13
II.1.4.3. Test de Mann Kendall	14
II.1.4.4. Test de normalité de Shapiro-Wilk	15
II.2. Méthodes	16
II.2.1. Préparation des données	16
II.2.2. Homogénéisation des données	16
II.2.2.1. Homogénéité absolue	16
II.2.2.2. Homogénéité relative.....	17

II.2.2.2.1. Première approche : à partir de l'homogénéité absolue et corrélation	17
II.2.2.2.2. Deuxième approche : à partir des séries voisines initiales	18
II.2.2.2.3. Test d'homogénéité relative des séries de base	18
II.2.3. Recherche de meilleure méthode d'homogénéisation.....	20
II.2.4. Zonage et modélisation.....	20
PARTIE III: RESULTATS ET INTERPRETATIONS.....	21
III.1. Test de normalité.....	21
III.2. Test d'homogénéité absolue.....	21
III.2.1. Test de Buishand	21
III.2.2. Test de Mann Kendall	22
III.3. Ajustement	22
III.4. Résultat final d'homogénéité absolue après ajustement	26
III.5. Homogénéité relative	27
III.5.1. Première approche.....	27
III.5.1.1. Série avec un saut.....	27
III.5.1.2. Deux sauts	28
III.5.1.3. Tendance	29
III.5.1.4. Saut et tendance.....	30
III.5.2. Deuxième approche.....	32
III.6. Résultat de l'analyse en composante principale.....	32
III.7. Modélisation.....	36
III.7.1. Série moyenne de la sous zone A.....	36
III.8. Modélisation ARIMA	37
III.8.1. Etude de stationnarité de la série.....	37
III.8.2. Identification de modèle.....	38
III.8.2. Estimation du modèle.....	40
III.8.4. Prévision.....	42

II.9. Discussion :	46
CONCLUSION.....	48
REFERENCES BIBLIOGRAPHIQUES.....	50
ANNEXES	53

LISTE DES ACRONYMES

ACP	:	Analyse en Composantes Principales
AR	:	AutoRegressive
ARMA	:	AutoRegressive Moving Average
ARIMA	:	AutoRegressive Integrated Moving Average
FAC	:	Fonction d'Autocorrélation
FACP	:	Fonction d'Autocorrélation Partiel
ECMWF	:	European Climatic Medium Range Weather Forecast (Centre Européen pour les Prévisions Météorologiques à Moyen-Terme)
MA	:	Moving Average
MAPE	:	Mean Absolute Percentage Error
MEAP	:	Moyenne des Erreurs Absolues en Pourcentage
MATLAB	:	Matrix Laboratory
SNHT	:	Standard Normal Homogeneity Test

NOMENCLATURE

Lettres latines

F1	:	première axe factoriel
F2	:	deuxième axe factoriel
H	:	horizon de prévision
n	:	nombre de mesures
H	:	exposant de Hurst
K	:	Kelvin
°C	:	degré Celsius
°F	:	degré Fahrenheit
T_a^{\max}	:	statistique du test de SNHT pour un saut
T_{ab}^{\max}	:	statistique du test de SNHT pour une tendance
μ_0	:	moyenne de la série
μ_1	:	moyenne de la série avant la date de rupture
μ_2	:	moyenne de la série après la date de rupture
μ_3	:	moyenne entre les dates de ruptures

Lettres grecques :

σ_j	:	écart type de x_j
σ_j^2	:	variance de la variable x_j
ρ_j	:	coefficient de corrélation entre la série de base et la série de référence j
ε_t	:	bruit blanc

LISTE DES FIGURES

Figure 1 : Localisation de la zone d'étude.....	11
Figure 2 : Organigramme du test d'homogénéité absolue.....	17
Figure 3 : Organigramme du test d'homogénéité relative.....	19
Figure 4 : Résultats du test d'homogénéité absolue.....	22
Figure 5 : Exemple de série hétérogène	23
Figure 6 : Deuxième ajustement de la série C2	24
Figure 7 : Exemple de séries homogènes	25
Figure 8 : Résultats du test d'homogénéité relative première approche	27
Figure 9 : Exemple de série avec un saut.....	28
Figure 10 : Exemple de série avec 2 sauts	29
Figure 11 : Exemple de série avec une tendance	30
Figure 12 : Exemple de série avec un saut et une tendance.....	31
Figure 13 : Exemple de série homogène	31
Figure 14 : Résultats du test d'homogénéité relative deuxième approche.....	32
Figure 15 : Représentation des variables dans le plan (F1, F2)	33
Figure 16 : Coordonnées des individus.....	34
Figure 17 : Division de la zone d'étude en sous zone	35
Figure 18 : Série moyenne de la sous zone A.....	36
Figure 19 : Courbe d'autocorrélation et intervalle de confiance des séries de la sous zone A	37
Figure 20 : Séries après différenciation de premier ordre.....	38
Figure 21 : Correlogramme de la fonction d'autocorrélation simple et partielle des série moyenne.....	39
Figure 22 : Correlogramme de la fonction d'autocorrélation simple et partielle de résidus	41
Figure 23 : Prévision des séries de la sous zone A	43
Figure 24 : Prévision finale de la série absolue de la sous zone A.....	44
Figure 25 : Prévision finale de la série absolue de la zone entière	45

LISTE DES TABLEAUX

Tableau 1 : Valeurs remarquables de la température.....	3
Tableau 2 : Classification de la qualité de prévisions.....	10
Tableau 3 : Position de série de base et des séries de références.....	18
Tableau 4 : Résultats des tests de Buishand	21
Tableau 5 : Facteur de correction des séries hétérogènes	22
Tableau 6 : Résultat du test d'homogénéité absolue après ajustement.....	26
Tableau 7 : Corrélation de la série de base avec la série de référence.....	26
Tableau 8 : Série avec un saut.....	27
Tableau 9 : Série avec 2 sauts.....	28
Tableau 10 : Série avec une tendance	29
Tableau 11 : Série avec un saut et une tendance.....	30
Tableau 12 : Valeur propre	32
Tableau 13 : Exposant de Hurst et effet mémoire des séries de la sous zone A.....	36
Tableau 14 : Résultat du test de stationnarité	37
Tableau 15 : Résultat du test de stationnarité après différenciation	38
Tableau 16 : Modèle possible pour la modélisation	40
Tableau 17 : Modèles estimés.....	40
Tableau 18 : Corrélation des résidus	42
Tableau 19 : Résultat des tests de normalités des résidus.....	42
Tableau 20 : Résultat des tests d'homoscédasticités des résidus.....	42
Tableau 21 : Valeurs des prévisions de la température moyenne.....	45

LISTE DES ANNEXES

<i>ANNEXE A : TABLEAU DES VALEURS CRITIQUES D'ALEXANDERSSON</i>	<i>i</i>
<i>ANNEXE B : TEST DE NORMALITE</i>	<i>i</i>
<i>ANNEXE C : TEST DE MANN KENDALL</i>	<i>i</i>
<i>ANNEXE D : COSINUS CARRES DES VARIABLES.....</i>	<i>ii</i>
<i>ANNEXE E : MODELISATION DES SERIES MOYENNES DES TEMPERATURES DE LA ZONE ENTIERE</i>	<i>iii</i>

INTRODUCTION

La climatologie est une discipline scientifique dynamique à usages multiples dans un large champ d'application. De nouvelles techniques sont mises au point, des études sont entreprises pour savoir comment le climat s'applique à de nombreux domaines, notamment l'agriculture et l'énergie [1]. Elle étudie les phénomènes atmosphériques à diverses échelles ; comme l'échelle d'une région, d'un pays, ou de la planète entière. Elle utilise : des relevés de température, de précipitations, de vents, de pression sur une période d'au moins trente ans [2]. Ces relevés sont généralement obtenus à partir des stations de mesures.

Les conditions de mesure d'une station peuvent varier au cours du temps. Les modifications des conditions de mesure (déplacement de stations, remplacement d'instruments de mesure, changement des heures d'observations ou encore modification de l'environnement immédiat de l'instrument de mesure) peuvent introduire des ruptures artificielles dans les données qui ne reflètent pas les variations réelles du climat [3].

Par conséquent, il se peut que des décisions soient prises en se basant sur des données qui contiennent des erreurs [4]. Pourtant, c'est justement cela qu'il faut éviter, d'où la nécessité de l'homogénéisation des données.

L'homogénéisation est un processus de détection et de correction des ruptures (qu'on appelle également inhomogénéité) d'origine non climatique. En homogénéisation, on compare souvent les données d'une station avec celles des stations voisines afin d'éviter qu'un changement climatique ne soit interprété comme une inhomogénéité [3].

Plusieurs techniques existent pour homogénéiser les données de température : méthode des doubles accumulations, analyse graphique des séries de différences, méthode de Tayanc, régression multiple, régression multiphasé, méthode d'Alexandersson, test de Student séquentiel, méthode de Boroneant et Tomozeiu, et la méthode de Datsenko [4].

Beaulieu et al [3][4] ont généré des séries synthétiques afin de vérifier la performance et la robustesse de quelques méthodes d'homogénéisations. La méthode d'Alexandersson en fait partie. Pour cette méthode d'Alexandersson, la consultation

de l'information historique est complémentaire et le nombre de séries de référence est libre au choix de l'utilisateur.

Il y a deux tests distincts pour les deux types d'inhomogénéités : test pour un saut et test pour une tendance. Les sauts sont facilement ajustables alors que les mathématiques du test pour les tendances sont beaucoup plus compliquées que pour les sauts. Les deux types s'automatisent facilement [4].

Pour cela, ce travail de mémoire est axé sur l'homogénéisation des données et prévisions qui s'intitule : « homogénéisation des données de températures moyenne dans la région d'Atsimo Andrefana de Madagascar. Modèle de prévision de ces températures ».

Dans cette étude, nous allons utiliser des données des températures. Et nous nous demandons si ces données de températures qu'on va utiliser contient des ruptures ; si ces ruptures sont d'origine climatique ; et si on peut avoir des modèles fiables pour ces températures ?

Nous avons décidé de diviser le travail en trois (3) grandes parties, afin d'atteindre l'objectif, qui est d'avoir des données et des prévisions fiables :

- La première partie énoncera quelques généralités concernant : la température, l'homogénéisation et le modèle de prévision ARIMA.
- La seconde partie comprendra la localisation de la zone d'étude, suivi de l'évaluation des outils utilisés et les explications des procédures de traitements.
- La troisième partie consistera à la présentation, interprétation et analyse des résultats obtenus.

PARTIE I : GENERALITES

I.1. Température

I.1.1. Définition

Etant une grandeur physique, la température est mesurée à l'aide d'un thermomètre et étudiée en thermométrie. Selon les conditions atmosphériques (vent, pluie, ensoleillement...), sa perception varie d'un individu à l'autre [5]. En générale, elle est reliée aux sensations de froid et de chaud, provenant du transfert thermique entre le corps humain et son environnement. En physique, c'est une variable d'état qui, du point de vue de la structure de la matière, caractérise le degré d'agitation de ses particules. C'est toujours la température d'une boule au voisinage d'un point que l'on mesure, et que l'on désigne par « température en ce point ».

I.1.2. Echelle

Plusieurs échelles ont été définies pour déterminer la grandeur physique de la température. Telle l'échelle Celsius ($^{\circ}\text{C}$), Fahrenheit ($^{\circ}\text{F}$) et kelvin (K).

La majorité des pays utilise le degré Celsius. Le 0°C de cette échelle correspond au point de congélation de l'eau, et le 100°C au point d'ébullition de l'eau [1].

Le kelvin est l'unité légale du système internationale.

Tableau 1 : Valeurs remarquables de la température suivant l'échelle Kelvin, Celsius et Fahrenheit

Quelques valeurs remarquables			
Description	Echelle		
	Kselvin	Celsius	Fahrenheit
Zéros absolu (température la plus basse dans l'univers)	0 K	-273,15 $^{\circ}\text{C}$	-459,67 $^{\circ}\text{F}$
Température de fusion de l'eau (à la pression standard)	273,15 K	0 $^{\circ}\text{C}$	32 $^{\circ}\text{F}$
Température de fusion de l'eau (à la pression standard)	373,15 K	100 $^{\circ}\text{C}$	212 $^{\circ}\text{F}$

I.2. Homogénéisation

Comme nous avons déjà précisé dans l'introduction, l'homogénéisation est un processus de détection et de correction des ruptures d'origine non climatique. Une meilleure approche d'homogénéisation est de considérer l'information livrée par plusieurs stations voisines en évaluant la série à homogénéiser.

I.2.1. Méthode

Les techniques d'homogénéisations varient selon l'objectif pour lesquels on les applique et la philosophie de chaque équipe de travail.

Malgré la grande diversité des méthodes d'homogénéisations, elles peuvent néanmoins être classifiées en deux catégories principales :

- *Méthodes subjectives* :

Lorsque l'emplacement d'une discontinuité est détecté à l'œil nu sur un graphique, la méthode appartient à la classe subjective même si des tests statistiques sont appliqués par la suite. Par exemple : l'analyse des doubles accumulations, les méthodes basées sur les déviations cumulées et l'analyse graphique des déviations cumulées.

- *Méthodes objectives* :

Quant aux méthodes objectives, elles ne dépendent pas du jugement de l'utilisateur pour localiser les inhomogénéités. Exemple : régression multiple, méthode de Jaruskova et méthode d' Alexandersson (SNHT).

I.2.2. Métadonnée [3]

La source d'information la plus importante pour appuyer toutes les méthodes existantes de détection des inhomogénéités dans des séries climatiques provient de la consultation des métadonnées. Elles sont formées des archives historiques propres à chaque station. Les métadonnées contiennent les enregistrements de la station, des annuaires météorologiques, des fiches d'inspection, des photographies de la station et de son environnement, etc. Une entrevue avec la personne responsable d'une station constitue également une source d'information. Les études comparatives réalisées sur les instruments peuvent aussi donner une bonne idée de l'effet d'un changement d'instrument dans une série de données.

I.2.3. Rupture [3]

Une rupture désigne une modification subite dans les propriétés d'un processus aléatoire. Elle suppose que ses propriétés sont de part et d'autre de l'année de rupture. Il existe deux types de d'inhomogénéités :

I.2.3.1. Saut de moyenne

Les sauts dans une série de données climatiques peuvent être occasionnés par plusieurs types de changements :

- changement d'instrument,

- déménagement d'une station,
- changement de méthode de calcul des statistiques comme les moyennes mensuelles,
- changement de technicien
- changement des heures de mesure.

Les sauts abrupts sont faciles à détecter tandis que les sauts de faibles amplitudes sont plus problématiques.

I.2.3.1. Tendance

Une tendance (au sens mathématique) correspond à une certaine orientation prise par les valeurs d'une série de données en fonction du temps, c'est-à-dire à une évolution du processus observé en fonction du temps.

Elle peut être causée par :

- une modification dans l'environnement immédiat de la station (urbanisation et industrialisation).
- reforestation graduelle autour de la station.
- vieillissement de l'appareil de mesure.

Une tendance est plus délicate à quantifier car il faut identifier correctement son début, sa fin et ainsi que son amplitude.

I.2.4. Série de base

La série qui peut potentiellement contenir des inhomogénéités est appelée : série de base.

I.2.5. Séries de référence

Une série de référence est une fonction d'une ou plusieurs stations climatiquement similaires à la station de base. Le rôle de telles séries est d'empêcher qu'une variation climatique régionale soit classifiée comme une inhomogénéité. La plupart des méthodes d'homogénéisation incluent une ou plusieurs séries voisines pour différencier les changements du climat régional des inhomogénéités observées à la station de base.

I.2.6. Série homogène

Une série est homogène si ces variations répondent exclusivement aux causes climatiques.

I.2.7. Série hétérogène

Une série est hétérogène si elle contient des ruptures d'origine non climatique.

I.3. Test statistique [6][7]

I.3.1. Généralité

Le test statistique est le moyen de comparer les deux hypothèses H_0 et H_1 . C'est une valeur numérique calculée à partir de la série de données testée (on calcule la valeur d'une certaine variable appelée la « statistique » du test). Le test permet de choisir, parmi les deux hypothèses émises, celle qui a le plus de chances d'être vraie. Toutefois, en statistique, les hypothèses ne sont jamais « acceptées ». Il est préféré l'expression « on ne peut rejeter l'hypothèse H_0 à un niveau de confiance » ou bien, dans le cas de l'alternative, « on doit rejeter l'hypothèse H_0 et retenir l'hypothèse H_1 à un niveau de confiance ».

I.3.2. Quelques définitions

En considérant une hypothèse nulle (privilégiée) H_0 , une alternative H_1 , et en bâtissant une règle permettant de décider de rejeter ou pas H_0 . L'erreur de première espèce est de rejeter H_0 alors qu'elle est vraie. L'erreur de seconde espèce est de ne pas rejeter H_0 alors qu'elle est fausse.

- α est la probabilité de rejeter à tort H_0 (risque de première espèce).
- β est la probabilité de ne pas rejeter H_0 alors que H_1 est vraie (risque de première espèce).
- la région d'acceptation du test représente un intervalle regroupant $(1-\alpha)\%$ des valeurs que peut prendre la statistique si l'hypothèse H_0 est vraie et son complémentaire c'est la région critique.
- la p valeur (p-value) c'est la probabilité que la statistique théorique sous H_0 soit supérieure à la statistique calculée (expérimentale).

I.3.3. Etape du test statistique

- 1- Etablir l'hypothèse nulle H_0 et l'hypothèse alternative H_1 .
- 2- Choisir la statistique du test,
- 3- Choisir la valeur du risque de première espèce α ,
- 4- Déterminer la région critique en fonction de α
- 5- Calculer la valeur expérimentale de la statistique.
- 6- Conclure : rejet ou validation de l'hypothèse nulle H_0

I.3.4. Tests paramétriques et tests non paramétriques

I.3.4.1. Tests paramétriques

Les tests paramétriques présument que les données des séries chronologiques suivent une distribution particulière (généralement une distribution normale). *Cette condition doit être remplie pour pouvoir appliquer ces tests.* Ils sont utiles car ils permettent de quantifier les changements dans les données (amplitude du changement de moyenne ou gradient de tendance) et sont *généralement plus puissants que les tests non-paramétriques.*

I.3.4.2. Tests non paramétriques

Les tests non-paramétriques ne nécessitent pas d'hypothèse sur la distribution des données des séries chronologiques. Ces tests permettent de détecter des tendances ou des changements mais ne sont généralement pas appropriés pour quantifier les pentes des tendances ou l'amplitude des changements.

I.4. Exposant de Hurst [8][9]

L'exposant de Hurst permet de classifier les séries temporelles en fonction de la nature de leur mémoire (structure de dépendance). Elle se définit comme l'étendue des sommes partielles des écarts d'une série temporelle à sa moyenne divisée par son écart type. C'est une mesure de persistance absolue.

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

$$Q_T = R / S_T$$

$$\text{Où } R = \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T)^2 - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T)^2 ,$$

$$S_T = \left[\frac{1}{T} \sum_{j=1}^k (X_j - \bar{X}_T)^2 \right]^{\frac{1}{2}}$$

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 :

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

I.5. Modélisation ARIMA des séries chronologique [10][11]

p étant l'ordre de processus autorégressif, d l'ordre d'intégration et q l'ordre de moyenne mobile. Les modèles ARIMA(p, d, q) permettent de combiner trois types de processus temporels :

- ***Processus autorégressifs AR(p)*** (AR-AutoRegressive),

Un processus (X_t) est dit autorégressif d'ordre p , AR(p), si l'observation présente X_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période sous la forme suivante :

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t$$

α_i $i=0, \dots, p$ sont des paramètres positifs ou négatifs à estimer.

(ε_t) est un bruit blanc.

- ***Processus intégrés (I-Integrated)***,

L'ordre d'intégration d est le nombre de fois que la série initiale a été différenciée pour obtenir la stationnarité.

- ***Moyennes mobiles MA(q)*** (MA-MovingAverage).

Dans un processus (X_t) de moyenne mobile d'ordre q , chaque observation X_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période dans le passé.

$$X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

I.5.1. Box & Jenkins [10][12]

Identification :

Toute composante saisonnière étant supposée éliminée, l'identification consiste à spécifier les trois paramètres p , d , q du modèle ARIMA (p, d, q). D'abord, on teste la stationnarité du modèle. Ensuite, on effectue tour à tour : une étude graphique, de corrélogrammes. Si la série n'est pas stationnaire, alors, il convient de la transformer (en général par différenciation) pour obtenir une série stationnaire.

En ayant une série stationnaire, on analyse la fonction d'autocorrélation (FAC) et la fonction d'autocorrélation partielle (FACP) afin de déterminer les paramètres p, q du modèle.

-Estimations :

Le logiciel Matlab permet d'estimer les coefficients du modèle identifiés dans l'étape précédente, les paramètres p, q et d étant clairement spécifiés.

-Validité du modèle :

Il faut que les résidus soient non corrélés, homoscédaste et normalement distribué de moyenne et variance constante.

Les caractéristiques des résidus doivent correspondre à celles d'un bruit blanc. La statistique Q de Box et Ljung, connu encore sous le nom de statistique de Box et Pierce modifiée est couramment utilisée pour tester le bruit blanc.

-Prévision :

On cherche à prévoir la valeur que va prendre la variable aléatoire X_{n+h} , et pour cela on s'appuie sur les variables X_n, X_{n-1}, \dots que l'on suppose avoir été observées. On dit que h est l'horizon de la prévision.

I.5.2. Les processus ARMA (p,q) (Autorégressive Moving Average)

Les modèles ARMA sont représentatifs de processus générés par une combinaison des valeurs passées et des erreurs passées.

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.1)$$

On peut aussi écrire le modèle ARMA (p,q) sous la forme :

$$\alpha(L)X_t = \Theta(L)\varepsilon_t \quad (1.2)$$

où L est l'opérateur de décalage et (ε_t) est le processus de bruit blanc.

I.5.3. Les processus ARIMA (p,d,q)

Un processus (X_t) est dit ARIMA (p,d,q), p,d,q positifs ou nuls si le processus $(1-L)^d X_t$ est un processus ARMA(p,q) stationnaire.

On peut écrire le modèle ARIMA (p,d,q) sous la forme :

$$\alpha(L)(1-L)^d X_t = \Theta(L)\varepsilon_t \quad (1.3)$$

I.6. Moyenne des Erreurs Absolues en Pourcentage (M.E.A.P.) ou Mean Absolute Percentage Error (MAPE) [13]

Le MAPE est utilisée pour comparer les prévisions le long de différentes séries temporelles. C'est un modèle qui permet de suggérer une meilleure prévision.

Elle est définie par la formule :

$$MAPE_n = \frac{100}{n} \sum_{t=1}^n \left| \frac{E_t}{D_t} \right| \quad (2)$$

E_t l'erreur de la prévision définie par la formule suivante :

$$E_t = F_t - D_t$$

F_t : valeur exacte

D_t : valeur estimé

Le tableau 2 ci-dessous montre la classification de la qualité de prévisions selon les valeurs de MAPE.

Tableau 2 : Classification de la qualité de prévisions [89]

MAPE _n (en %)	Qualité de la prévision
<10	Excellent
> 10 et <20	Bon
> 20 et <30	Moyen
>30	Mauvais

PARTIE II: MATERIELS ET METHODES

II.1. Matériels

II.1.1. Localisation de la zone d'étude

Notre zone d'étude est la région Sud-Ouest de Madagascar, elle s'étend entre les longitudes $42,5^{\circ}\text{E}$ à $45,5^{\circ}\text{E}$ et les latitudes $21,5^{\circ}\text{S}$ à $26,5^{\circ}\text{S}$, noté zone 1 (Fig.1).

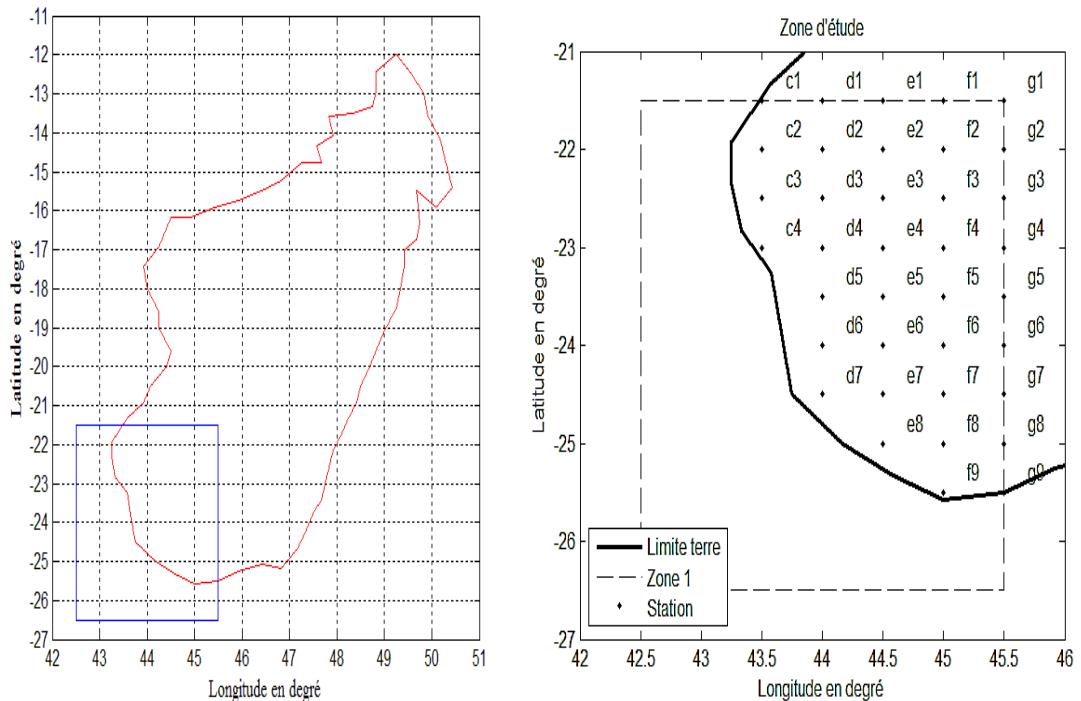


Figure 1 : Localisation de la zone d'étude

II.1.2. Données utilisées

Les données sont des températures journalières de l'année 1979 à 2015. Elles ont été téléchargées sur le site « www.ecmwf.int », avec une extension netcdf et un pas de grille de $0,5^{\circ}\times 0,5^{\circ}$. Ce sont des données de réanalyses. L'unité de la température est le kelvin « K ».

II.1.3. Outils informatique

Quatre logiciels sont utilisés :

- Logiciel Matlab pour la programmation.
- Logiciel Excel, pour faciliter la manipulation des données et des représentations graphiques.
- Logiciel Xlstat, pour les tests de tendance, de stationnarité, de normalité et l'analyse en composante principale.

- Logiciel AnClim afin de vérifier des résultats obtenus avec Matlab sur les tests de ruptures.

II.1.4. Outils statistiques

II.1.4.1. Test de Buishand [15][16][17][18][19]

Le test de Buishand est un test paramétrique .Il est appliqué au chronique X_1, X_2, \dots, X_N disposant d'au moins 10 valeurs et dont la distribution est normale.

- L'hypothèse nulle H_0 est l'absence de rupture.
- L'hypothèse alternative H_1 est la présence de rupture au temps t qui correspond à la valeur de U ci-dessous.

Ce test est basé sur les écarts cumulés à la moyenne jusqu'au rang k .

$$U = \sum_{k=1}^{N-1} \frac{\left(\frac{S_k}{D_x}\right)^2}{N(N+1)} \quad (3.1)$$

Avec

$$S_k = \sum_{i=1}^k (X_i - \bar{X}) \quad (3.2)$$

Et

$$D_x = \sum_{i=1}^N \frac{(X_i - \bar{X})^2}{N} \quad (3.3)$$

\bar{X} est la moyenne de la série.

Des valeurs critiques de la statistique U ont été données par Buishand en 1982 et en 1984.

II.1.4.2. Test de SNHT [4][20][21][22]

La méthode d'Alexandersson ou plus communément appelée SNHT (standard normal homogénéité test) a été développé pour homogénéiser des séries de précipitations totales annuelles ou de températures moyennes annuelles. Une ou plusieurs séries voisines homogènes et complètes sont utilisées pour créer une série de différences :

$$Q_i = Y_i - \left\{ \left[\sum_{j=1}^k \rho_j^2 [X_{ji} - \bar{X}_j + \bar{Y}_j] \right] / \sum_{j=1}^k \rho_j^2 \right\} \quad (4)$$

Y : série de base, X_{ji} : j-ième série de référence, Y_i : valeur de l'année i de Y

X_{ji} : valeur de l'année i de X_j et ρ_j : coefficient de corrélation entre la série de base et la série de référence j .

Le test de SNHT est appliqué à la série standardisée Z_i :

$$Z_i = (Q_i - \bar{Q}) / \sigma_Q \quad (4.1)$$

II.1.4.2.1. Test pour un saut

Sous l'hypothèse nulle, les différences standardisées Z_i , sont distribués normalement avec une moyenne nulle et un écart-type de 1. La contre-hypothèse est qu'il y a un changement de moyenne à partir d'un certain point, noté par a , dans la série de base.

$$H_0: Z_i \sim N(0,1) \quad i \in \{1, \dots, n\}$$

$$H_1: \begin{cases} Z_i \sim N(\mu_1, 1) & i \in \{1, \dots, a\} \\ Z_i \sim N(\mu_2, 1) & i \in \{a+1, \dots, n\} \end{cases}$$

La statistique du test est :

$$T_a^{\max} = \max_{1 \leq a \leq n-1} \{a\bar{z}_1^2 + (n-a)\bar{z}_2^2\} \quad (5)$$

\bar{z}_1 et \bar{z}_2 sont les moyennes des différences standardisées avant et après le saut respectivement et n représente le nombre total de l'année dans la série. Si la statistique T_a^{\max} dépasse sa valeur critique alors il semble y avoir un saut de moyenne au temps a . L'amplitude de ce saut est estimée par la différence des moyens avant et après le saut $\bar{q}_2 - \bar{q}_1$. La série se corrige en ajoutant le premier segment par ce différence.

$$\bar{q}_1 = \sigma_Q \bar{z}_1 + \bar{Q} \quad (5.1)$$

$$\bar{q}_2 = \sigma_Q \bar{z}_2 + \bar{Q} \quad (5.2)$$

II.1.4.2.2. Test pour une tendance

$$H_0: Z_i \sim N(0,1) \quad i \in \{1, \dots, n\}$$

$$H_1: \begin{cases} Z_i \sim N(\mu_1, 1) & i \in \{1, \dots, a\} \\ Z_i \sim N(\mu_1 + (i-a)(\mu_2 - \mu_1) / (b-a), 1) & i \in \{a+1, \dots, b\} \\ Z_i \sim N(\mu_2, 1) & i \in \{b+1, \dots, n\} \end{cases}$$

La statistique du test s'écrit :

$$T_{ab}^{\max} = \max_{\substack{a,b \\ 1 \leq a < b \leq n}} \left\{ -a\mu_1^2 + 2a\mu_1\bar{z}_1 - \mu_1^2SB - \mu_1^2SA + 2\mu_1SZB + 2\mu_2SZA - 2\mu_1\mu_2SAB - (n-b)\mu_2^2 + 2(n-b)\mu_2\bar{z}_2 \right\} \quad (6)$$

Avec

$$SA = \sum_{i=a+1}^b (i-a)^2 / (b-a)^2 \quad (6.1)$$

$$SB = \sum_{i=a+1}^b (b-i)^2 / (b-a)^2 \quad (6.2)$$

$$SZA = \sum_{i=a+1}^b z_i(i-a) / (b-a) \quad (6.3)$$

$$SZB = \sum_{i=a+1}^b z_i(b-i) / (b-a) \quad (6.4)$$

$$SAB = \sum_{i=a+1}^b (b-i)(i-a) / (b-a)^2 \quad (6.5)$$

$$\mu_1 = \frac{a\bar{z}_1 + SZB - SL \times SAB}{a + SB + SK \times SAB} \quad (6.6)$$

$$\mu_2 = \mu_1 SK + SL = \mu_1 \frac{(-SAB)}{SA + n - b} + \frac{(n - b)\bar{z}_2 + SAZ}{SA + n - b} \quad (6.7)$$

Si $b=a+1$, $T_{ab}^{\max} = T_a^{\max}$. Dans ce cas, μ_1 et μ_2 sont égales \bar{z}_1 et \bar{z}_2 dans le test pour un saut.

Les valeurs critiques pour le test de saut et tendance sont données par Alexanderson en 1994 (Annexe A).

Si la valeur de la statistique du test d'une série est inférieure à la valeur critique au seuil de 95%, la série est homogène, dans le cas contraire, hétérogène.

II.1.4.3. Test de Mann Kendall [19][23]

Le test de Mann Kendall est un test non paramétrique qui sert à détecter une tendance.

- L'hypothèse nulle H_0 est l'absence de la tendance dans la série.
- Hypothèse alternative H_1 est la présence d'une tendance dans la série étudiée.

La statistique S de Kendall et sa variance sont données par :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(X_j - X_k) \quad (7)$$

$$\text{Var}(S) = \frac{n(n-1)(2n-5)}{18} \quad (7.1)$$

Où n est le nombre de données de la série, et les X_i ($i=1 \dots n$) sont les observations, supposées indépendantes.

Dès que l'échantillon contient une dizaine de données, la loi de la statistique de test Z ci-dessous peut être approchée par une gaussienne centrée réduite.

$$Z = \begin{cases} \frac{S-1}{(\text{Var}(S))^{\frac{1}{2}}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{(\text{Var}(S))^{\frac{1}{2}}} & \text{si } S < 0 \end{cases} \quad (7.2)$$

La statistique S est liée étroitement à la valeur τ de Kendall par :

$$\tau = \frac{S}{D} \quad (7.3)$$

Où

$$D = \left[\frac{1}{2} n(n-1) - \frac{1}{2} \sum_{j=1}^p t_j(t_j-1) \right]^{\frac{1}{2}} \left[\frac{1}{2} n(n-1) \right]^{\frac{1}{2}} \quad (7.4)$$

II.1.4.4. Test de normalité de Shapiro-Wilk [19][24]

Le test de Shapiro-Wilk teste l'hypothèse nulle selon laquelle un échantillon provient d'une population normalement distribuée.

La statistique calculée est :

$$W = \frac{\left(\sum_{i=1}^n a_i X_i' \right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (8)$$

Où \bar{X} est la moyenne de l'échantillon, X' est le vecteur contenant les données triées (X_i' est donc la i-ème valeur la plus petite), n est le nombre d'observations et :

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} m)^{\frac{1}{2}}}, \text{ où } m = (m_1, \dots, m_n)^T \text{ et les } m_i \text{ sont les valeurs attendues}$$

d'une distribution normale pour un échantillon de taille n et V est la matrice variance-covariance correspondante.

La p-value du test est une valeur exacte pour $n=3$, sinon des approximations différentes sont utilisées pour $4 \leq n \leq 11$ d'une part et $n \geq 12$ d'autre part.

II.2. Méthodes

II.2.1. Préparation des données

Les données téléchargées du site de l'ECMWF sont enregistrées dans une matrice à trois dimensions (11x7x13514). Ce sont des données journalières des 77 points de grilles et durant 13514 jours (du 1er janvier 1979 au 31décembre 2015).

Nous nous intéressons seulement sur les points qui sont sur le continent (Fig.1).

On calcule avec Matlab, les moyennes mensuelles et annuelles en tenant compte des années bissextiles.

II.2.2. Homogénéisation des données

II.2.2.1. Homogénéité absolue

Elle consiste à tester la normalité et l'homogénéité de chaque série par le test de Buishand si la distribution est normale et par le test de Pettit [25] sinon. Dans le cas où la série n'est pas homogène, on teste la stabilité de la moyenne par le test de Mann Kendall. On passe à l'ajustement si le dernier test est significatif, sinon la série est considérer homogène (Fig.2).

Les sauts significatifs sont corrigés en ajustant le début de la série jusqu'au point de discontinuité. Le dernier segment de la série est considéré homogène et la correction débute avec le segment le plus récent en reculant dans le temps. Le facteur de correction correspond à la différence des moyennes entre les 12 observations de chaque côté du saut.

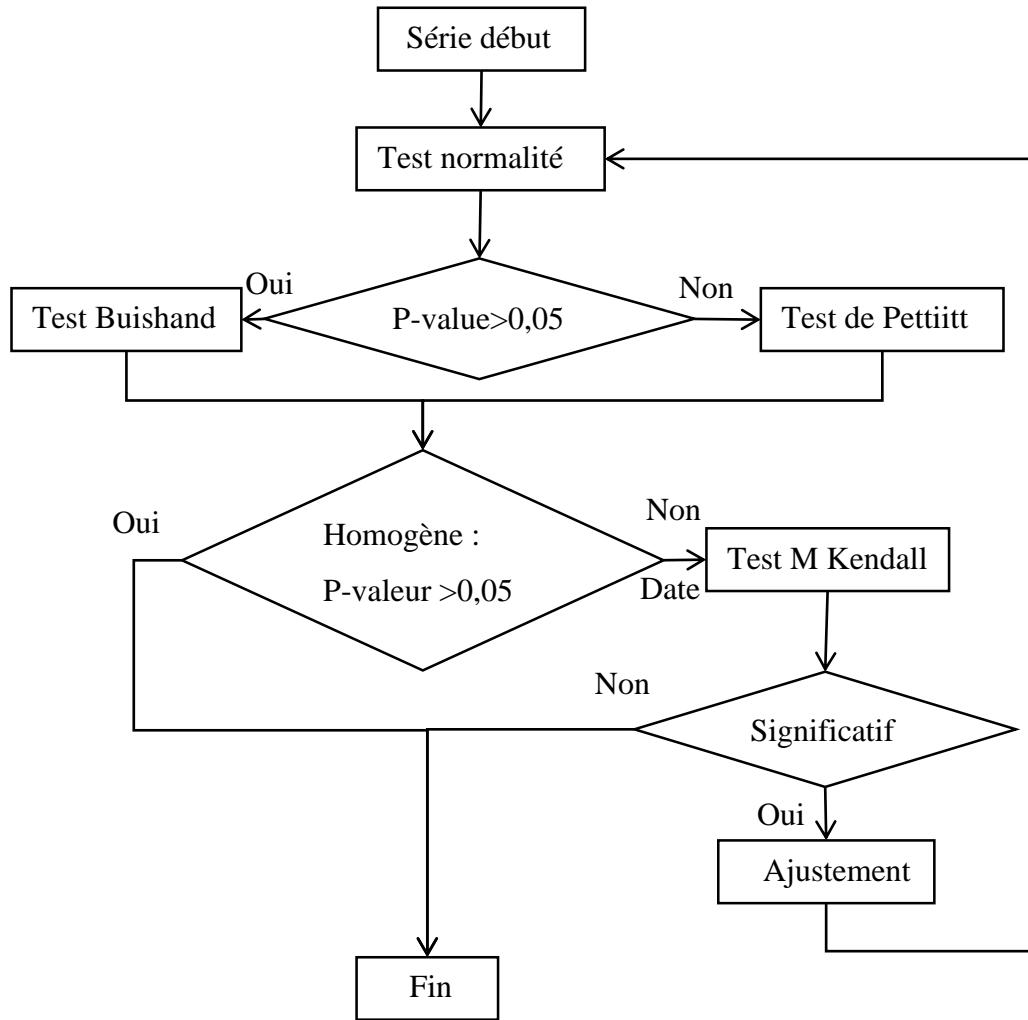


Figure 2 : Organigramme du test d'homogénéité absolue

II.2.2.2. Homogénéité relative [26]

Création série de référence

Chaque point de grille est considéré comme une station. Toutes les stations ont au moins une station voisine à une distance environ de 55 km.

II.2.2.2.1. Première approche : à partir de l'homogénéité absolue et corrélation

On utilise les résultats de l'homogénéité absolue pour créer les séries de référence.

Les moyennes des données homogènes (précédente) de ces stations voisines sont utilisées comme série de référence, si elles sont complètes [4] et ayant des corrélations avec la série de base supérieur à 0,8 [27]. Sinon, on abandonne cette série de base.

Le tableau ci-dessous indique les positions des séries de base et série de référence :

Tableau 3 : Position d'une série de base et ses séries de références

$S_{i,j}$	$S_{i,j+1}$	$S_{i,j+2}$
$S_{i+1,j}$	$S_{i+1,j+1}$	$S_{i+1,j+2}$
$S_{i+2,j}$	$S_{i+2,j+1}$	$S_{i+2,j+2}$

$S_{i+1,j+1}$: Série de base.

$S_{i,j}$: Séries des stations voisines $i \in \{1, \dots, 9\}$ et $j \in \{1, \dots, 5\}$.

Le nombre de stations voisines est entre 3 à 8.

II.2.2.2.2. Deuxième approche : à partir des séries voisines initiales

Les positions des séries de bases et de référence sont la même que les précédentes (Tableau 3). Les séries de références sont toutes les séries brutes voisines de la série de base.

II.2.2.2.3. Test d'homogénéité relative des séries de base

Nous utilisons les tests de SNHT pour les sauts et tendances. On calcule la statistique du test (Eq.5.1, Eq.6.1) et on compare avec la valeur critique donné par Alexanderson (ANNEXE A) selon la longueur de la série étudiée. Si la statistique dépasse la valeur critique correspondante alors la série est hétérogène. Sinon elle est homogène (Fig.3).

La série de base hétérogène est subdivisée en deux parties :

- *pour les sauts* :

La première partie est du début de la série jusqu'à la date de rupture. La deuxième partie est à la date de rupture jusqu'à la fin de la série.

- *pour les tendances* :

La première partie est du début de la série jusqu'au début de la tendance. La deuxième partie à la fin de la tendance jusqu'à la fin de la série.

Si la longueur de la partie est inférieure ou égale à 10, le test se termine pour cette partie. Et si elle est supérieure à dix, nous recommençons le test pour cette partie et ainsi de suite. Voici le schéma qui explique les procédures suivis pour le test relatif.

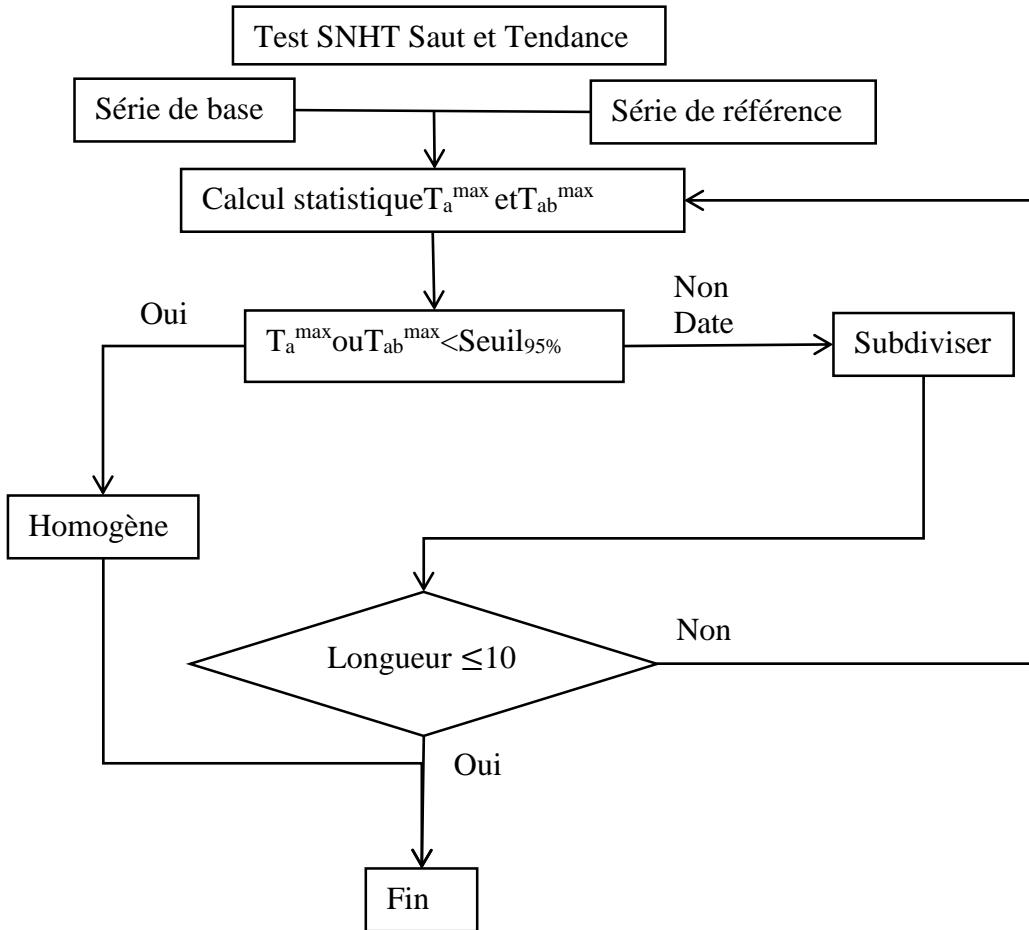


Figure 3 : Organigramme du test d'homogénéité relative

- **Identification de rupture :**

Il est nécessaire d'identifier les types de ruptures avant de passer à la correction pour ne pas fausser les données. En fait, si la longueur de la tendance b-a est supérieur ou égal à 5 et la statistique du test de tendance T_{ab}^{\max} est strictement supérieur à la valeur critique au seuil de 95%, la rupture est une tendance [28]. Dans le cas contraire c'est un saut.

- **Ajustement :**

La série se corrige en ajoutant le premier segment par la différence $\bar{q}_2 - \bar{q}_1$ pour le saut. Pareil pour la tendance mais \bar{q}_1 et \bar{q}_2 sont obtenues en remplaçant respectivement \bar{z}_1 par μ_1 dans l'équation (5.1) et \bar{z}_2 par μ_2 dans l'équation (5.2).

II.2.3. Recherche de meilleure méthode d'homogénéisation

Si les dates de ruptures sont avant 2008, les données sont divisées en 2 parties. La première partie est formée par les températures moyennes annuelles allant de 1979 à 2008 et la deuxième partie celles de 2009 à 2015. Sinon les données ne sont pas divisées.

La deuxième partie est utilisée pour identifier la meilleure méthode d'homogénéisation.

La meilleure méthode est celle qui aura la minimale de moyenne des erreurs absolues après la prévision.

II.2.4. Zonage et modélisation

On utilise l'ACP (Analyse en Composantes Principales) [9][29][30][31][32] pour diviser la zone d'étude en quelques sous zones pour les quatre données: (i) données initiales (avant homogénéisation), (ii) données après homogénéisation absolue, (iii) données après homogénéisation relative de la première approche, (iv) données après homogénéisation de la deuxième approche.

Pour chaque donnée :

- Les variables sont les 30 valeurs moyennes annuelles des températures de l'année 1970 à 2008 si les dates de ruptures sont avant 2008, sinon elles sont les 37 valeurs moyennes annuelles des températures de l'année 1970 à 2015,
- Les individus sont les 37 stations (point de grille).

Ainsi, un tableau de données est une matrice de 30 ou 37 lignes fois 37 colonnes.

On utilise la température moyenne d'une sous zone et de la zone entière comme série chronologique à modéliser. Au total nous avons 8 séries à modéliser.

Pour cette série, on calcule l'exposant de Hurst pour vérifier si le phénomène présente une dépendance à long terme ou non. Si la série présente un effet mémoire longue (exposant de Hurst supérieur à 0,5). On adopte un modèle ARIMA pour prévoir les futures températures annuelles.

PARTIE III: RESULTATS ET INTERPRETATIONS

III.1. Test de normalité

Après le test de normalité de Shapiro-Wilk, toutes les séries ont une p-value supérieure au seuil 0,05 (ANNEXE B). La p-value minimale est 0,209 (série C3) tandis que la maximale est 0,996 (série G5). Donc toutes les données suivent une loi normale.

III.2. Test d'homogénéité absolue

III.2.1. Test de Buishand

Le tableau 4 et la figure 4 représentent les résultats du test d'homogénéité absolue :

10 séries sont homogènes (27 %) et 27 hétérogènes (73 %). Les dates des ruptures sont entre 1995 et 2001. Les séries homogènes ont une p-value qui varie de 0,057 (série G1) à 0,130 (série G3), avec un seuil de 0,05.

Tableau 4 : Résultats des tests de Buishand

Série	P-value	Série	P-value	Série	P-value
C1	0,001	E3	0,039	F7	0,022
C2	0,001	E4	0,04	F8	0,014
C3	0,003	E5	0,075	F9	0,021
C4	0,015	E6	0,023	G1	0,057
D1	0,01	E7	0,013	G2	0,094
D2	0,007	E8	0,009	G3	0,13
D3	0,008	F1	0,023	G4	0,097
D4	0,012	F2	0,099	G5	0,085
D5	0,03	F3	0,095	G6	0,058
D6	0,023	F4	0,072	G7	0,035
D7	0,014	F5	0,063	G8	0,025
E1	0,013	F6	0,034	G9	0,035
E2	0,029				

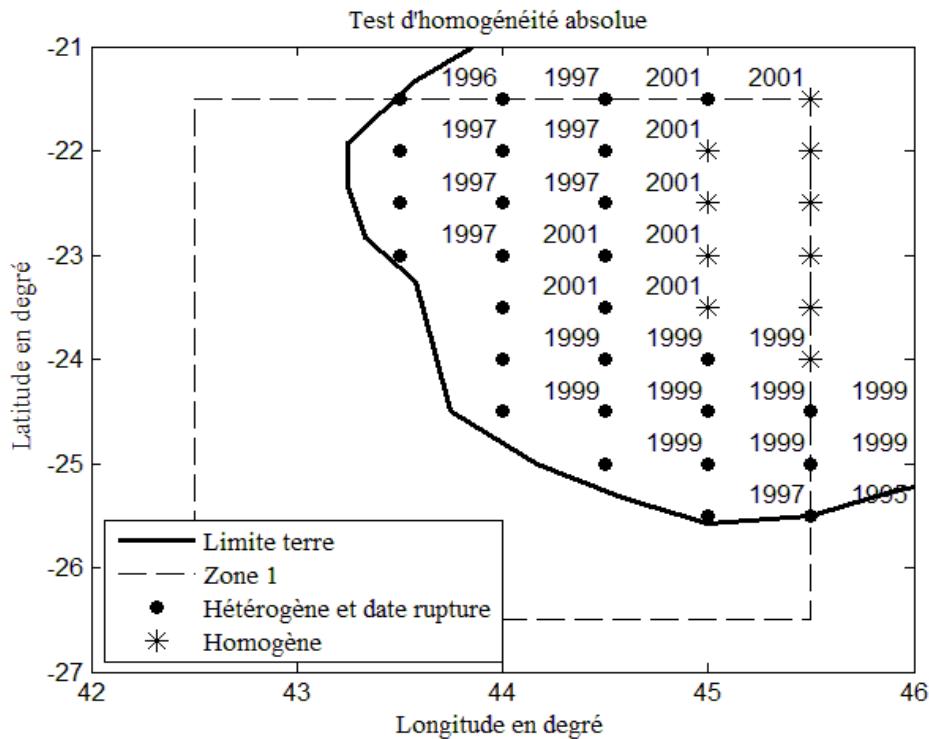


Figure 4 : Résultats du test d'homogénéité absolue.

III.2.2. Test de Mann Kendall

Le test de Mann Kendall appliquée aux série hétérogènes indique la présence de tendance dans ces séries car les p-values sont inférieurs à 5% (ANNEXE C).

Qui implique que les moyennes ces séries ne sont pas stables. Donc les corrections des ruptures détectées par le test de Buishand sont raisonnables.

III.3. Ajustement

Facteur de correction

Les facteurs de correction des séries hétérogènes sont donnés par le tableau 5. Ils varient de 0,314 (série G9) à 0,798 (série E5).

Tableau 5 : Facteur de correction des séries hétérogènes

Série	Facteur de correction	Série	Facteur de correction	Série	Facteur de correction
C1	0,395	D6	0,641	E8	0,692
C2	0,429	D7	0,674	F1	0,536
C3	0,447	E1	0,478	F6	0,67
C4	0,506	E2	0,463	F7	0,733
D1	0,475	E3	0,471	F8	0,671
D2	0,503	E4	0,568	F9	0,39
D3	0,472	E5	0,798	G7	0,644
D4	0,571	E6	0,739	G8	0,632
D5	0,73	E7	0,768	G9	0,314

La figure 5 représente 3 exemples de séries hétérogènes et ces ajustements, à savoir la série C2, la série C4 et la série F1. Ces 3 séries contiennent des sauts aux dates respectives : 1997, 2001 et 2001. Les courbes avec un petit cercle représentent les séries, les lignes horizontales moins et plus grasses représentent respectivement les moyennes des séries avant et après ruptures. Les courbes avec un petit carré représentent les séries finales après correction.

Puisque la correction des ruptures se fait en ajoutant à la série le facteur de correction, du début jusqu'à la date de rupture, on remarque une augmentation de la série corrigée jusqu'à la date de rupture.

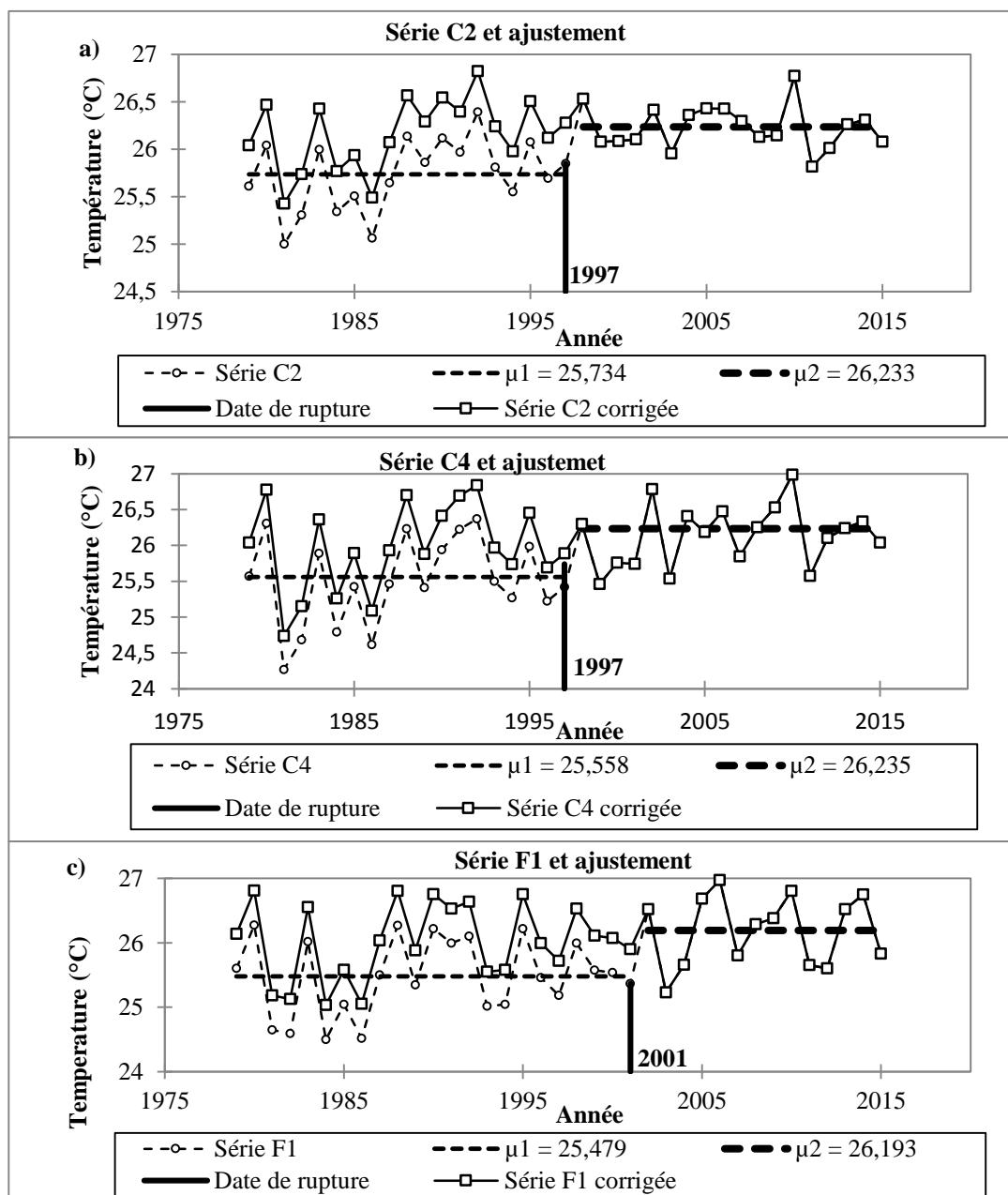


Figure 5 : Exemple de série hétérogène : a) série C2 ; b) série C4 ; c) série F1

Après le premier ajustement (Fig.5.a), la série C2 est encore hétérogène et la date de rupture est 1987, un deuxième ajustement est alors nécessaire (Fig.6).

La figure 8 représente l'ajustement de la deuxième rupture de la série C2.

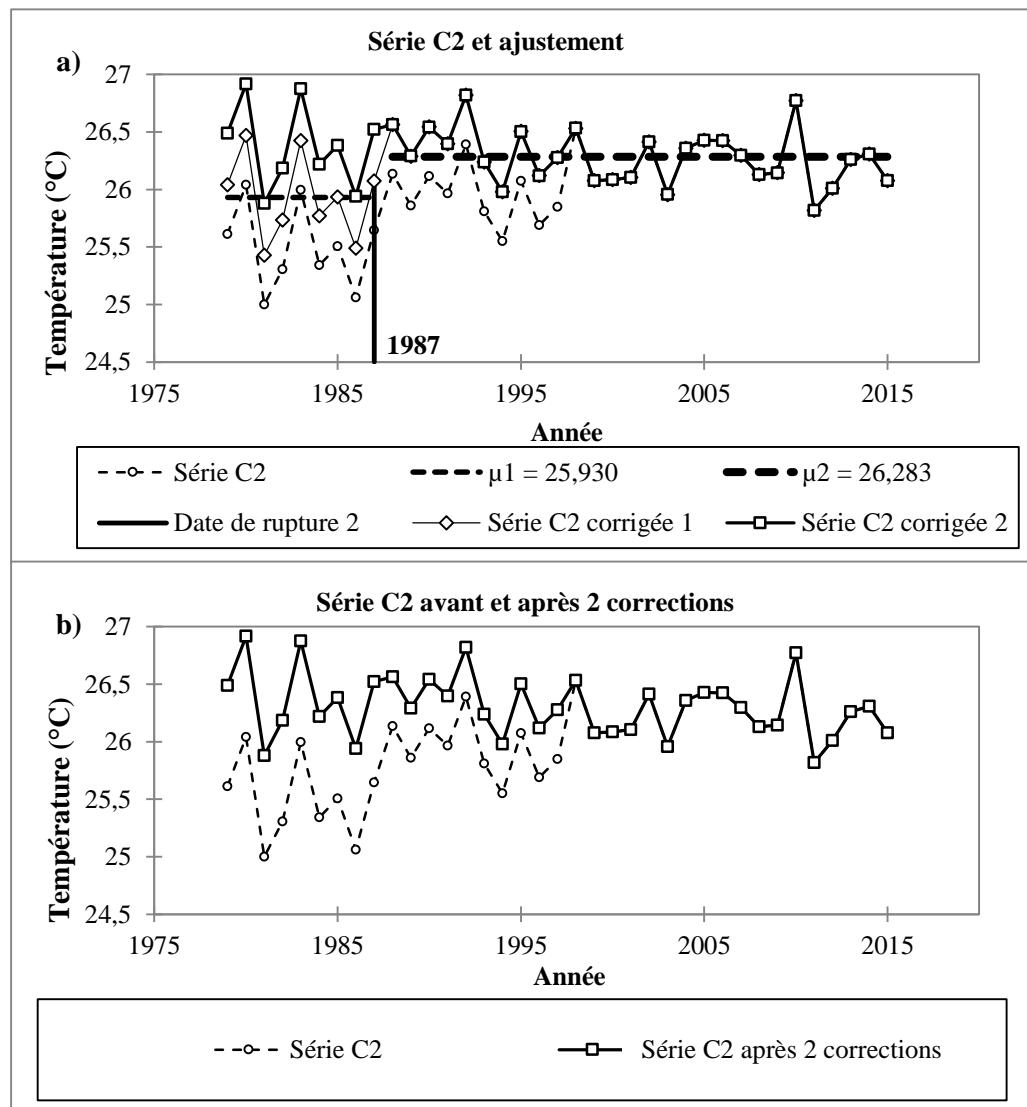


Figure 6 : Deuxième ajustement de la série C2 : a) série C2 et ajustement, b) série C2 avant et après ajustement

La figure 7 représente 3 exemples de séries homogènes, à savoir C2, C4 et F1. Les courbes avec un petit cercle représentent les séries, les lignes horizontales représentent respectivement les moyennes des séries.

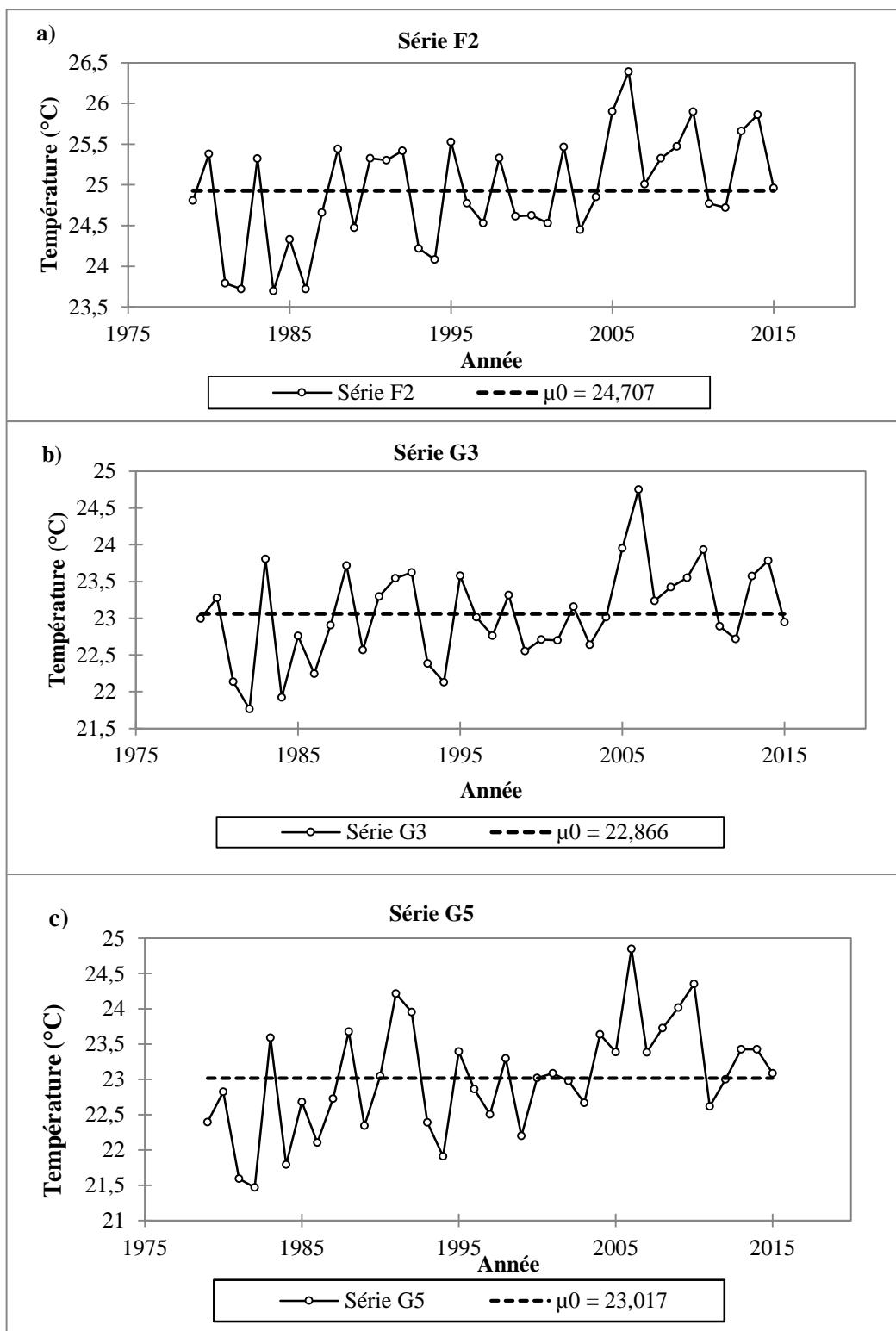


Figure 7 : Exemple de séries homogènes : a) série F2 ; b) série G3 ; c) série G5

III.4. Résultat final d'homogénéité absolue après ajustement

Après ajustement, les séries sont toutes homogènes.

Le tableau 6 montre que toutes les séries sont homogènes après ajustement car les p-values sont supérieur au seuil 0,05.

Tableau 6 : Résultat du test d'homogénéité absolue après ajustement

Série	P-value	Série	P-value	Série	P-value
C1	0,150	D6	0,345	E8	0,258
C2	0,311	D7	0,269	F1	0,361
C3	0,124	E1	0,079	F6	0,147
C4	0,238	E2	0,096	F7	0,148
D1	0,137	E3	0,130	F8	0,180
D2	0,099	E4	0,190	F9	0,232
D3	0,118	E5	0,186	G7	0,145
D4	0,126	E6	0,245	G8	0,157
D5	0,331	E7	0,212	G9	0,268

Le coefficient de corrélation de la série C1 avec sa série de référence est inférieur à 0,8 (Tableau 7). On ne sait pas la propriété de la série C1 vis-à-vis de l'homogénéité relative. Pour la suite nous enlevons la série C1 dans la zone d'étude.

Tableau 7 : Corrélation de la série de base avec la série de référence

Série	Corrélation	Série	Corrélation	Série	Corrélation
C1	0,564	E3	0,975	F7	0,936
C2	0,835	E4	0,97	F8	0,896
C3	0,865	E5	0,953	F9	0,855
C4	0,87	E6	0,933	G1	0,981
D1	0,9	E7	0,906	G2	0,984
D2	0,907	E8	0,88	G3	0,988
D3	0,913	F1	0,982	G4	0,99
D4	0,919	F2	0,987	G5	0,993
D5	0,929	F3	0,993	G6	0,974
D6	0,909	F4	0,99	G7	0,953
D7	0,904	F5	0,98	G8	0,904
E1	0,942	F6	0,97	G9	0,827
E2	0,961				

III.5. Homogénéité relative

III.5.1. Première approche

Après le test de SNHT pour un saut, la subdivision si la série est hétérogène et l'identification de rupture (saut de moyenne ou tendance) Les résultats sont représentées par la figure 8 suivantes :

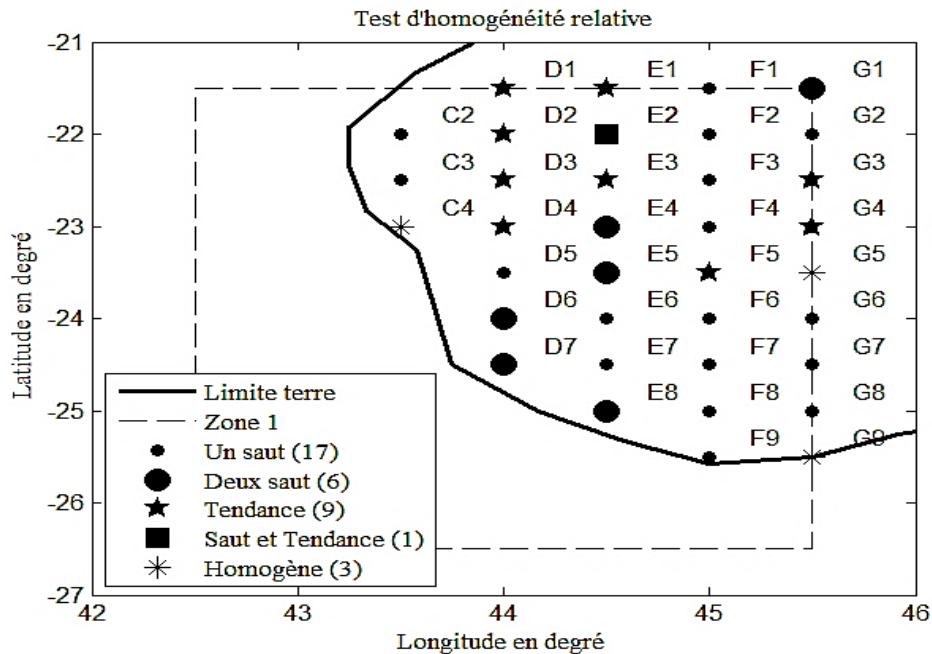


Figure 8 : Résultats du test d'homogénéité relative première approche

III.5.1.1. Série avec un saut

17 séries ont un saut de moyenne d'amplitude entre -0,12 (série G2) et 0,76 (série E7).

Les séries avec les dates de ruptures ainsi que les amplitudes sont données par le tableau 8 suivant.

Tableau 8 : Série avec un saut

Série	Date	Amplitude	Série	Date	Amplitude	Série	Date	Amplitude
C2	1997	0,3659174	F2	2002	0,24080706	F9	1999	0,26916146
C3	1997	0,2829992	F3	2002	0,17725359	G2	1988	-0,12726409
D5	2002	0,62059338	F4	2002	0,21158023	G6	2000	0,38305041
E6	2000	0,73397421	F6	2000	0,52042609	G7	2000	0,53977283
E7	2000	0,76791462	F7	2000	0,6857248	G8	2000	0,68736656
F1	2002	0,24521664	F8	2000	0,68676209			

Série avec un saut :

Pour l'homogénéisation absolue, la série C2 présente deux sauts (Fig.6). Pour l'homogénéisation relative de première approche, elle présente seulement un saut (Fig.9) d'amplitude égal à 0,365.

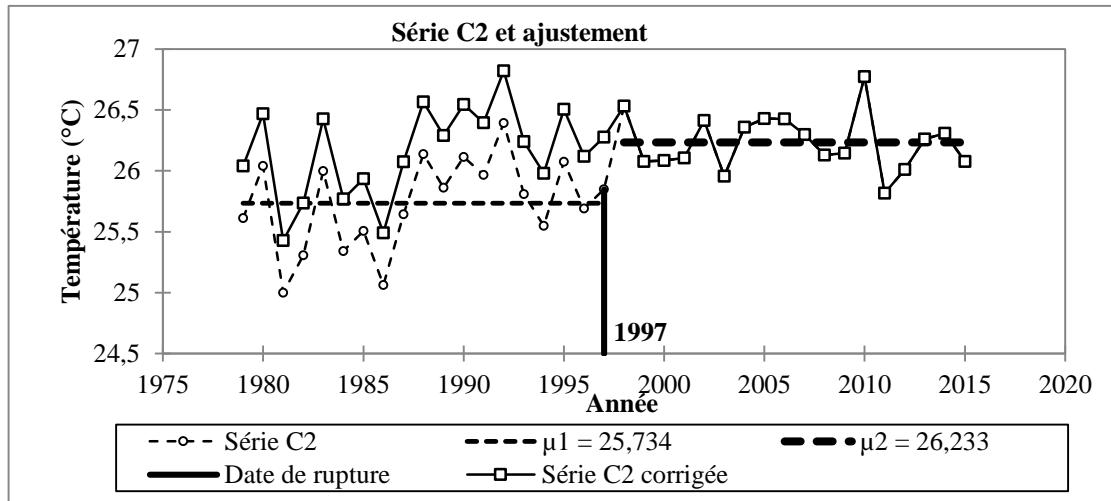


Figure 9 : Exemple de série avec un saut : série C2

III.5.1.2. Deux sauts

Pour l'homogénéisation relative de première approche, 6 séries ont 2 sauts de moyennes d'amplitude entre 0,11 et 0,66. Les séries avec les dates de ruptures et les amplitudes sont représentées par le tableau 9 ci-dessous.

Tableau 9 : Série avec 2 sauts

Série	Première date	Amplitude	Seconde date	Amplitude
D6	1988	0,54608268	2000	0,64705843
D7	1988	0,43880042	2000	0,63319228
E4	1995	0,31757645	2002	0,39843054
E5	1988	0,65307081	2002	0,6146107
E8	2000	0,66174677	2005	0,33857082
G1	1992	0,11597953	2002	0,16058539

Série avec deux sauts :

La série G1 est un exemple de série avec deux sauts (en 1992 et 2002) pour l'homogénéisation relative de première approche. Elle est homogène pour l'homogénéisation absolue (Tableau 4). La série G1 avec les ajustements des deux sauts sont représentés par la figure 10.

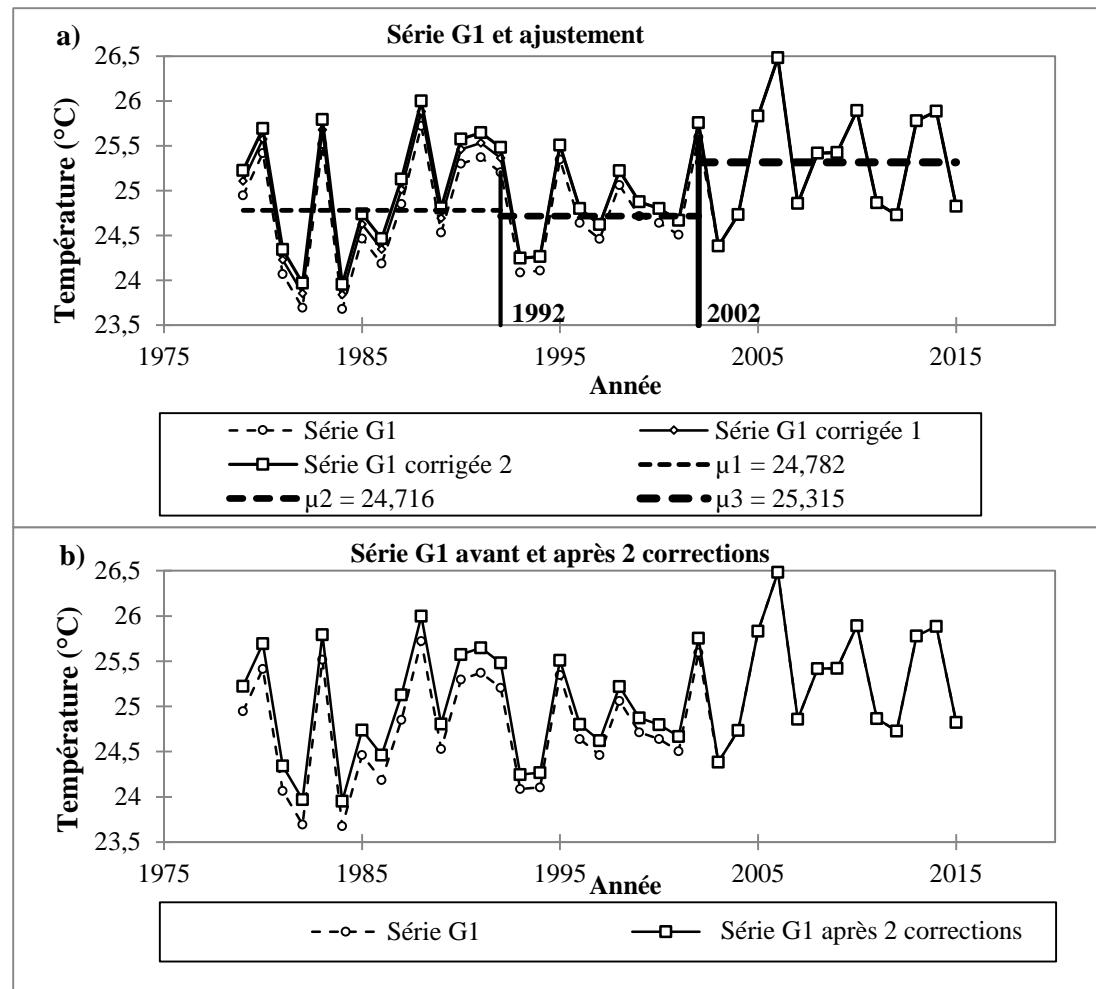


Figure 10 : Exemple de série avec 2 sauts : a) série G1 et ajustement, b) série G1 avant et après ajustement

III.5.1.3. Tendance

9 séries présentent une tendance d'amplitude entre -0,17 et 0,56. Les séries avec les dates du début et fin de tendance ainsi que les amplitudes de ces tendances sont données par le tableau 10 ci-dessous. Les amplitudes en valeurs absolues varient de 0,152 (série G4) à 0,574 (série D4).

Tableau 10 : Série avec une tendance

Série	Date de début	Date de fin	Amplitude
D1	1996	2002	0,51312553
D2	1996	2002	0,54762608
D3	1996	2003	0,56227263
D4	1997	2003	0,57457977
E1	1997	2002	0,45294946
E3	1993	2006	0,38595623
F5	1999	2004	0,40160105
G3	1986	1991	-0,16972117
G4	1986	1990	-0,15256397

Série avec une tendance :

La série E3 est un exemple de série avec une tendance pour l'homogénéisation relative de première approche. La série E3 avec l'ajustement, les moyennes avant et après la tendance, l'ajustement de la série, sont représentés par la figure 11.

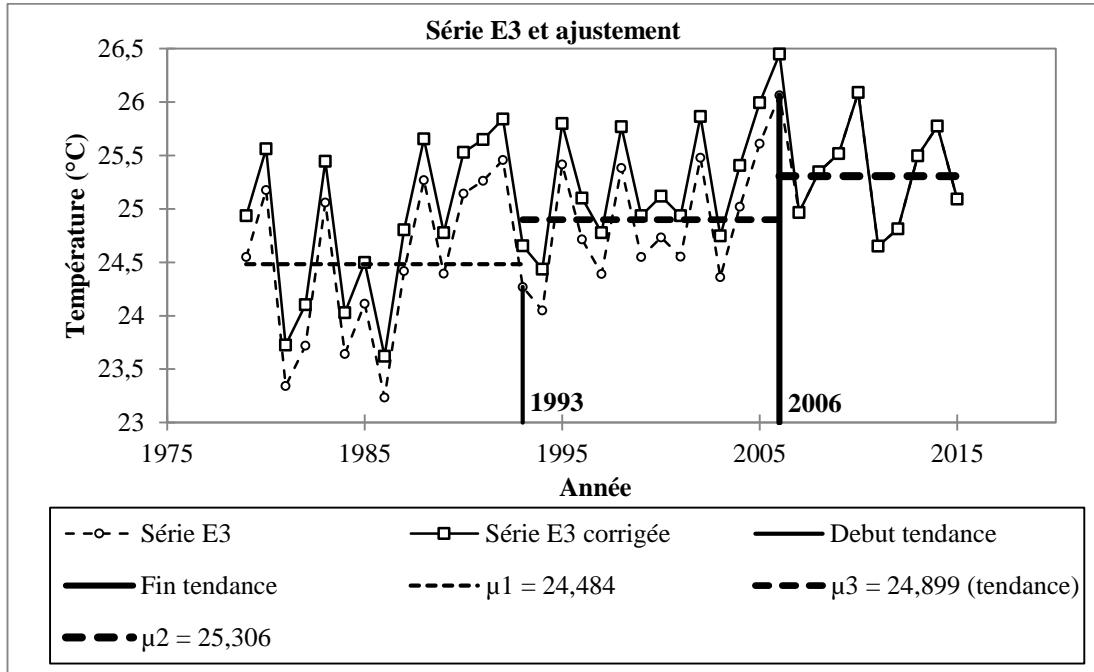


Figure 11 : Exemple de série avec une tendance: série E3

III.5.1.4. Saut et tendance

Une série ont un saut et une tendance à la fois, d'amplitude représenté par le tableau 11ci-dessous.

Tableau 11 : Série avec un saut et une tendance

Série	Date saut	Amplitude	Début tendance	Fin tendance	Amplitude
E2	1990	0,35378712	1993	2005	0,45041535

Série avec un saut et une tendance :

La figure 12 représente la série E2, les ajustements de saut et de la tendance. La courbe en trait avec un petit cercle représente l'ajustement de la tendance. La courbe moins grasse avec un petit losange représente l'ajustement de la tendance. La courbe plus grasse avec un petit carré donne la série E2 après ajustement de saut.

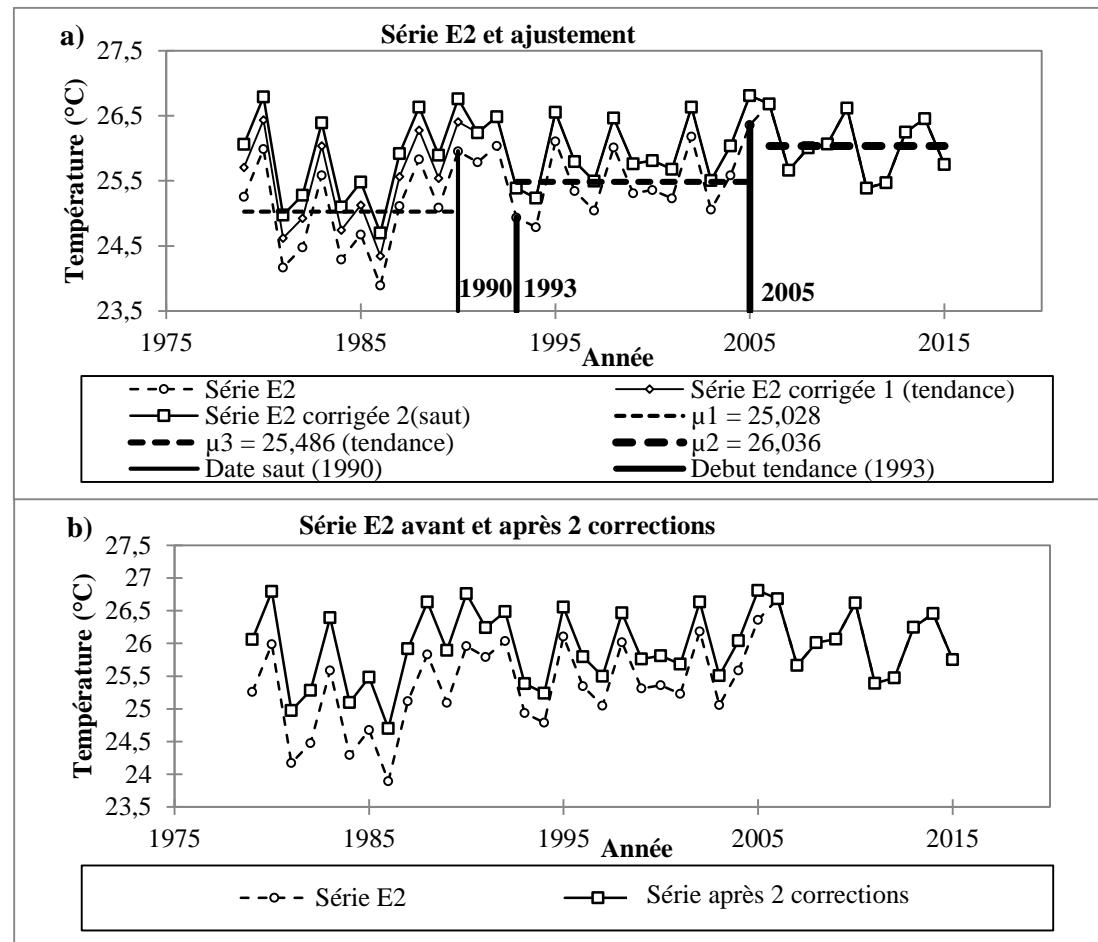


Figure 12 : Exemple de série avec un saut et une tendance:a) série E2, b)série E2 après correction

Série homogène :

3 séries sont homogènes, à savoir C4,G5 et G9.

La figure 13 représente un exemple de série homogène pour l'homogénéisation relative de première approche.

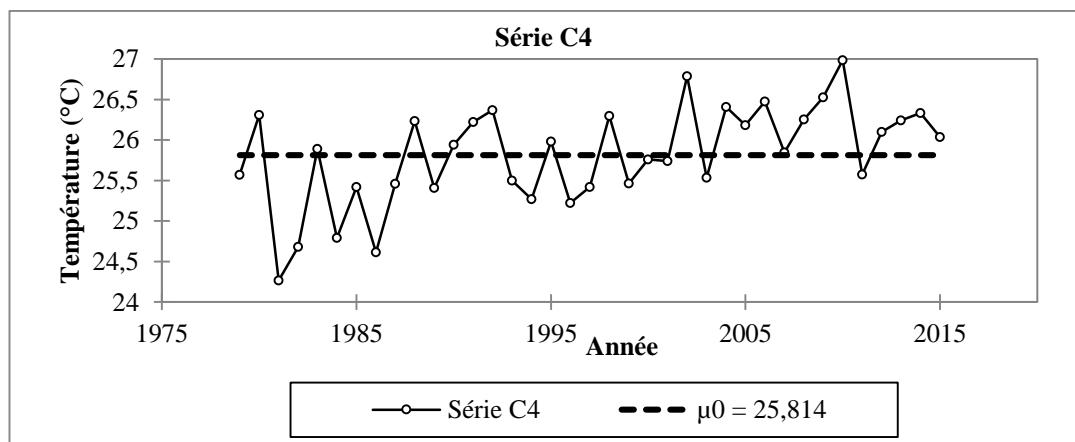


Figure 13 : Exemple de série homogène : série C4

III.5.2. Deuxième approche

La figure 14 représente les résultats de l'homogénéisation relative de deuxième approche. 14 séries présentent un saut, 2 séries présentent deux sauts, 3 séries ont une tendance et 17 séries sont homogènes.

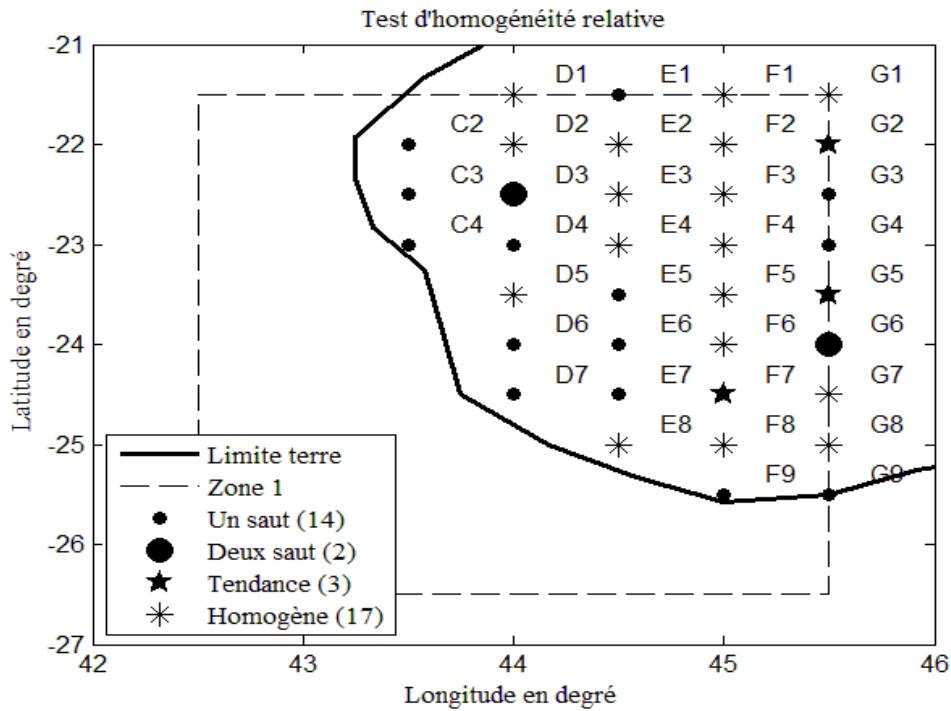


Figure 14 : Résultats du test d'homogénéité relative deuxième approche

III.6. Résultat de l'analyse en composante principale

Le tableau 12 représente les valeurs propres avec les axes principales F1, F2 et F3 pour toutes les données avant et après homogénéisation.

Tableau 12 : Valeur propre

Axes	Données initiales			Données après homogénéisation absolue		
	Valeur propre	Variabilité (%)	% cumulé	Valeur propre	Variabilité (%)	% cumulé
F1	27,65	92,17	92,17	28,21	94,04	94,04
F2	1,31	4,36	96,53	0,84	2,80	96,85
F3	0,74	2,46	98,98	0,60	2,00	98,84
Axes	Données après homogénéisation relative 1			Données après homogénéisation relative 2		
	Valeur propre	Variabilité (%)	% cumulé	Valeur propre	Variabilité (%)	% cumulé
F1	27,94	93,12	93,12	27,63	92,11	92,11
F2	0,99	3,31	96,43	1,30	4,32	96,43
F3	0,58	1,95	98,38	0,72	2,39	98,82

En utilisant le critère de Kaiser, pour les données initiales, nous avons retenus les 2 premiers facteurs F1 et F2.

La Figure 15 met en évidence que toutes les variables sont corrélées positivement avec l'axe F1. En plus, les variables sont toutes bien représentées sur cet axe. Donc, on peut déduire que l'axe F1 représente la température. Par suite, les individus qui se trouvent dans la partie positive de F1 ont de température supérieure à la moyenne ($25,188^{\circ}\text{C}$). L'ensemble de ces individus forme la zone A. Les individus qui sont dans la partie négative de F1 ont de température inférieure à la moyenne et forment la zone B (Fig.16.a). Les données homogénéisées ont les mêmes caractéristiques que les données initiales pour le zonage (Fig.16.a, Fig.16.b et Fig.16.b).

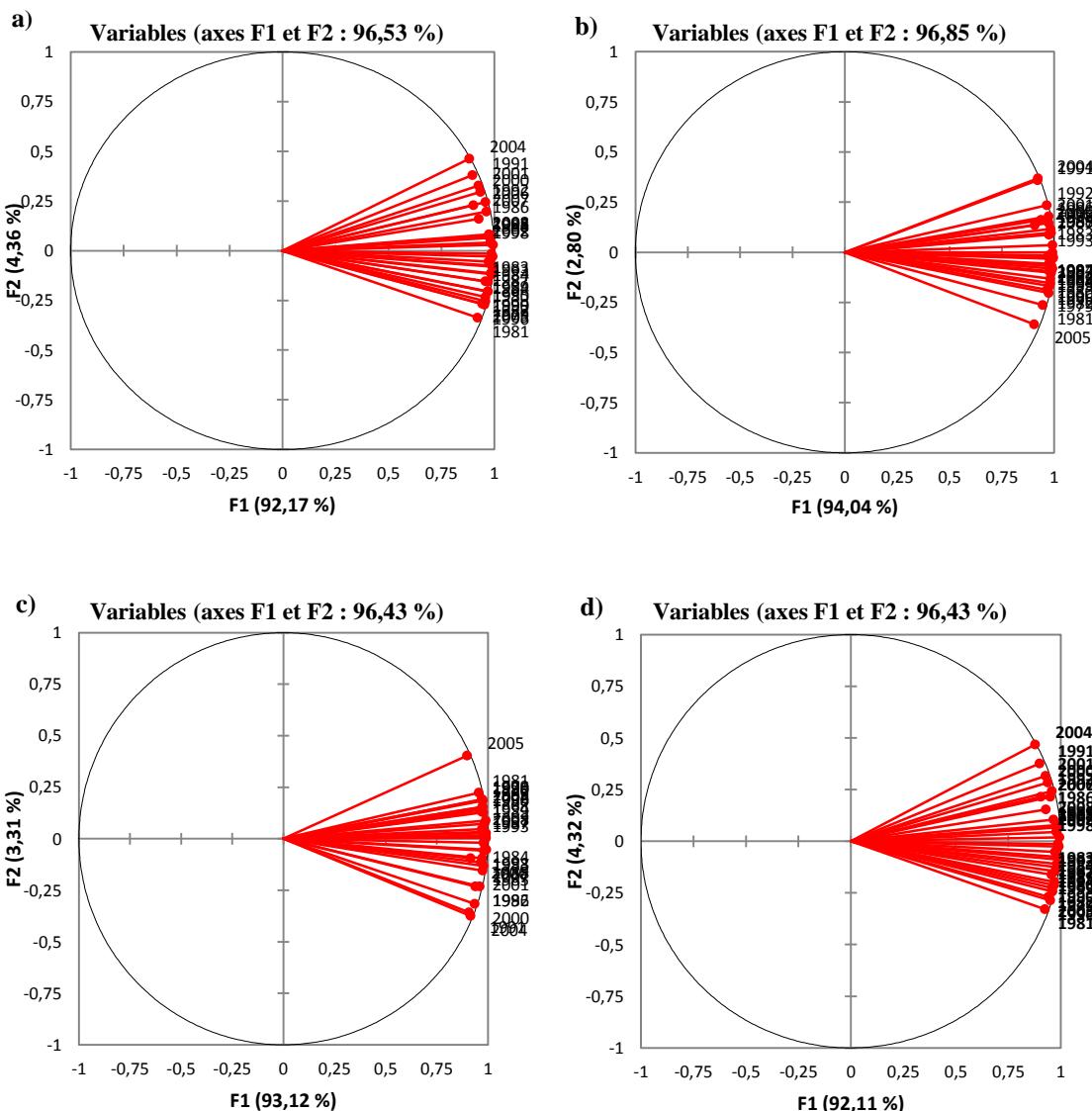


Figure 15 : Représentation des variables dans le plan ($F1, F2$): a) Données initiale, b) Données après homogénéisation absolue, c) Données homogénéisation relative première approche et d) Données homogénéisation relative deuxième approche

La figure 16 représente les coordonnées des individus par rapports aux 2 premières axes principales pour toutes les données avant et après homogénéisation.

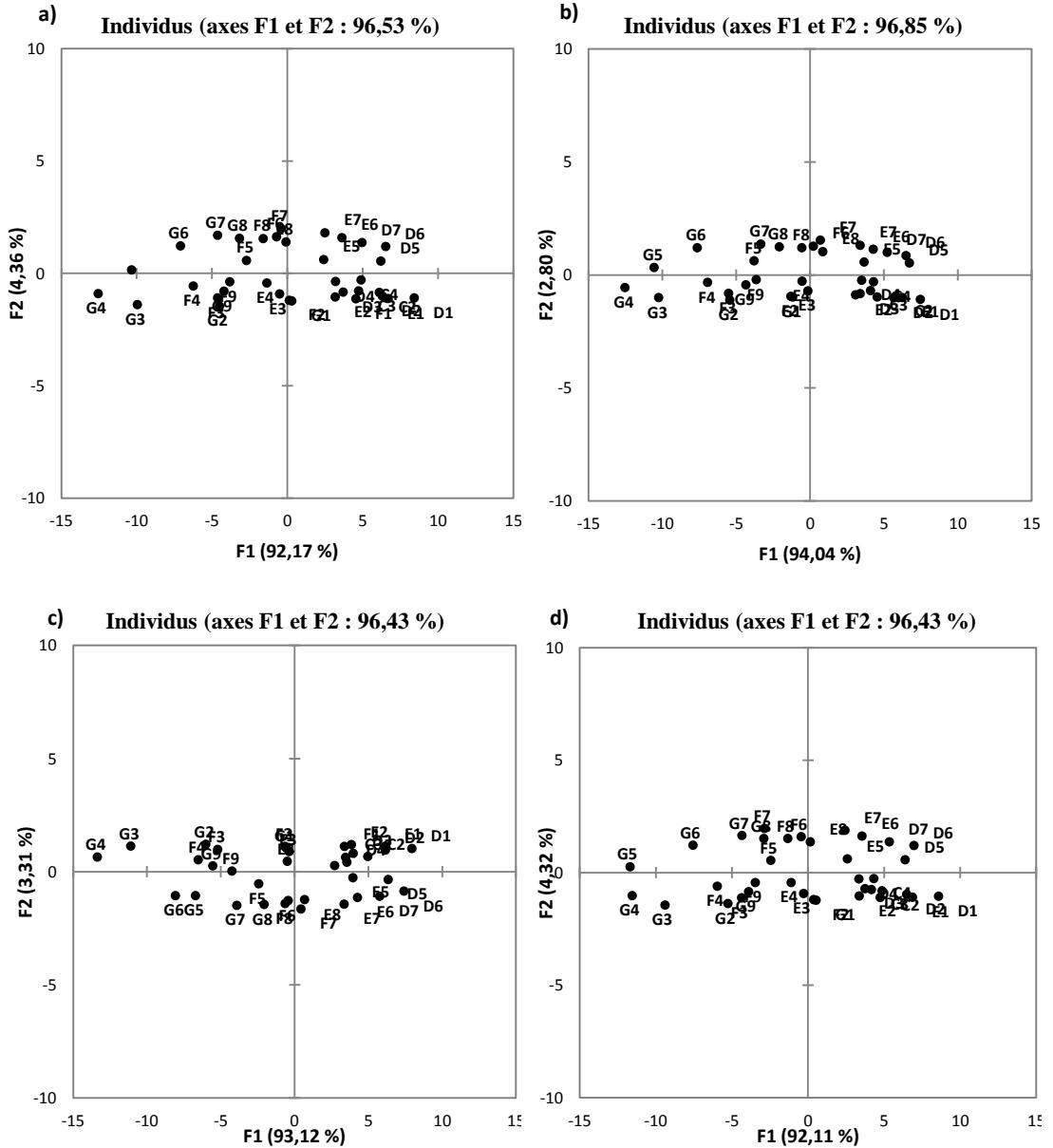


Figure 16 : Coordonnées des individus : a) Données initiale, b) Données après homogénéisation absolue, c) Données homogénéisation relative première approche et d) Données homogénéisation relative deuxième approche

La figure 17 représente les sous zones en divisant la zone d'étude pour les 4 données. Pour les 4 jeux de données, on trouve les mêmes sous zones : B à l'Est et A à l'Ouest. Pour les données avant homogénéisation, les cercles verts représentent les individus de la sous zone A, quant aux cercles rouges, ils représentent les individus de la sous zone B, et les cercles bleus représentent les individus non classées (Fig.17.a). Il en est de même pour les données après homogénéisation (Fig.17.b, Fig.17.c et Fig.17.d).

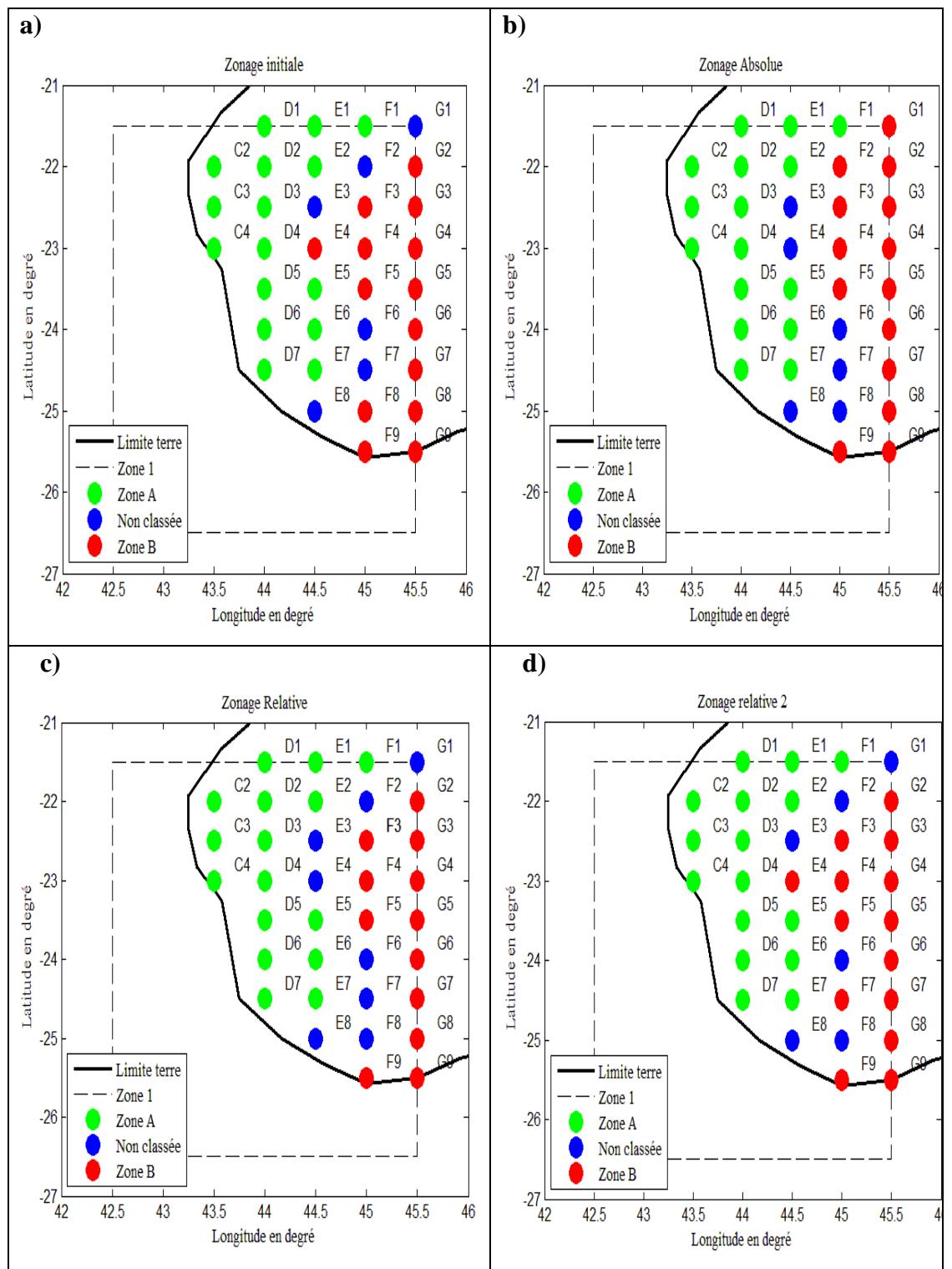


Figure 17 : Division de la zone d'étude en sous zone : a) Avec les données initiale, b) Avec les données après homogénéisation absolue, c) Avec les données après homogénéisation relative première approche et d) Avec les données après homogénéisation relative deuxième approche

III.7. Modélisation

III.7.1. Série moyenne de la sous zone A

La série relative 2 se rapproche de la série initiale et elles sont presque confondues de l'ordre de 10^{-2} (Fig.18). La cause est que pendant l'homogénéisation relative 1, la série initiale a été prise comme série de référence. Les amplitudes des ruptures sont faibles et en calculant la moyenne des données homogènes et homogénéisées, ces amplitudes tendent vers zéro. De même pour la série absolue et la série relative 1.

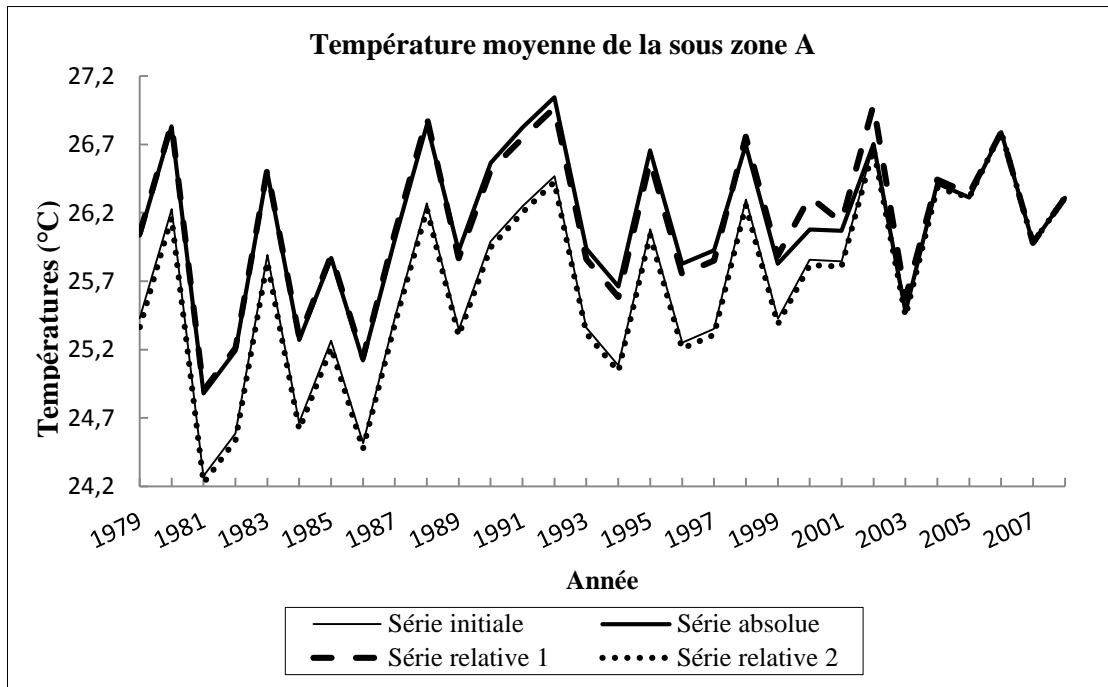


Figure 18 : Série moyenne de la sous zone A

Exposant de Hurst :

Pour toutes les séries, l'exposant de Hurst minimum est 0,5587 et le maximum est de 0,65. Ils sont supérieurs à 0,5. Donc les phénomènes présentent de dépendance à long terme. Les séries absolue et relative 1 ont le même effet mémoire de 3 ans et celles de l'initiale et de la relative 2 ont le même 7 ans (Tableau 13).

Tableau 13 : Exposant de Hurst et effet mémoire des séries de la sous zone A

Série moyenne de la zone A	Initiale	Absolue	Relative 1	Relative 2
Exposant de Hurst	0,644	0,566	0,5587	0,65
Effet mémoire	3	7	3	7

La figure 19 représente la courbe d'autocorrélation et intervalle de confiance des séries de la sous zone A. Les courbes continues moins et plus grasse représentent respectivement la série initiale et la série absolue. La courbe discontinue en trait plus grasse représente la série relative 1. La courbe discontinue en pointillé représente la série relative 2. La courbe discontinue en trait moins grasse représente l'intervalle de confiance.

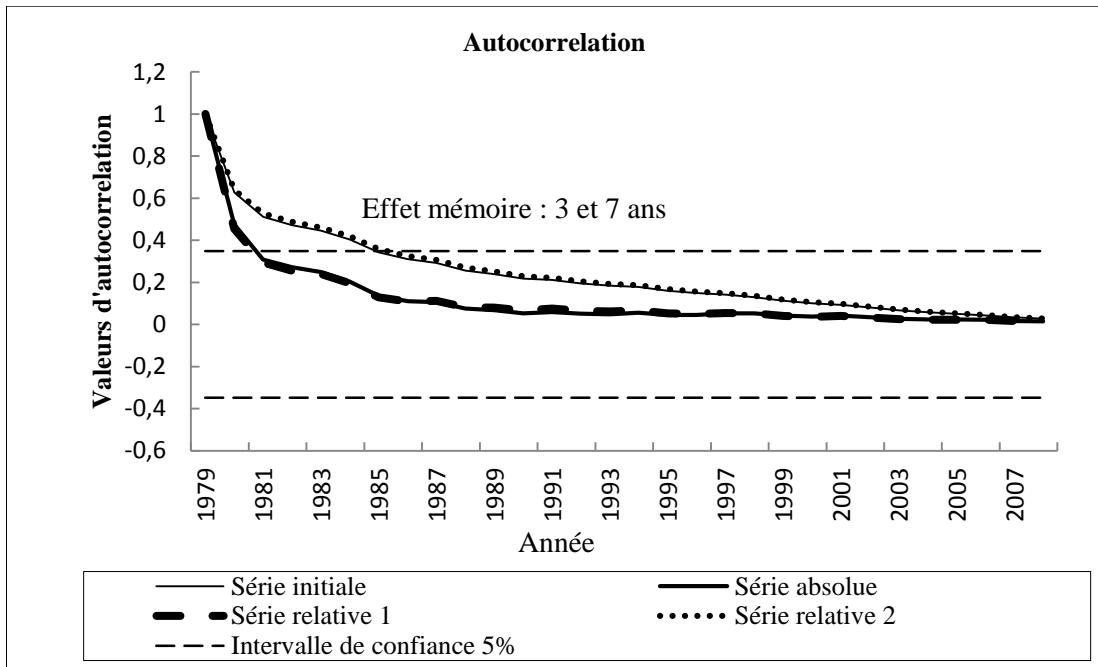


Figure 19 : Courbe d'autocorrélation et intervalle de confiance des séries de la sous zone A

III.8. Modélisation ARIMA

III.8.1. Etude de stationnarité de la série

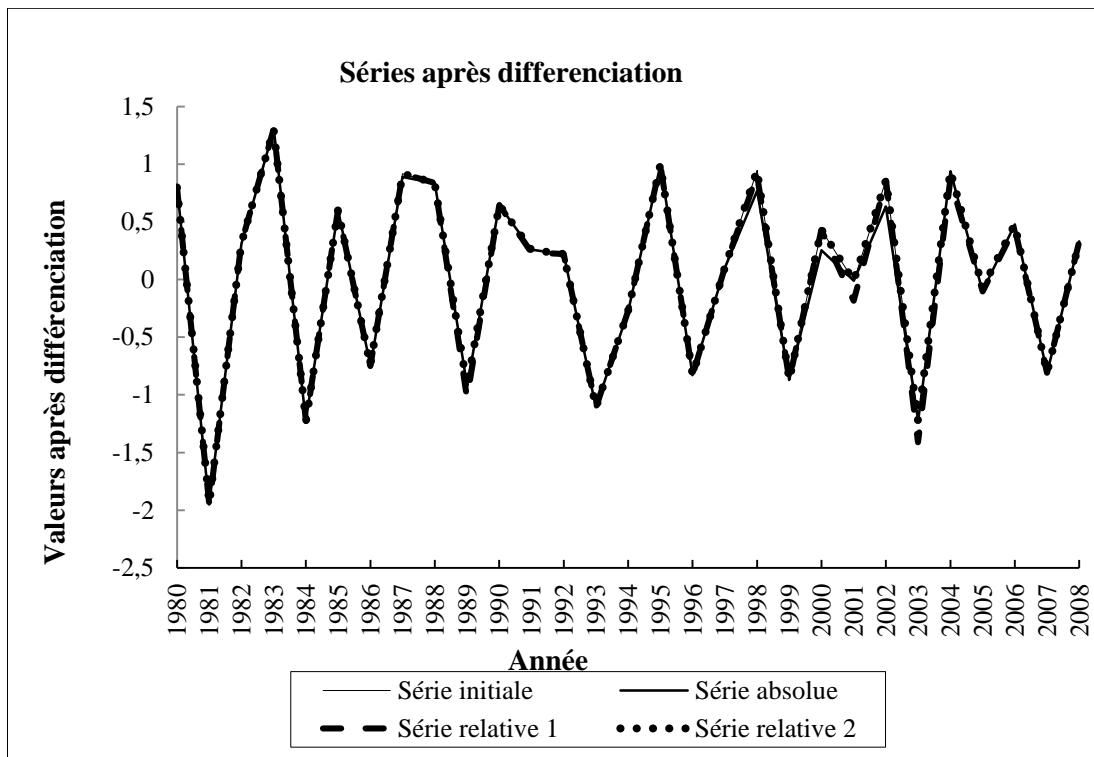
Test de Dickey-Fuller

Le tableau 14 ci-dessous montre que tous les séries ne sont pas stationnaire car leurs p-valeurs dépassent largement la valeur d'alpha 0,05.

Tableau 14 : Résultat du test de stationnarité

	Série initiale	Série absolue	Série relative 1	Série relative 2
Tau (Valeur observée)	-2,066	-2,169	-2,223	-2,059
Tau (Valeur critique)	-0,598	-0,598	-0,598	-0,598
p-value	0,527	0,472	0,445	0,531
alpha	0,05	0,05	0,05	0,05

La figure 20 représente les séries de la sous zone A après différenciation. Puisque ces séries ne sont pas stationnaire (Tableau 14), donc une différenciation est nécessaire.



Après différenciation, toutes les séries sont stationnaires car les p-values sont inférieur à la valeur d'alpha (Tableau 15).

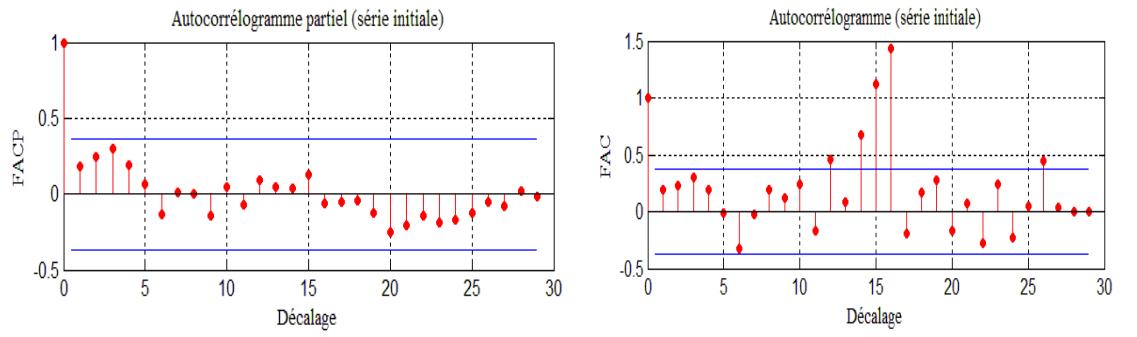
Tableau 15 : Résultat du test de stationnarité après différenciation

	Série initiale	Série absolue	Série relative 1	Série relative 2
Tau (Valeur observée)	-4,168	-4,405	-4,610	-4,145
Tau (Valeur critique)	-0,591	-0,591	-0,591	-0,591
p-value	0,011	0,006	0,004	0,012
alpha	0,05	0,05	0,05	0,05

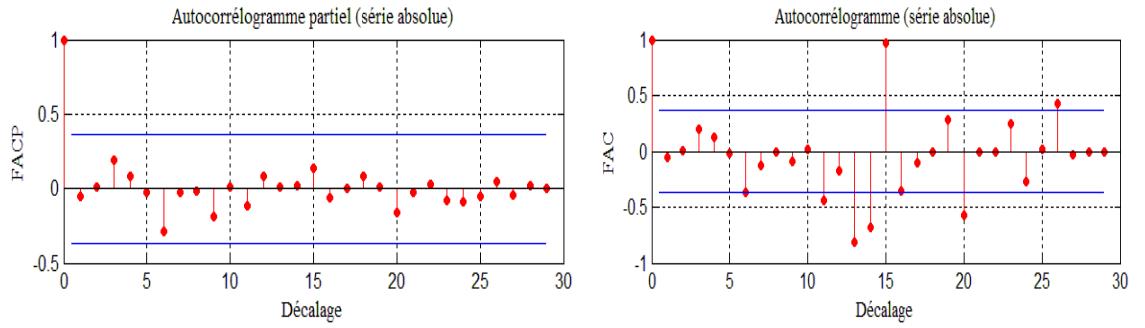
III.8.2. Identification de modèle

La figure 21 montre que pour toutes les 4 séries : leurs correlogrammes de la fonction d'autocorrélation n'ont pas de pics significatifs. Et leurs correlogrammes de la fonction d'autocorrélation partiels indiquent qu'au décalage 26 se trouve le dernier pic significatif. Nous avons donc : $p=0$ pour le modèle AR(p) et l'ordre maximal q du modèle MA (q) est égale à 26.

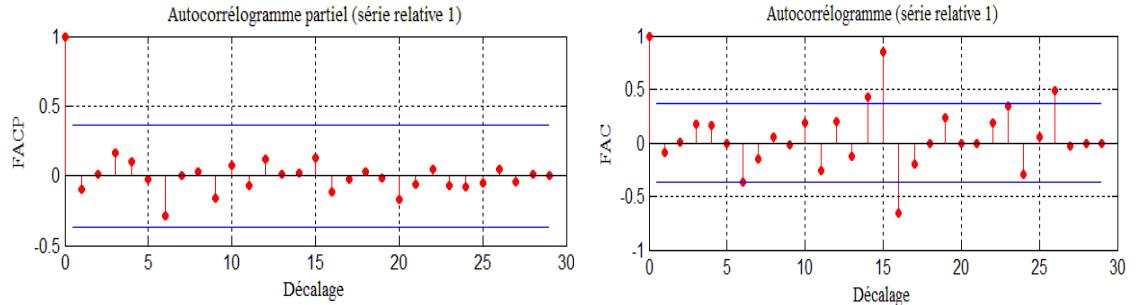
a)



b)



c)



d)

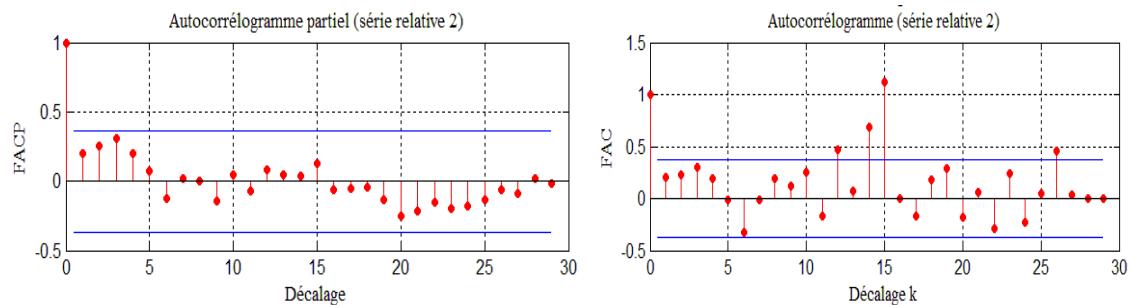


Figure 21 : Corrélogramme de la fonction d'autocorrélation simple et partielle des séries moyenne : a) série initiale , b) série absolue , c) série relative 1 et d) série relative 2.

En fixant p égale à 0 du modèle MA(p) et en variant q de 1 à 26, du modèle MA(q). Nous avons 4 modèles possibles pour la série initiale, un seul pour la série absolue, 3 pour la série relative 1 et 2 pour la série relative 2. Ces modèles vérifient l'hypothèses des bruits blancs, ils sont représentés par le tableau 16 ci-dessous.

Modèle candidat :

Tableau 16 : Modèle possible pour la modélisation

Séries	Modèle	MAPE
Initiale	ARIMA(0,1,15)	1,3426
	ARIMA(0,1,16)	1,3195
	ARIMA(0,1,17)	1,3196
	ARIMA(0,1,18)	1,3228
Absolue	ARIMA(0,1,13)	1,2807
Relative 1	ARIMA(0,1,13)	1,2484
	ARIMA(0,1,18)	1,1589
	ARIMA(0,1,20)	1,1595
Relative 2	ARIMA(0,1,18)	1,3287
	ARIMA(0,1,19)	1,3289

III.8.2. Estimation du modèle

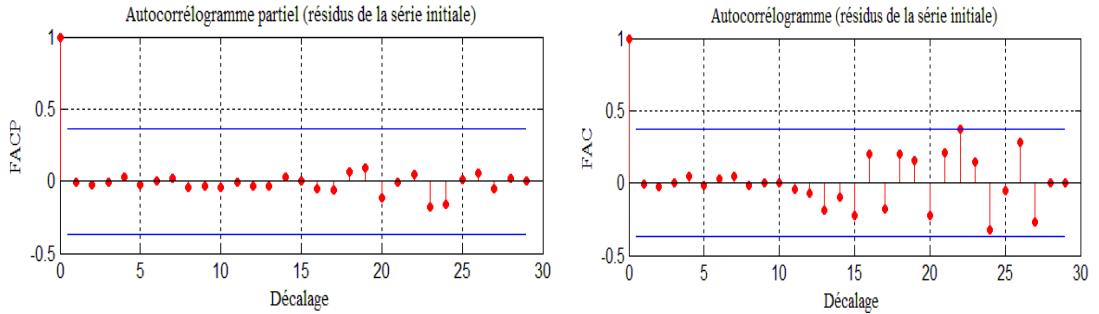
Le modèle ayant la plus petite valeur de MAPE est estimé comme bon modèle. Les bons modèles pour les séries de la sous zone A sont donnés par le tableau 17.

Tableau 17 : Modèles estimés

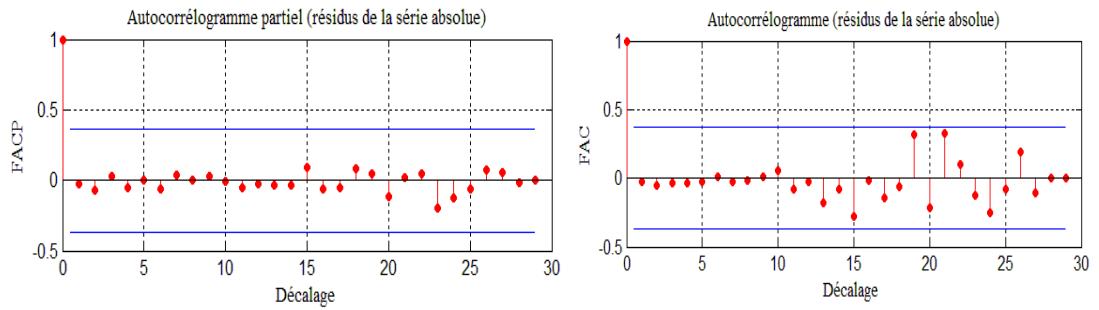
Séries	Initiale	Absolue	Relative 1	Relative 2
Modèle	ARIMA(0,1,16)	ARIMA(0,1,13)	ARIMA(0,1,18)	ARIMA(0,1,18)

Les correlogrammes de la fonction d'autocorrélation (FAC) et fonction d'autocorrélation partielle (FACP) des résidus des 4 série à modéliser indiquent que ces résidus ne sont pas corrélés(Fig.22).

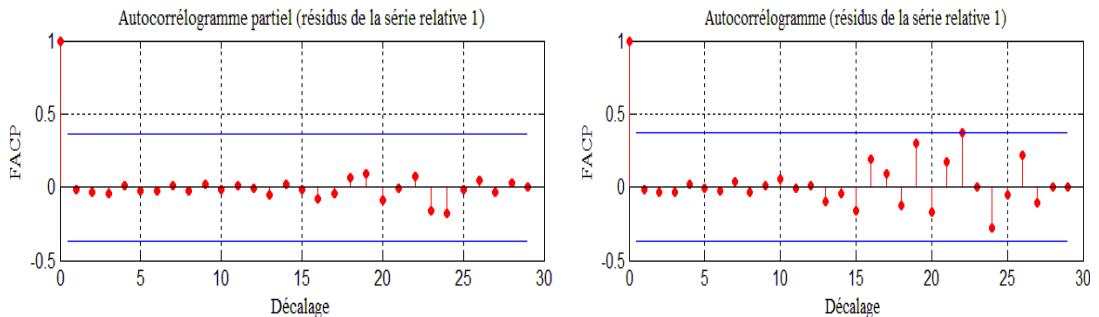
a)



b)



c)



d)

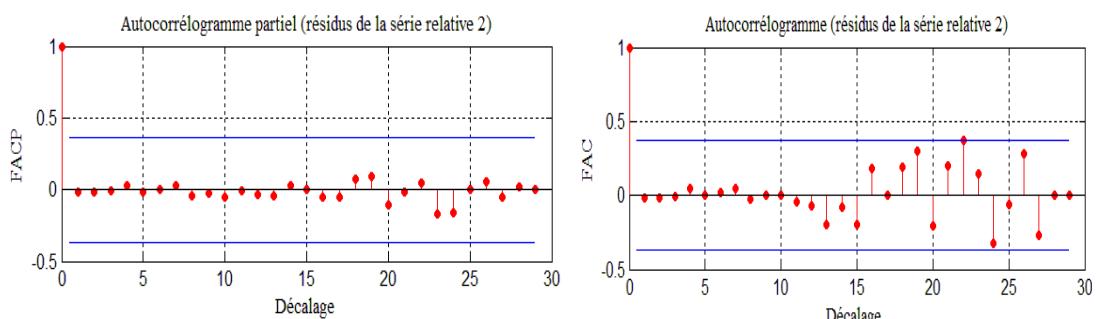


Figure 22 : Corrélogramme de la fonction d'autocorrélation simple et partielle de résidus : a) résidus de la série initiale , b) résidus de la série absolue , c) résidus de la série relative 1 et d) résidus de la série relative 2

Vérification de la qualité de l'ajustement du modèle :

- Corrélation des résidus

Les p-values du test de Ljung-Box sont supérieurs au seuil 0,05 (Tableau 18). Donc les résidus ne sont pas corrélé.

Tableau 18 : Corrélation des résidus

Test de Ljung-Box		
Résidu	p-value	Conclusion
Résidu initiale	0,9514	Non corrélé
Résidu absolue	0,8793	Non corrélé
Résidu relative 1	0,9202	Non corrélé
Résidu relative 2	0,9364	Non corrélé

- Normalité des résidus

Le tableau 19 montre les p-values des tests de normalités, ces p-values sont supérieurs à 0,05. Donc les résidus suivent une loi normale.

Tableau 19 : Résultat des tests de normalités des résidus

Test de normalité de Shapiro-Wilk		
Résidu	p-value	Conclusion
Résidu initiale	0,101	Normale
Résidu absolue	0,299	Normale
Résidu relative 1	0,130	Normale
Résidu relative 2	0,117	Normale

- Homoscédasticité des résidus [33][34]

Tableau 20 : Résultat des tests d'homoscédasticité des résidus

	Résidu initiale	Résidu absolue	Résidu relative 1	Résidu relative 2	Conclusion
p-value	0,125	0,107	0,184	0,118	Homoscédaste
alpha	0,05	0,05	0,05	0,05	

Les résultats de tests de normalités (Tableau 19) et homoscédasticité (Tableau 20) des résidus montrent que les p-valeurs sont supérieurs à 0,05. Qui impliquent que résidus suivent une distribution normale et homoscédaste. Donc les résidus des séries à modéliser peuvent être assimilés à des bruits blancs. On peut alors conclure que les modèles estimés sont bien ajustés.

III.8.4. Prévision

La figure 23 représente les prévisions des températures moyennes de la sous zone A de l'année 2009 à 2015. Ces prévisions sont utilisées pour comparer les valeurs exactes des séries et les valeurs calculées à partir de ces modèles. Le modèle ARIMA(0,1,13) de la série moyenne des données après homogénéisation absolue a la minimale

de moyenne des erreurs absolue 0,0647. Les valeurs calculées à partir de ce modèle sont les plus proches des valeurs exactes par rapport aux autres modèles.

La courbe continue représente la série des températures moyennes de la sous zone A après homogénéisation absolue.

Le modèle ARIMA de la série absolue est représenté par la courbe discontinue moins grasse. L'intervalle de prévision au seuil de 95% est représenté par la courbe plus grasse discontinue.

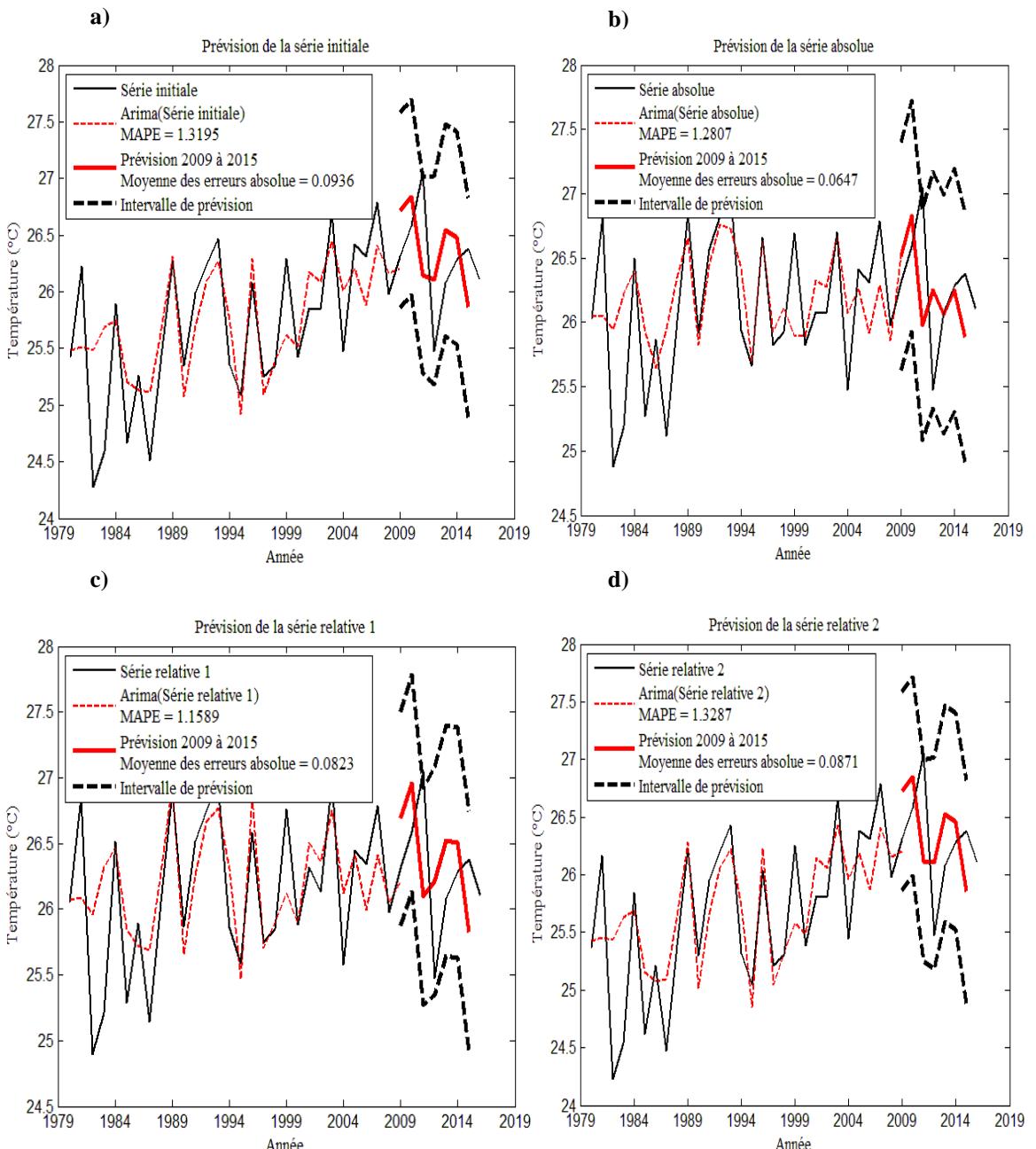


Figure 23 : Prévision des séries de la sous zone A : a) série initiale, b) série absolue, c) série relative 1 et d) série relative 2

Pour la sous zone A, la figure 24 représente la prévision finale fiable des températures moyenne de. La courbe continue en noire représente la série absolue, c'est-à-dire la série à modéliser pour les données après homogénéisation absolue. La courbe continue en rouge donne le modèle ARIMA correspondant au série absolue, le MAPE de ce modèle est 1,28%. Donc ce modèle est excellent selon la classification de MAPE (Tableau 2). La courbe en grasse noire représente l'intervalle au seuil de 95%. La courbe grasse en rouge représente les températures moyennes calculées à partir du modèle ARIMA.

Il en est de même pour la zone entière (figure 25).

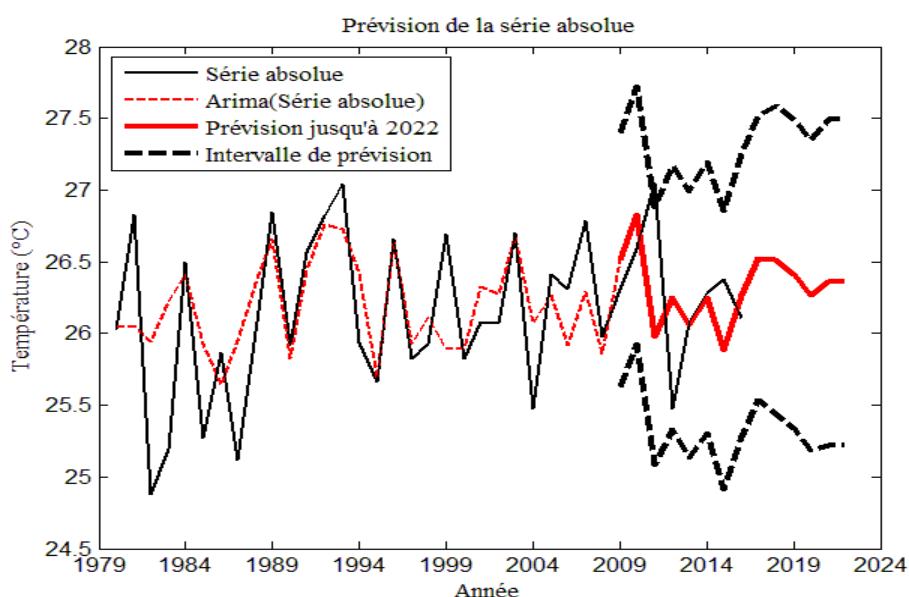


Figure 24 : Prévision finale de la série absolue de la sous zone A

Equation du modèle de la sous zone A :

D'après l'équation (1.3), nous avons l'équation suivante :

$$(1-L) X_t = (1 - 1,14908L + 0,152307L^2 + 0,235855L^3 - 0,090956L^4 - 0,335056L^5 - 0,0507922L^6 + 0,00892051L^7 + 0,299803L^8 - 0,546728L^9 + 0,482197L^{10} - 0,123531L^{11} + 0,510357L^{12} - 0,393293L^{13}) \varepsilon_t \quad (9)$$

X_t : série moyenne de la sous zone A après homogénéisation absolue

ε_t : bruit blanc

L : opérateur de décalage tel que $L^m X_t = X_{t-m}$

Pour la modélisation et la prévision des séries des températures moyennes de la zone entière, nous avons utilisés les mêmes procédures que les séries de températures moyennes de la sous zone A (ANNEXE E).

Et les résultats montrent que la prévision plus fiable est aussi d'utilisé les données après homogénéisation absolue car elle a la minimale de moyenne des erreurs absolue qui est 0,0875 (ANNEXE E).

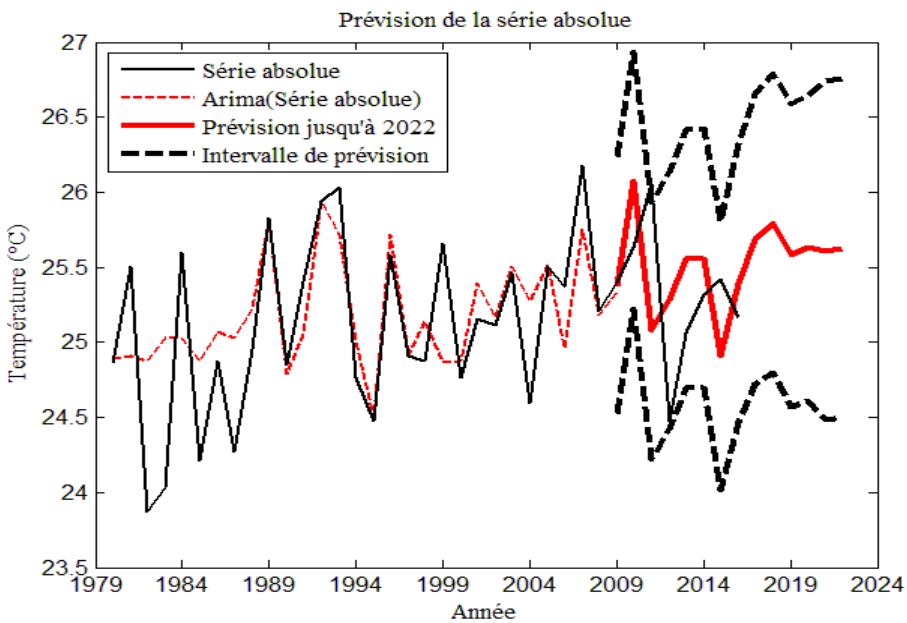


Figure 25 : Prévision finale de la série absolue de la zone entière

Equation du modèle de la zone entière :

D'après l'équation (1.3), nous avons l'équation suivante :

$$(1-L) X_t = (1 - 1.07903L + 0.0949029L^2 - 0.00990391L^3 - 0.075832L^4 - 0.0201689L^5 - 0.186885L^6 - 0.0177837L^7 + 0.596432L^8 - 0.59472L^9 + 0.118355L^{10} + 0.0214469L^{11} + 0.731555L^{12} - 0.578373L^{13}) \varepsilon_t \quad (10)$$

X_t : série moyenne de la zone entière après homogénéisation absolue

ε_t : bruit blanc et L est un opérateur de décalage

Le tableau 21 représente les valeurs des prévisions de la température moyenne de l'année 2016 à 2022 pour la sous zone A et la zone entière. Ces valeurs sont comprises dans l'intervalle de confiances au seuil de 5%.

Tableau 21 : Valeurs des prévisions de la température moyenne de l'année 2016 à 2022

Année	Sous zone A	Zone entière
2016	26,264	25,395
2017	26,523	25,688
2018	26,513	25,794
2019	26,414	25,579
2020	26,269	25,6297
2021	26,357	25,611
2022	26,364	25,627

II.9. Discussion :

Lorsqu'on évalue l'homogénéité d'une série, on ne sait pas préalablement si la série de base sera homogène ou si elle contiendra une ou plusieurs inhomogénéités.

La méthode d'Alexanderson a été choisie car elle utilise les séries voisines existant autour de la série de base et elle permet de distinguer les deux types de ruptures. La méthode de Buishand et Pettitt sont fréquemment utilisés pour étudier l'homogénéité des séries de donnée climatiques. En plus pour ces trois méthodes, une série inhomogène s'ajuste facilement et elles sont rapidement exécutables.

Les données sont des données de réanalyse qui sont censé être homogène. En plus nous avons rencontré des problèmes sur la décision pour les ajustements puisqu'il n'y a pas de metadonnées. Nous avons supposé que les ruptures ne sont pas d'origine climatique. Nous avons corrigé toutes les ruptures dans les séries.

Après le test de normalité de Shapiro-Wilk, toutes les séries suivent une loi normale même après ajustement donc seul le test de Buishand est utilisé. Les dates de rupture varient de l'année 1987 à 2004 pour les séries hétérogènes.

Pour faire une approche plus que possible de la réalité, l'information des séries voisines sont utilisés pour l'homogénéité relative.

L'analyse en composante principale a montré que la zone d'étude peut être divisée en deux sous zones pour les données avant et aussi après l'homogénéisation. Mais ces deux sous zones sont tous différents. Nous avons pris la sous zone A et la zone entière pour faire la modélisation.

En effet, l'évidence devrait le fait que les types de rupture dans les séries sont différents, de même les dates et les amplitudes. Les dates de ruptures ne dépassent pas l'année 2008 pour tous les tests. Donc à partir de l'année 2009 les séries sont homogènes. Les amplitudes des ruptures pour l'homogénéité absolue sont entre -0,75 et -1,25 écart types. Elles sont négatives car il y avait une diminution de la valeur de la moyenne des douze avant et après la rupture des séries. Ces ruptures se traduisent donc par des chutes brusques des moyennes. Et pour l'homogénéité relative, les amplitudes des ruptures sont entre 0,25 et 0,75 écarts types. Elles sont dues à l'augmentation brusque des moyennes.

En fait, les séries sont homogènes (hétérogènes) selon deux types de test et d'autres sont homogènes pour l'un et hétérogènes pour l'autre.

Nous posons que la méthode d'homogénéisation performante est la méthode qui a la minimale d'erreur de prévision.

Après la modélisation et la prévision des données, on trouve que la méthode d'homogénéité absolue a la minimale d'erreur de prévision. Pour une série contenant des inhomogénéités, la correction réduit l'erreur de prévision. Pour les données de températures de la sous zone A : avant la correction, l'erreur est 0,0936 et après les erreurs sont 0,0647, 0,0823 et 0,0871. Pour les données de la zone entière : avant la correction, l'erreur est 0,2114 et après les erreurs sont 0,0875, 0,2113 et 0,092. Nous pouvons donc dire qu'en générale les ruptures dans les séries de base sont d'origine non climatique (artificielle). Et la méthode d'homogénéité absolue (Buishand) qui est plus performante que la méthode d'homogénéité relative (SNHT).

CONCLUSION

En guise de conclusion, on peut dire que ce travail nous a permis de puiser beaucoup de connaissance sur l'homogénéisation des données. En particulier les températures dans la région Atsimo Andrefana de Madagascar.

Nous avons utilisé quatre tests durant ce travail : le test de normalité de Shapiro-Wilk, le test de l'homogénéité absolue de Buishand, le test de Mann Kendall pour la stabilité des moyennes des séries et le test d'homogénéité relative d'Alexanderson.

- Le test de Shapiro-Wilk a montré que les températures suivaient une distribution normale.
- Le test de Buishand et de Mann Kendall nous a permis de connaître les caractéristiques des séries de températures annuelles sur l'homogénéité absolue.
- Le test de Mann Kendall nous renseigne sur la stabilité de la moyenne des séries.
- La méthode d'Alexanderson (SNHT) a donné des informations sur l'homogénéité relative telles que : les nombres, les types ainsi que l'amplitude des ruptures. Elle a permis de faciliter les ajustements des séries hétérogènes.

Nous avons détecté trois types de ruptures (saut, tendance, saut avec tendance) et réussi à corriger ces ruptures.

Nous sommes parvenus à développer et comparer trois approches pour l'homogénéisation, d'abord en utilisant l'homogénéité absolue : sans série de référence (Buishand), ensuite l'homogénéité relative, en utilisant comme série de référence les résultats de l'homogénéité absolue, enfin en utilisant toutes les séries voisines de la série de base comme référence.

L'analyse en composantes principales nous a permis de connaître que la zone d'étude peut être divisé en deux sous zones pour les données avant et après homogénéisation. Le calcul de l'exposant de Hurst et de l'autocorrélation des séries à modéliser évoque un effet mémoire à long terme.

La modélisation ARIMA a été utilisée pour les quatre séries afin d'évaluer la prévisibilité de la température et de comparer les méthodes d'homogénéisation. Nous avons pu calculer la température de l'année 2016 jusqu'à 2022. Les modèles sont excellents selon la classification de MAPE.

La prévision indique que c'est la méthode d'homogénéisation absolue qui nous donne des données fiables parmi ces méthodes.

En guise de perspective, la méthode d'homogénéisation pourrait être améliorée si l'on disposait plus d'information sur les données (métadonnées) ou de longues séries de données.

On pourra faire l'étude de l'homogénéité de ces données à l'échelle mensuelle pour avoir une longue série de données, ou encore utiliser d'autres méthodes d'homogénéisation puis comparer les résultats pour savoir lesquelles de ces méthodes sont plus appropriée pour les données de températures dans la région d'Atsimo Andrefana de Madagascar.

REFERENCES BIBLIOGRAPHIQUES

- [1] Guide des pratiques climatologiques Organisation météorologique mondiale, OMM-N°100, 2011.
- [2] Marcel Mateescu, étude statistique des chroniques des paramètres climatiques en Europe dans la période instrumentale, 2011.
- [3] Claudie Beaulieu, Taha B.M.J. Ouarda, Ousmane Seidou, Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations, *Hydrological Sciences–Journal–des Sciences Hydrologiques*, 52(1), 2007.
- [4] Claudie Beaulieu, Taha B.M.J. Ouarda, Ousmane Seidou, revue bibliographique des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations, *Rapport de recherche No R-811 Janvier 2005*.
- [5] <http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/parametres-observees/temperature>. Consulté le 13 Février 2017.
- [6] R. Ramousse, M. Le Berre, L. Le Guelte, Introduction aux statistiques <http://www.cons-dev.org/elearning/stat/index.html>. Consulté le 13 Février 2017
- [7] Jean Vaillant. Principe des tests statistiques, 2016.
- [8] Mignon Valérie. Méthodes d'estimation de l'exposant de Hurst. Application aux rentabilités boursières. In: *Économie & prévision*. Numéro 132-133, 1998-1-2. pp. 193-214.
- [9] FRIKHA MOUNA, Persistance de la performance des fonds de couverture: approche non paramétrique, 2008.
- [10] O.ROUSTANT, O.RO Introduction aux séries chronologiques, axe méthodes statistiques et applications, 51 p, 2008.
- [11] Desbois Dominique, Une introduction à la méthodologie de Box et Jenkins : l'utilisation de modèles ARIMA avec SPSS, 2005.
- [12] G. M. JENKINS, "Practical Experiences with Modelling and Forecasting Time Series" in *forecasting*, O.D. Anderson editor, North-Holland. 1979.
- [13] RAZANATOMPOHARIMANGA.N. H. Prévision des Périodes de Pluies (début et fin) dans la partie Centre-Ouest de Madagascar. THESE de Doctorat en PHYSIQUE, 2016.
- [14] AnClim – software for times series analysis (for windows). Stepanek, P. Brno, Czech Republic, 2005.

- [15] Lemaitre Fabien, Travail de fin d'études, 2002.
- [16] H. Lubès-Niel et al. Variabilité climatique et statistiques. Etude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité des chroniques.
- [17] Buishand, T. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, 58, 11-27, 1982.
- [18] Buishand, T. Test for detecting a shift in the mean of the hydrological time series. *J. Hydrol.*, 73, 51-69, 1984.
- [19] N. Croiset, B.Lopez (BRGM) Hype : Outils d'analyse statistique des séries temporelles d'évolution de la qualité des eaux sous terraines.
- [20] Alexandersson H and Moberg A. Homogenization of swedish temperature data. Part 1:Homogeneity test for linear trends. *International journal of climatology* 17:25-34, 1997.
- [21] M. N. Khaliq and T. B. M. J. Ouarda. Short Communication On the critical values of the standard normal homogeneity test (SNHT). *International journal of climatology*.27:681-687, 2007.
- [22] Alexandersson H. A homogeneity test applied to precipitation data. *Journal of climatology* 6:661-675, 1986.
- [23] Thorsten Pohlert. Non-Parametric Trend Tests and Change-Point Detection, 2016.
- [24] Shapiro S.S. et Wilk M.B. An analyse of variance test for normality (complete samples). *Biometrika* ,55,2 and 3, 591-597, 1965.
- [25] PettittA.N, A non-parametricapproach to the change-point problem. *AppliedStatistics*, 28, n°2, pp 126-135, 1979.
- [26] Sinan Sahin and H. Kerem Cigizoglu : Homogeneity analysis of Turkish meteorological data set. *HYDROLOGICAL PROCESSES*. *Hydrol. Process.* 24, 981–992 (2010).
- [27] T.C.Peterson et al. Homogeneityadjustments of in situ atmospheric climate data: areview. *International journal of climatology*.18:1493-1517, 1998.
- [28] Alexandersson H and Moberg A, Homogenization of swedish temperature data. Part 2 : Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International journal of climatology*, vol. 17, 35–54 , 1997.
- [29] C. Duby, S. Robin, Département O.M.I.P. :« Analyse en Composantes

Principales ».

[30] R. RAKOTOMALALA : "Analyse en Composantes Principales(ACP) principe et pratique de l'ACP".

[31] Gemtech,"Interprétation d'une ACP," campus-douai.gemtech. SP-ADD2 http://campus-douai.gemtech.fr/pluginfile.php/13713/mod_resource/content/0/SP-ADD2.pdf

[32] J.F. DELMAS, Saad SALAM, Analyse en composante principales. 2006.

[33] www.xlstat.com

[34] RAKOTOVAO N.A, Influence de la variabilité intrasaisonnière de la convection dans le canal de Mozambique, des phénomènes enso et mjo sur la pluviométrie de la côte ouest de Madagascar. THESE de Doctorat en PHYSIQUE, 2014.

ANNEXES

ANNEXE A : TABLEAU DES VALEURS CRITIQUES D'ALEXANDERSSON

n	10	20	30	40	50
T ₉₀	5,05	6,1	6,65	7	7,25
T ₉₅	5,7	6,95	7,65	8,1	8,45
T _{95,5}	6,25	7,8	8,65	9,25	9,65

ANNEXE B : TEST DE NORMALITE

Série	P-value	Série	P-value	Série	P-value
C1	0,946	E3	0,437	F7	0,727
C2	0,214	E4	0,375	F8	0,771
C3	0,209	E5	0,562	F9	0,74
C4	0,283	E6	0,538	G1	0,811
D1	0,293	E7	0,574	G2	0,851
D2	0,305	E8	0,829	G3	0,9
D3	0,268	F1	0,534	G4	0,722
D4	0,222	F2	0,399	G5	0,996
D5	0,688	F3	0,487	G6	0,966
D6	0,638	F4	0,649	G7	0,969
D7	0,785	F5	0,91	G8	0,916
E1	0,431	F6	0,559	G9	0,605
E2	0,571				

ANNEXE C : TEST DE MANN KENDALL

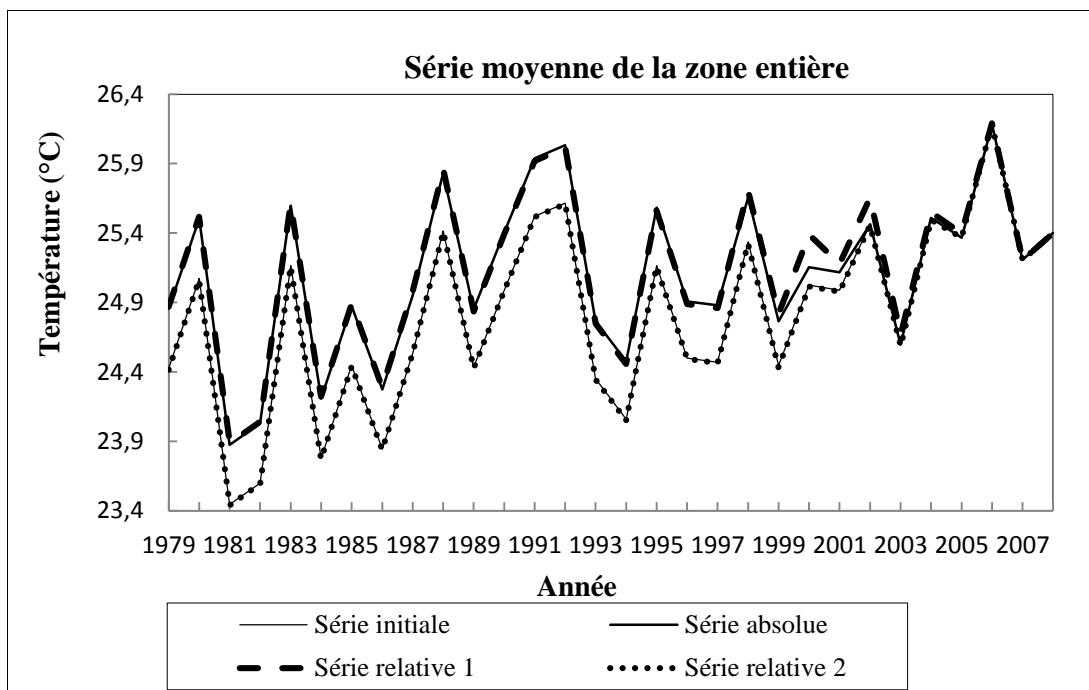
Série	p-value(en %)	Série	p-value(en %)
C1	0,01	E4	0,05
C2	0,01	E5	0,12
C3	0,01	E6	0,05
C4	0,07	E7	0,11
D1	0,0004	E8	0,16
D2	0,02	F1	0,22
D3	0,02	F6	0,08
D4	0,04	F7	0,16
D5	0,31	F8	0,1
D6	0,15	F9	0,15
D7	0,16	G7	0,11
E1	0,02	G8	0,12
E2	0,03	G9	1,28
E3	0,03		

ANNEXE D : COSINUS CARRES DES VARIABLES

	F1	F2	F3	F4	F5	F6
1979	0,948	0,041	0,003	0,000	0,005	0,000
1980	0,963	0,017	0,014	0,001	0,002	0,000
1981	0,894	0,069	0,027	0,007	0,000	0,001
1982	0,933	0,035	0,029	0,001	0,000	0,001
1983	0,985	0,001	0,002	0,001	0,003	0,003
1984	0,949	0,001	0,047	0,000	0,001	0,000
1985	0,968	0,009	0,014	0,000	0,008	0,000
1986	0,880	0,026	0,065	0,020	0,002	0,004
1987	0,991	0,001	0,000	0,000	0,005	0,000
1988	0,969	0,010	0,013	0,000	0,006	0,001
1989	0,974	0,005	0,017	0,000	0,002	0,000
1990	0,964	0,024	0,009	0,001	0,001	0,000
1991	0,847	0,129	0,015	0,002	0,000	0,000
1992	0,931	0,055	0,005	0,002	0,000	0,001
1993	0,985	0,000	0,006	0,003	0,001	0,004
1994	0,984	0,007	0,006	0,001	0,000	0,000
1995	0,954	0,009	0,032	0,003	0,000	0,000
1996	0,956	0,028	0,001	0,003	0,008	0,000
1997	0,989	0,001	0,007	0,000	0,000	0,000
1998	0,965	0,014	0,008	0,006	0,000	0,003
1999	0,972	0,010	0,002	0,001	0,001	0,010
2000	0,949	0,021	0,005	0,007	0,005	0,005
2001	0,950	0,032	0,006	0,003	0,001	0,001
2002	0,952	0,003	0,026	0,000	0,010	0,002
2003	0,972	0,001	0,013	0,005	0,001	0,005
2004	0,851	0,136	0,001	0,001	0,001	0,008
2005	0,817	0,129	0,047	0,002	0,000	0,000
2006	0,823	0,018	0,128	0,023	0,003	0,003
2007	0,952	0,008	0,011	0,008	0,018	0,002
2008	0,944	0,000	0,039	0,001	0,006	0,006

ANNEXE E : MODELISATION DES SERIES MOYENNES DES TEMPERATURES DE LA ZONE ENTIERE

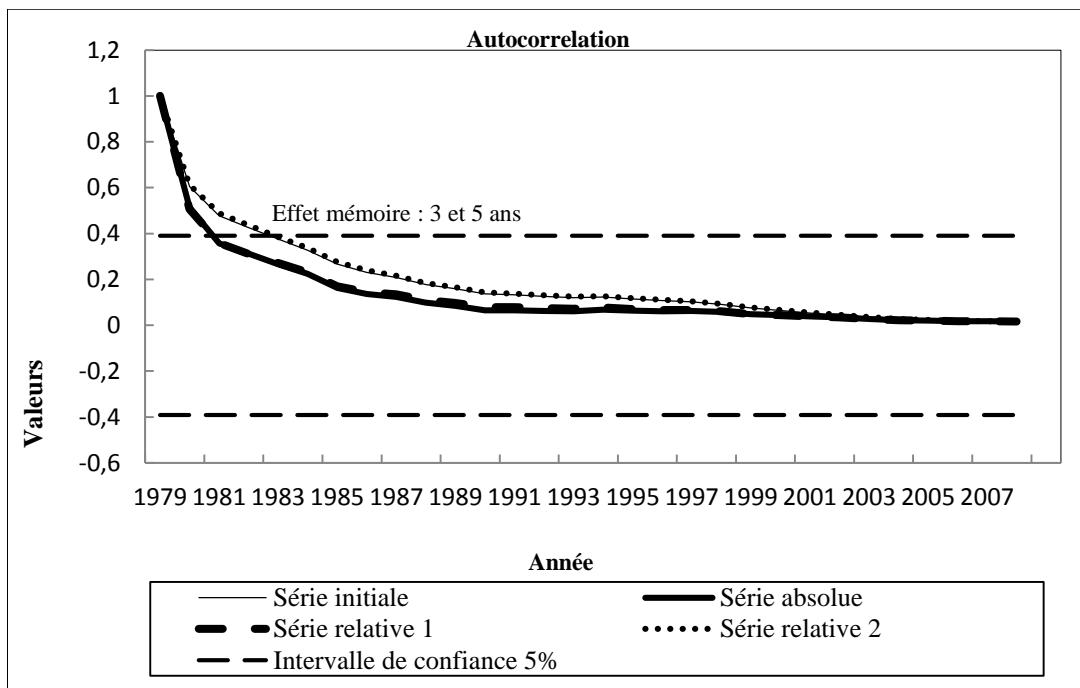
1. Série moyenne de la zone entière



2. Exposant de Hurst et effet mémoire

Série moyenne	Initiale	Homogénéité absolue	Homogénéité relative approche 1	Homogénéité relative approche 2
Exposant de Hurst	0,5975	0,5645	0,5667	0,6006
Effet mémoire	5	3	3	5

3. Courbe d'autocorrélation et intervalle de confiance

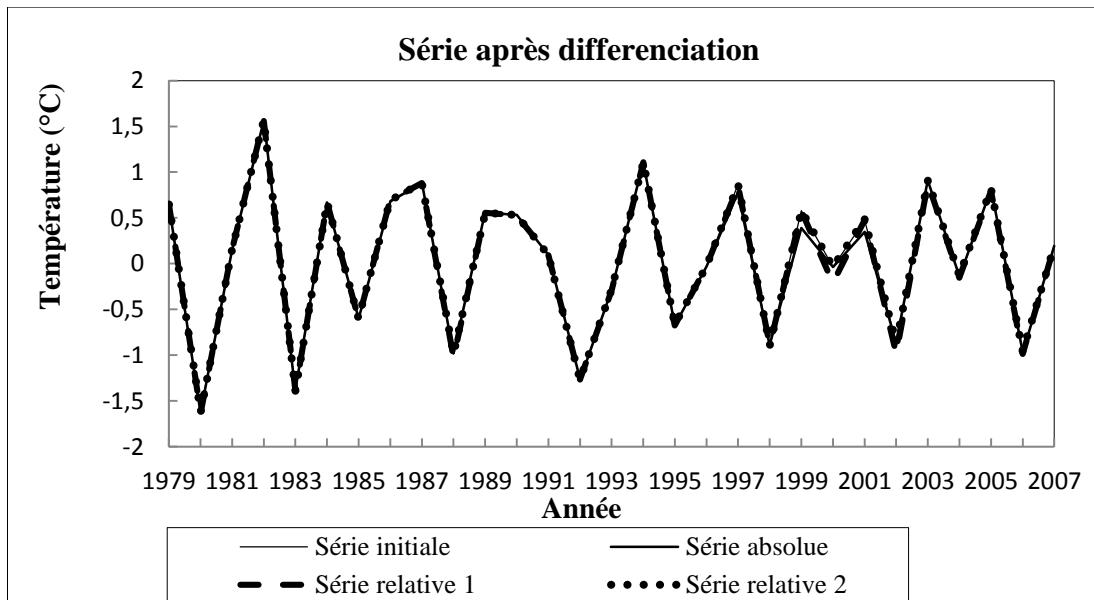


4. Test de stationnarité des séries à modéliser

Toutes les séries sont non stationnaires.

	Série initiale	Série absolue	Série relative 1	Série relative 2
Tau (Valeur observée)	-2,080	-2,174	-2,254	-2,071
Tau (Valeur critique)	-0,508	-0,508	-0,508	-0,508
p-value (unilatérale)	0,508	0,460	0,420	0,513
alpha	0,05	0,05	0,05	0,05

5. Séries après différenciation de premier ordre

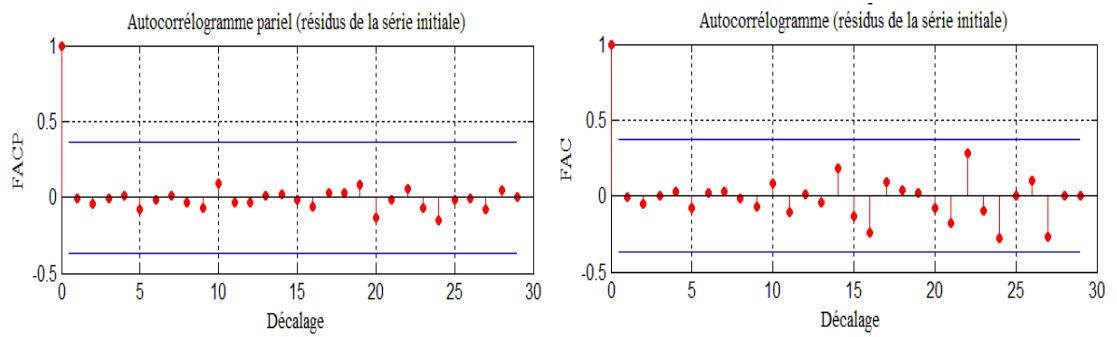


6. Test de stationnarité après différenciation

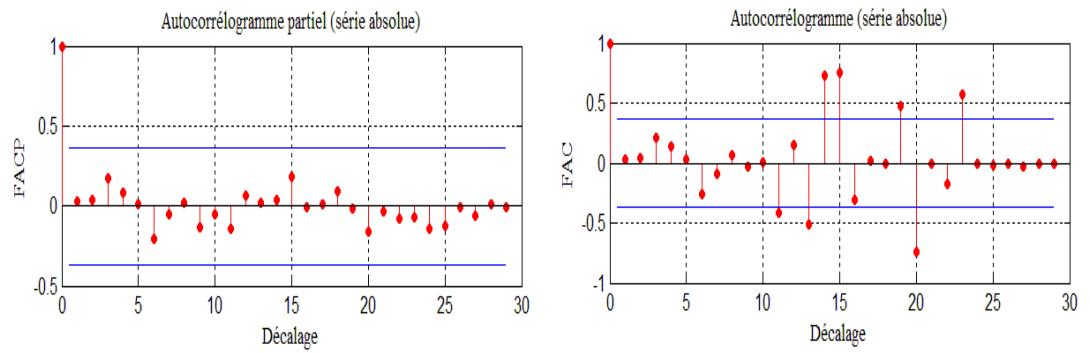
	Série initiale	Série absolue	Série relative 1	Série relative 2
Tau (Valeur observée)	-3,741	-3,787	-3,903	-3,745
Tau (Valeur critique)	-0,484	-0,484	-0,484	-0,484
p-value (unilatérale)	0,036	0,033	0,026	0,035
alpha	0,05	0,05	0,05	0,05

7. Identification de modèle

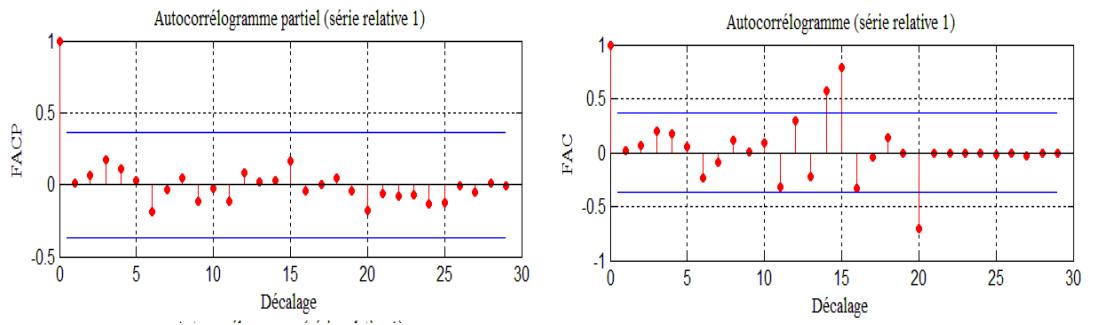
a)



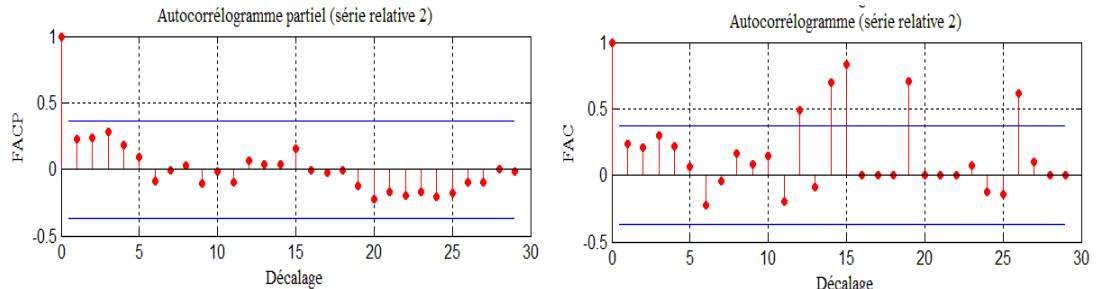
b)



c)



d)



Correlogramme de la fonction d'autocorrélation simple et partielle des série moyenne : a) Initiale , b) Relative , c) Relative 1 et d) Relative 2

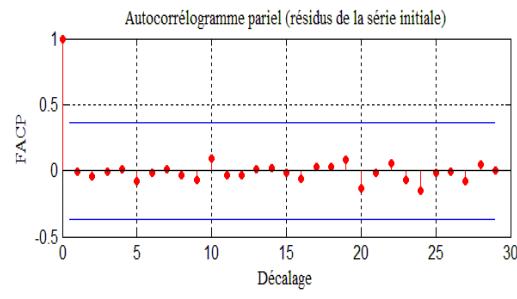
8. Modèle candidat pour la modélisation

Séries	Modèle	MAPE
Initiale	ARIMA(0,1,13)	1,3806
	ARIMA(0,1,17)	1,3105
	ARIMA(0,1,26)	1,233
Absolue	ARIMA(0,1,13)	1,2536
Relative 1	ARIMA(0,1,15)	1,3431
	ARIMA(0,1,19)	1,2273
Relative 2	ARIMA(0,1,17)	1,3235

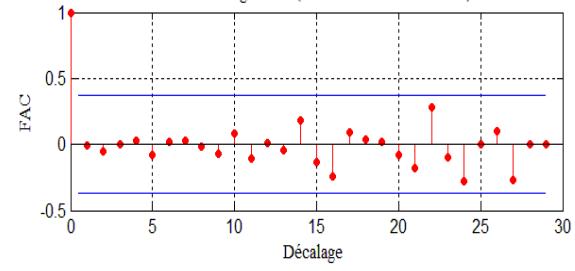
9. Modèle estimé

Séries	Initiale	Absolue	Relative 1	Relative 2
Modèle	ARIMA(0,1,26)	ARIMA(0,1,13)	ARIMA(0,1,19)	ARIMA(0,1,17)

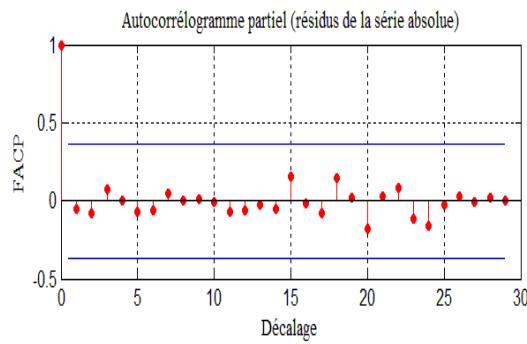
a)



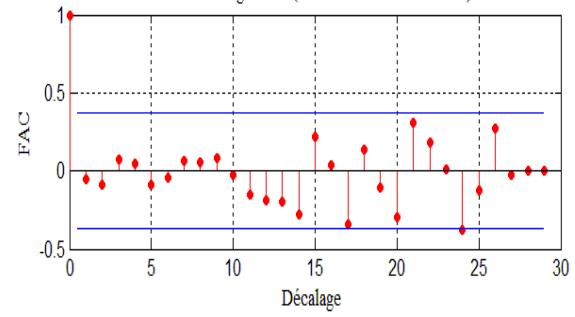
Autocorrélogramme (résidus de la série initiale)



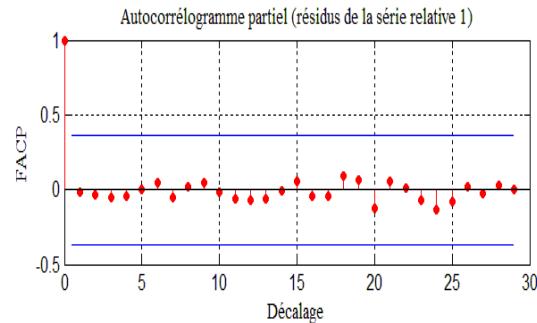
b)



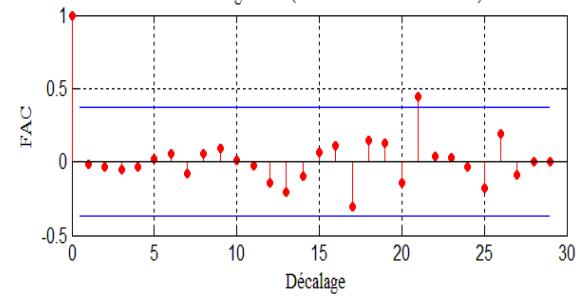
Autocorrélogramme (résidus de la série absolue)



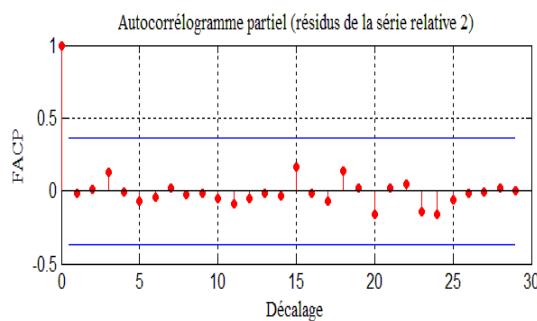
c)



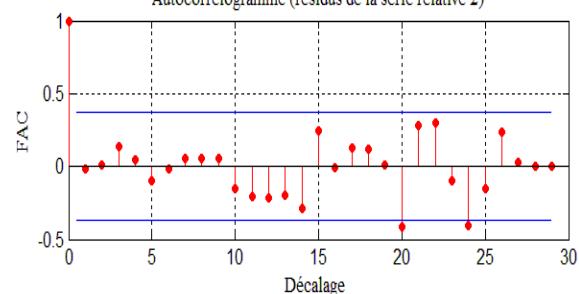
Autocorrélogramme (résidus de la série relative 1)



d)



Autocorrélogramme (résidus de la série relative 2)



Correlogramme de la fonction d'autocorrélation simple et partielle des résidus des séries : a) Initiale , b) Relative , c) Relative 1 et d) Relative 2

10. Vérification de la qualité de l'ajustement du modèle :

- Corrélation des résidus

Test de Ljung-Box		
Résidu	p-value	Conclusion
Résidu initiale	0,999	Non corrélé
Résidu absolue	0,987	Non corrélé
Résidu relative 1	0,999	Non corrélé
Résidu relative 2	0,999	Non corrélé

- Normalité des résidus

Test de normalité de Shapiro-Wilk		
Résidus	p-value	Conclusion
Résidu initiale	0,191	Normale
Résidu absolue	0,137	Normale
Résidu relative 1	0,196	Normale
Résidu relative 2	0,223	Normale

- Homoscédasticité des résidus

	Résidu initiale	Résidu absolue	Résidu relative 1	Résidu relative 2	Conclusion
p-value	0,242	0,153	0,172	0,092	Homoscédaste
alpha	0,05	0,05	0,05	0,05	Homoscédaste

11. Moyenne des erreurs absolues

Série	Série initiale	Série absolue	Série relative 1	Série relative 2
Moyenne des erreurs absolue	0,2114	0,0875	0,2113	0,092

HOMOGÉNÉISATION DES DONNÉES DE TEMPÉRATURES MOYENNE DANS LA RÉGION D'ATSIMO ANDREFANA DE MADAGASCAR. MODÈLE DE PRÉVISION DE CES TEMPÉRATURES.

Résumé : Ce travail consiste à étudier l'homogénéité des données de températures moyennes annuelles de 1979 à 2015 dans la région d'Atsimo Andrefana de Madagascar et de concevoir des modèles de prévision de ces données. La zone d'étude est située entre 21,5°S à 26,5°S de latitude et 42,5°E à 45,5°E de longitude. L'objectif de ce travail est d'avoir des données et prévisions fiables. Pour cela, nous avons utilisé la méthode d'homogénéisation absolue de Buishand et la méthode d'homogénéisation relative d'Alexanderson. Les données homogènes sont 27% pour l'homogénéité absolue, 8% pour l'homogénéité relative de première approche et 47% pour la deuxième approche. Les ruptures sauts, tendance, saut avec tendance sont détecté et tous corrigées. Pour les données avant et après homogénéisation : la zone d'étude a été divisée en 2 sous zones (Zone A à l'Ouest et Zone B à l'Est) par l'Analyse en composantes principales. Par la suite, les séries des températures moyennes de la sous zone A et de la zone entière ont été modélisé. Après comparaison de ces modèles, les données après homogénéisation absolue sont les plus fiables. Nous avons pu prévoir la température moyenne de la zone A et de la zone entière jusqu'à 2022.

Mots clé : série homogène, série hétérogène, homogénéité absolue, homogénéité relative, Analyse en Composantes Principales, test statistique, ARIMA.

HOMOGENIZATION OF AVERAGE TEMPERATURE DATA IN THE ATSIMO ANDREFANA AREA OF MADAGASCAR. MODEL FORECASTING THESE TEMPERATURES.

Abstract : This work consists of studying the homogeneity of the annual mean temperature data from 1979 to 2015 in the Atsimo Andrefana region of Madagascar and design models for predicting this data. The study area is located between 21.5 ° S and 26.5 ° S latitude and 42.5 ° E at 45.5 ° E longitude. The objective of this work is to have reliable data and forecasts. For this, we used the method of absolute homogenization of Buishand and the method of relative homogenization of Alexanderson. The homogeneous data are 27% for the absolute homogeneity, 8% for the relative homogeneity of first approach and 47% for the second approach. Shifts breaks, trend, trend with shift are detected and all corrected. For data before and after homogenization: the study area was divided into 2 sub-zones (Zone A West and Zone B East) by Principal Component Analysis. Subsequently, the average temperature series of sub-zone A and entire zone were modeled. After comparison of these models, the data after absolute homogenization are the most reliable. We were able to predict the average temperature of zone A and entire zone up to 2022.

Key words : homogeneous series, heterogeneous series, absolute homogeneity, relative homogeneity, Principal Component Analysis, statistical test, ARIMA.

Encadreur : Mr RABEHARISOA Jean Marc, Maître de conférences Université d'Antananarivo Email : jmrabeharisoa@gmail.com Tél : +261 33 14 410 15	Impétrant : RAFANOMEZANTSOA Herimalala Edmond Email : herimalalaedmond@gmail.com Adresse : CU Ambohipo Bloc 40 P A1 Tél : +261 34 91 764 64
---	--