Table des matières

1 Série Temporelle - La Théorie

1.1 Auto-Régression (AR)

1.1.1 Définition

Un processus (X_t) est Auto-Régressif quand sa valeur à l'instant t n'est expliquée que par ses anciennes valeurs $(X_{t-1},...,X_{t-i})$, où $i \in \{2,...,\infty\}$ et non par d'autres processus.

$$X_t = \phi_1 X_{t-1} + ... + \phi_i X_{t-i} + \epsilon_t, \quad i \in \{2,...,\infty\},$$

où les ϵ_t sont des bruits blancs, indépendants et identiquement distribués, notés $\epsilon_t \to iid(0,\sigma^2)$, $\forall t$.

1.1.2 Ordre d'un AR

Un processus AR est d'ordre p, noté AR(p), quand sa valeur à l'instant t est expliquée par ses p anciennes valeurs:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \epsilon_t,$$

où $\epsilon_t \to iid(0,\sigma^2)$, $\forall t$.

1.1.3 Opérateur de retard

Pour une série temporelle $(X_t)_t$, on définit l'opérateur de retard, noté L, par une application qui à chaque élément X_t de la série, associe son observation précédente X_{t-1}:

$$LX_t = X_{t-1}, \quad \forall t > 1.$$

En particulier, $L^i(X_t) = X_{t-i}$.

1.1.4 Opérateur de différenciation

Pour une série temporelle $(X_t)_t$, on définit l'opérateur de différenciation, noté ∇, par une application qui à chaque élément X_t de la série, associe la différence $X_t - X_{t-1}$:

$$\nabla X_t = X_t - X_{t-1}, \quad \forall t > 1.$$

1.1.5 Polynôme caractéristique

Ayant définit l'opérateur de retard, on peut l'utiliser dans la définition d'un processus AR(p).
Ainsi, si $(X_t)_t$ est un AR(p), alors, on définit le polynôme caractéristique Φ d'un processus AR(p) de tel sorte que: $\Phi(L)X_t = \epsilon_t$,

$$\Phi(L) = 1 - \phi_1 L^1 + ... + \phi_p L^p.$$

1.1.6 Équation caractéristique

On appelle équation caractéristique d'un processus AR(p), l'équation déduit de Φ en remplaçant L par x:

$$(1 - \phi_1 x^1 + ... + \phi_p x^p).$$

1.2 Moyenne Mobile (MA)

1.2.1 Introduction

Une moyenne est dite mobile lorsqu'elle est recalculée de façon continue, en utilisant à chaque calcul un sous-ensemble d'éléments dans lequel un nouvel élément remplace le plus ancien ou s'ajoute au sous-ensemble.

1.2.2 Définition

Un processus est une Moyenne Mobile lorsqu'il est de la forme:

$$X_t = \theta_1 \epsilon_{t-1} + ... + \theta_i \epsilon_{t-i} + \epsilon_t, \quad i \in \{2,...,\infty\},$$

où $\epsilon_t \to iid(0,\sigma^2)$, $\forall t$.

1.2.3 Ordre d'un MA

Un processus MA est d'ordre q, noté MA(q), quand sa valeur à l'instant t est expliquée par ses q anciennes valeurs:

$$X_t = \theta_1 \epsilon_{t-1} + ... + \theta_p \epsilon_{t-q} + \epsilon_t,$$

où $\epsilon_t \to iid(0,\sigma^2)$, $\forall t$.

1.2.4 Polynôme caractéristique

Le polynôme caractéristique Θ d'un processus MA(q) est definit de tel sorte que: $X_t = \Theta(L)\epsilon_t$,

$$\Theta(L) = 1 + \theta_1 L^1 + ... + \theta_q L^q.$$

1.3 Auto-Regression et Moyenne Mobile (ARMA)

1.3.1 Définition

Un processus ARMA $(X_t)_t$ est comme son nom l'indique, un processus auto-regéssif et moyenne mobile. Il a une partie AR(p) et une partie MA(q) et est noté ARMA(p,q) selon la définition:

$$X_t := \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_p \epsilon_{t-q} + \epsilon_t,$$

où $\epsilon_t \to iid(0,\sigma^2)$, $\forall t$.

1.3.2 Stationnarité faible

Un processus $(X_t)_t$ est faiblement stationnaire si: $\forall t$, $\forall h < t$;
- $E(X_t) = \mu$, l'espérence est constante au cours du temps.
- $Var(X_t) = \sigma^2 < \infty$, la variance est constante et non infinie.
- $Cov(X_t, X_{t-h}) = \gamma(h)$, l'auto-corrélation entre X_t et X_{t-h} reste constante et ne dépend que de h.

1.3.3 Stationnarité forte

Un processus $(X_t)_t$ est fortement stationnaire si: $\forall t$, $\forall h$; $(X_1, X_2, ..., X_t)$ et $(X_{1+h}, X_{2+h}, ..., X_{t+h})$ ont même lois en probabilité.

1.3.4 Auto-Corrélation (ACF)

La fonction d'auto-corrélation ρ est définie par:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{E[(X_t - \mu)(X_{t+h} - \mu)]}{\sigma^2}.$$

1.3.5 ACF et Ordre de la MA

Si on observe le graphe de l'ACF, le dernier pic significatif nous donnera l'ordre q de la partie MA du processus.

1.3.6 Auto-Corrélation Partielle (PACF)

La fonction d'auto-corrélation partielle L est définie par:

$$L(0) = 1,$$
$$L(h) = \phi_{hh}, \forall h > 0,$$

avec $\phi_h = \Gamma_h^{-1} \gamma_h$, $\Gamma_h = (\gamma(i-j))_{1 \leq i,j \leq h}$, $\gamma_h = (\gamma(1), \gamma(2), ..., \gamma(h))$.

1.3.7 PACF et Ordre de l'AR

Si on observe le graphe de la PACF, le dernier pic significatif nous donnera l'ordre p de la partie AR du processus.

1.4 Auto-Regression, Integrée et Moyenne Mobile (ARIMA)

1.4.1 Définition

Les modèles ARIMA sont des modèles qui se réduisent à des ARMA une fois différentiés un nombre fini de fois. On a alors la notation ARIMA(p,d,q), où
- p: ordre de la partie AR,
- d: ordre de différentiation,
- q: ordre de la partie MA.

1.4.2 Définition 2

Un modèle ARIMA(p,d,q) est de la forme:

$$A(L)(1-L)^d X_t = B(L)\epsilon_t,$$

où A(L) et B(L) sont les polynômes caractéristiques respectives des parties AR et MA.

1.5 Tests avec les séries temporelles

1.5.1 p-value

En statistique, le p-value est une critère d'acceptation de l'hypothèse nulle (H_0):

- si p-value$>$ 0.05, (H_0) est acceptée,
- si p-value$\leq$ 0.05, (H_0) est rejetée.

1.5.2 Test de Dickey-Fuller Augmenté - ADF

Le test ADF est un test de racine unitaire:
(H_0) : Le processus admet une racine unitaire (non stationnaire), contre
(H_1) : Le processus est stationnaire.

1.5.3 Test de Box-Pierce

Le Test de Box-Pierce est un test statistique qui teste l'auto-corrélation:
(H_0) : Aucune corrélation dans les erreurs, contre
(H_1) : Avec corrélation dans les erreurs.

2 Série Temporelle - La Pratique

2.1 Introduction

En analyse théorique, l'étude d'une série temporelle commence par sa forme théorique. Par exemple, l'étude d'un processus ARMA(p,q) $(X_t)_t$ débute par sa forme:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_p \epsilon_{t-q} + \epsilon_t,$$

où les ϕ_i et θ_j sont des constantes données $\forall i < p$, $\forall j < q$.

Par contre, l'analyse pratique d'une série temporelle débute par un tableau de données, et grâce à ces données, on essaie d'estimer les paramètres du processus, et de déterminer les propriétés des résidus.

2.2 Les données

Pour cette analyse, nous allons étudier les variations annuelles de la quantité d'émission de $C0_2$ à Madagascar de 1960 à 2011.

Émissions de CO2 à Madagascar, de 1960 à 2011 (Tonnes métriques par habitant)

annee	emission	annee	emission	annee	emission	annee	emission
1960	0.078382804	1961	0.068094335	1962	0.07398844	1963	0.084212917
1964	0.086046133	1965	0.096613424	1966	0.097901339	1967	0.138895339
1968	0.146498695	1969	0.132919321	1970	0.147766199	1971	0.151879665
1972	0.185133213	1973	0.150219836	1974	0.161421774	1975	0.223628345
1976	0.127410591	1977	0.105468171	1978	0.123284325	1979	0.133181546
1980	0.185728923	1981	0.115518639	1982	0.110956194	1983	0.071306652
1984	0.098900574	1985	0.1083812	1986	0.115036211	1987	0.125664328
1988	0.122370908	1989	0.086393633	1990	0.085435789	1991	0.09030147
1992	0.085217189	1993	0.084686453	1994	0.100956889	1995	0.09758658
1996	0.101694951	1997	0.118742384	1998	0.117520568	1999	0.126134972
2000	0.119012988	2001	0.107283198	2002	0.073839439	2003	0.098664011
2004	0.101772992	2005	0.095231683	2006	0.089405156	2007	0.093705131
2008	0.099740761	2009	0.092499326	2010	0.098983365	2011	0.112992805

FIG. 1 – *Tonnes métriques d'émission de CO2 à Madagascar*

2.3 La serie temporelle

La serie temporelle (Y_t) obtenue à partir de ces données produit la graphe ci-dessous.

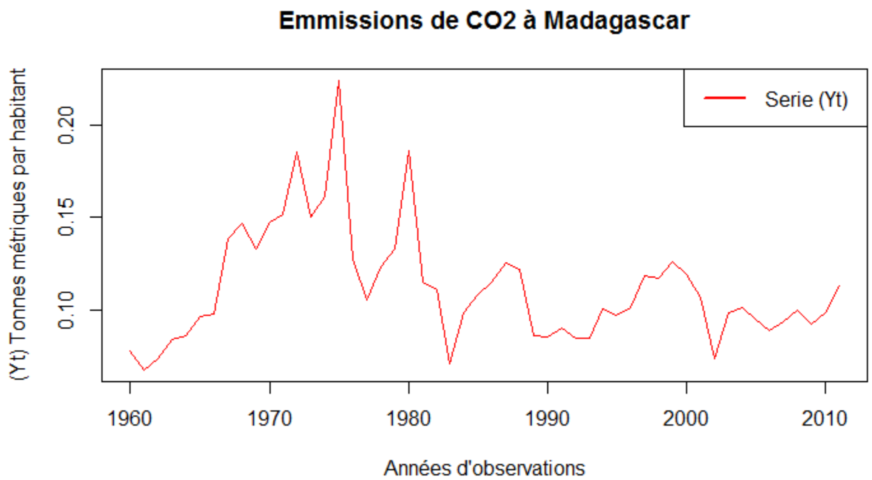

Fig. 2 – *Graphe, Tonnes métriques d'émission de CO2 à Madagascar*

De plus, elles produisent les graphes des ses corrélogrammes ci-dessous.

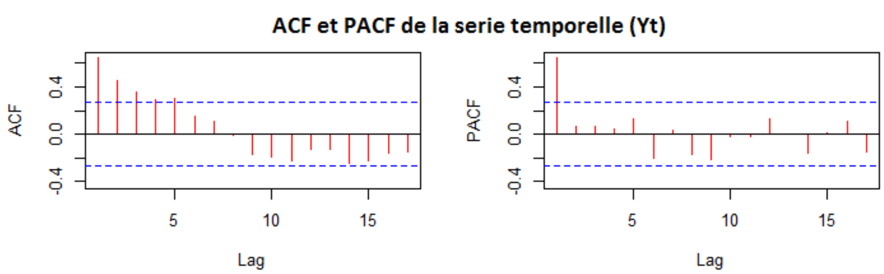

Fig. 3 – *Graphe, Tonnes métriques d'émission de CO2 à Madagascar*

2.3.1 Identification de la série

On peut constater à partir des corrélogrammes que:

- l'ACF donne q=5, ainsi on a MA(5), et
- la PACF donne p=1, alors on a AR(1),

donc, si les résidus sont iid, alors la série (Y_t) est ARMA(1,5).

2.3.2 Stationnarité de la série

Le Test de Dickey-Fuller Augmenté nous donne les résultats suivants:

```
        Augmented Dickey-Fuller Test

data:  Yt
Dickey-Fuller = -2.8825, Lag order = 3, p-value = 0.2194
alternative hypothesis: stationary
```

FIG. 4 – *Test de Dickey-Fuller sur la série (Yt)*

On constate ici que le p-value = 0.2194, ne nous permet pas de rejetter l'hypothèse nulle (H_0), ainsi, la série (Y_t) n'est PAS STATIONNAIRE.

2.4 La série temporelle differenciée

Continuons notre étude, mais utilisons cette fois ci la série différentiée (diff_Y_t) de (Y_t). On a alors sa graphe, ainsi que ses deux corrélogrammes, simple et partiel:

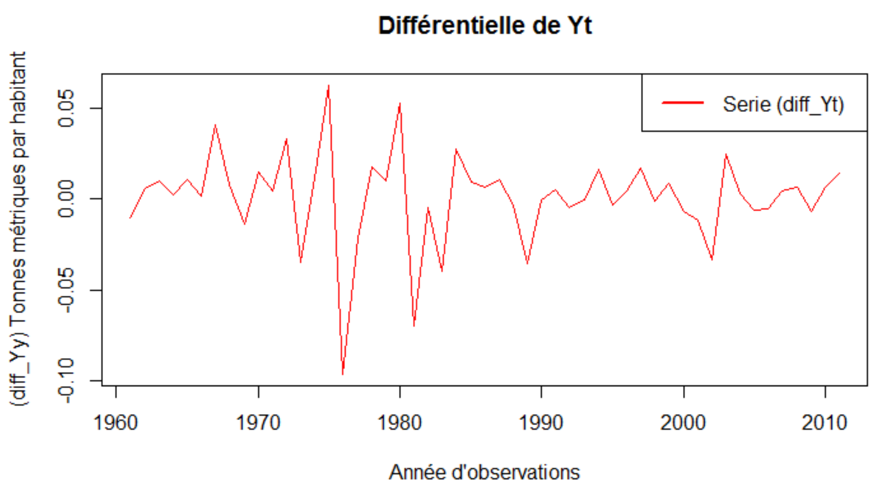

FIG. 5 – *Graphe de la serie (diff_Y_t)*

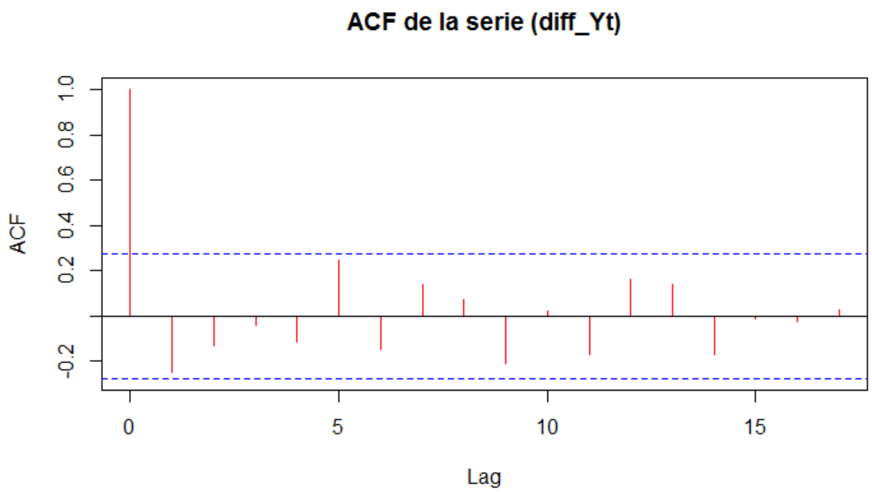

Fig. 6 – *L'ACF de la serie (diff_Y_t)*

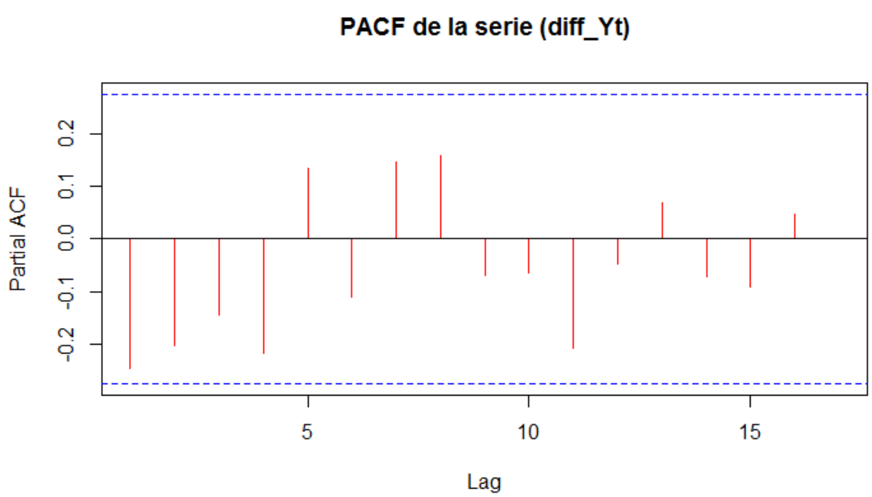

Fig. 7 – *La PACF de la série (diff_Y_t)*

2.4.1 Identification de la série différentiée

On peut constater à partir des corrélogrammes que:

- l'ACF donne q=1, ainsi on a MA(1), et
- la PACF donne p=0, alors on a AR(0),

donc, si les résidus sont iid, alors la série (diff_Y_t) est un processus ARMA(0,1) ou plus précisément un MA(1).

2.4.2 Stationnarité de la série différentiée

Le Test de Dickey-Fuller Augmenté nous donne les résultats suivants:

```
        Augmented Dickey-Fuller Test

data:  diff_Yt
Dickey-Fuller = -5.1604, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Fig. 8 – *Test de Dickey-Fuller sur la série (diff_Y_t)*

On constate ici que le p-value = 0.01, nous incite à rejeter l'hypothèse nulle (H_0), ainsi, la série (diff_Y_t) est STATIONNAIRE.

2.5 Les Résidus

On a donc une série temporelle différentiée (diff_Y_t) stationnaire, avec les ordres:

- p=0, d'après la PACF,
- d=1, l'ordre de différenciation,
- q=1, d'après l'ACF.

Ainsi, on a un processus ARIMA(0,1,1) si les résidus de cette série sont des bruits blanc.

2.5.1 Propriétés des résidus

Les figures ci-dessous sont le graphe et les corrélogrammes simple et partielles des residus du processus issus de (Y_t) et des études précédents.

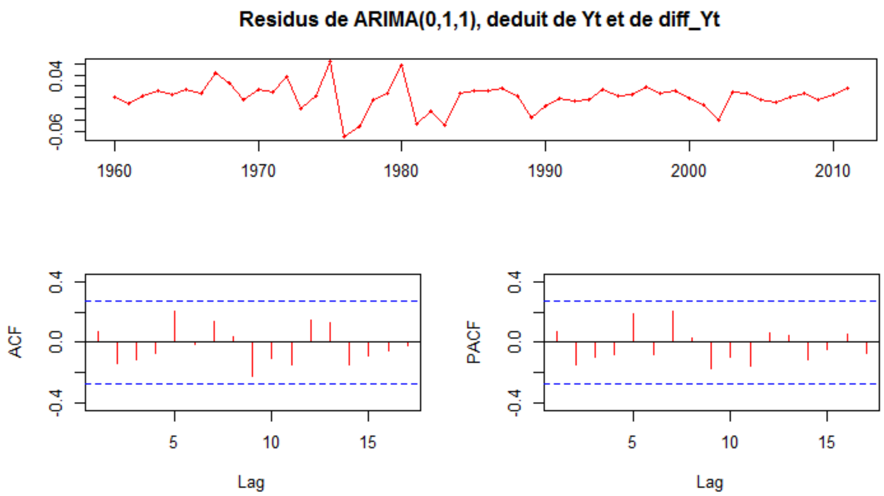

FIG. 9 – *Graphe, ACF et PACF des résidus*

On observe alors qu'il n'y a aucun pic significatif dans les corrélogrammes des résidus.

De plus, le test de Box-Pierce appliqué aux résidus nous donne:

```
        Box-Pierce test

data:  arima$residuals
X-squared = 0.229, df = 1, p-value = 0.6323
```

FIG. 10 – *Test de Box-Pierce sir les résidus*

On en déduit donc que les résidus sont des bruits blancs et, par conséquent le processus issus de (Y_t) et de notre étude est bien un ARIMA(0,1,1).

2.6 Le modèle

Notre étude nous a permis de constater que: $(1-L)^1 Y_t = (1+\theta L)\epsilon_t$, puisque

- le modèle est un ARIMA(0,1,1),
- les résidus sont des bruits blancs.

Il nous reste donc à déterminer θ. Or,

```
Call:
arima(x = Yt, order = c(0, 1, 1))

Coefficients:
          ma1
      -0.4141
s.e.   0.1566

sigma^2 estimated as 0.0005921:  log likelihood = 117.05,  aic = -230.1
```

FIG. 11 – *Propriétés du modèle*

Finalement, le modèle est:

$$(1 - L)X_t = (1 + \theta L)\varepsilon_t$$
$$(1 - L)X_t = (1 - 0.4141L)\varepsilon_t$$

2.7 La prévision

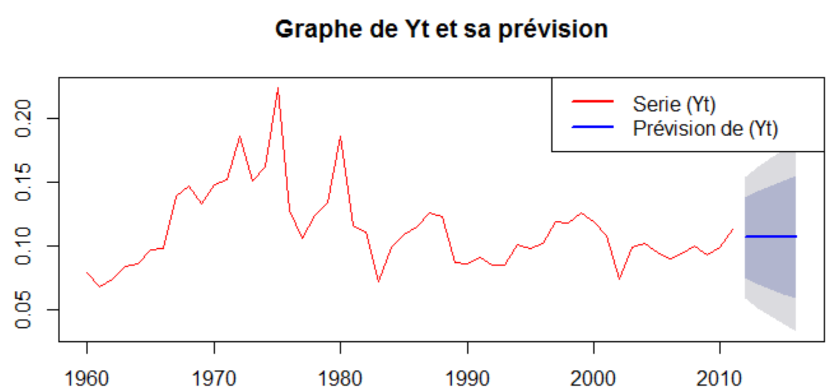

FIG. 12 – *Prévision de la série* (Y_t)

Après avoir vérifié que notre série est bien un ARMA(0,1,1) dont les résidus sont des bruits blancs, on peut trouver une intervalle de prévision pouvant contenir les prochaines valeurs de notre série (Y_t). Ceci est expliqué par le graphe ci-dessus:

2.8 La série ajustée

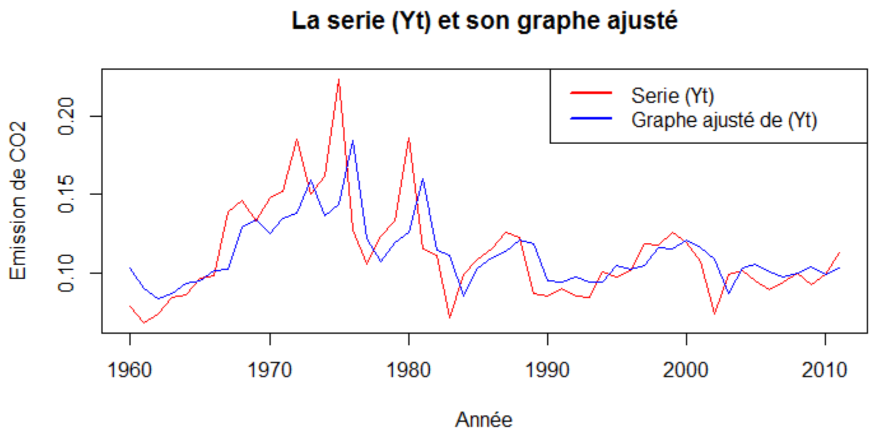

FIG. 13 – *Graphe ajusté de la série* (Y_t)

2.9 La prévision de la série ajustée

Le graphe ajusté de nous donne une intervalle de prévision selon le graphe suivant:

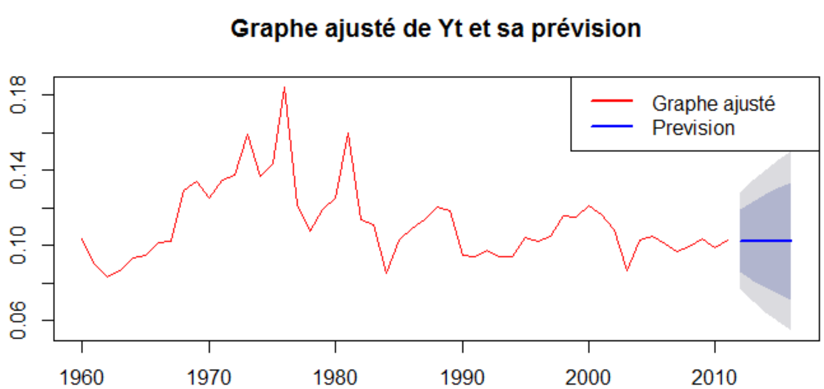

FIG. 14 – *Graphe ajusté de la série* (Y_t) *et sa prédiction*