International Journal of Applied Mathematics

Volume 28 No. 5 2015, 621-636
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v28i5.13

OPTIMIZATION AND ANALYSIS OF THE INDEX OF AIR QUALITY IN DAKAR BY THE PROCESS AUTO REGRESSIVE MOVING AVERAGE ARMA(2,1)

Lebede Ngartera[1] §, Salimata Diagne[2], Youssou Gningue[3]

[1]Department of Mathematics
Faculty of Exact and Applied Sciences N'Djamena University, Chad
B.P: 1739, Ndjamena, CHAD

[2]Department of Mathematics
Faculty of Science and Technology
Cheikh Anta Diop University, Dakar, SENEGAL

[3]Department of Mathematics and Computer Sciences
Laurentian University
Sudbury, ON, CANADA

Abstract: Numerous time-dependent phenomena govern our world. The time series are part of one of the methods often used to understand mathematically. Time series can be used to predict future events. In this article, we use the concept of time series to analyze and model the Index of air quality in Dakar to make his short-term forecast. To top this approach, the auto-regressive moving average(2,1) selected by the optimal Akaike information criterion and Bayesian information criterion was used to perform some simulations.

AMS Subject Classification: 78M50, 65C20, 62P12, 37M10
Key Words: AIC and BIC, air quality, optimization, auto-regressive moving average, simulation and estimation

1. Introduction

The statistical prediction is applied in many fields such as the atmospheric

Received: August 29, 2015

§Correspondence author

science, astronomy, economics, socio-political science, signal processing, etc. Note that a time series or time series is a sequence formed of observations over time. From the knowledge of the previous information, we can estimate the behavior of a system in the future. If the estimate of the future state of the system is accurate, we speak of a method of entirely deterministic prediction. In fact, several factors make the exact calculation of the future state of the system impossible. However, it is possible to generate a model that can be used to calculate the probability of a range of future behaviors between two specified limits. Such a model is called a stochastic model and stochastic processes. An important class of stochastic models is used for the description detrending called class stationary stochastic models. These models assume that time series properties are invariant under the time translation. These models include the autoregression models (AR), Moving Average models (MA) and autoregressive moving average models (ARMA). The processes used for the description of non-stationary time series (average, variance and others) are: ARIMA, SARIMA, ... ARIMA and SARIMA models are extensions of ARMA class in order to include more realistic dynamics, in particular, respectively, non stationarity in mean and seasonal behaviours.

For the construction of models, whatever their class, Box and Jenkins have introduced a methodology for obtaining a linear model that best adjusts to a time series. This methodology consists of three steps: identification of the model, parameter estimation and validation of the model, cf Fiordaliso [1].

The article is structured as follows: Section 2 is devoted to the basic concepts rest of ARMA. Ssection 3 presents the optimal selection criterion. Section 4 discusses the document, the Box and Jenkins approach to retain ARMA(2,1) as our simulation model and we end with the conclusion and some perspectives.

2. Process ARMA(AR+MA)

The ARMA models (also known as Box Jenkins models), are the most common type of time series model. They are mainly based on two principles highlighted by Slutsky and Yule, the autoregressive and moving average principles. Their application to the analysis and prediction of time series was widespread Box and Jenkins in 1970. They showed that this process could be applied to many areas and was easy to implement.

Given a time series X_t, the ARMA model is a tool to understand and attempt to predict possibly future values in this series. The model consists of two parts: an Autoregressive (AR) part and Moving-Average (MA) part. The

model is generally denoted ARMA (p, q), where p is the order of the AR part and q the order of the party with MA $p \geqslant 0$; $q \geqslant 0$.

An autoregressive model and moving-average orders (p, q) (abbreviated ARMA (p, q)) is a discrete time process $(X_t, t \in \mathbb{N})$ satisfying, cf. Jonathan [5]:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{2.1}$$

where the parameters φ_i and θ_i are constants, and the error terms ε_i are independent of the process.

- An autoregressive model AR(p) may be identified as an ARMA$(p, 0)$. In this case the series becomes :

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

  where $\varphi_1, \ldots, \varphi_p$ are the model parameters, c is a constant and ε_t white noise, cf Peter and al [6]. In some literature [5], the constant is often omitted, the process is then said to be centered. The first autoregressive processes were introduced by George Udny Yule. In his paper he uses the first autoregression model to model the time series of the number of sunspots than the Schuster periodogram method, cf. Jonathan and al [5].

- A moving average model MA(q) is an ARMA$(0, q)$. The series is then:

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

  where $\theta_1, \ldots, \theta_q$ are the parameters of the model $\varepsilon_t, \varepsilon_{t-1}, \ldots$ are again the error terms. Eugen Slutzky introduced this model for the first time in 1927 the moving average process in his article.

Note that the error terms of ε_i are generally assumed to be independent and identically distributed (iid) as a normal distribution with zero mean : $\varepsilon_t \sim N(0, \sigma^2)$ where σ^2 denotes the variance.

Condensed manner, using the delay operator, the ARMA may be written in condensed form as, for all $t \in \mathbb{Z}$,

$$\mathcal{A}(z) X_t = \mathcal{B}(z) \varepsilon_t \tag{2.2}$$

where we define for any $z \in \mathbb{C}$, polynomials $\mathcal{A}$ et $\mathcal{B}$ by, cf Jonathan et al [5]: $\mathcal{A}(z) = 1 - \theta_1 z - \ldots - \theta_p z^p$ and $\mathcal{B}(z) = 1 + \varphi_1 z + \ldots + \varphi_q z^q$.

In the case of AR(p), we have $\mathcal{B}(z) = 1$ whereas symmetrically, in the case of MA(q), we have $\mathcal{A}(z) = 1$. Modeling is referred to as "*minimal*" if $\theta_p \neq 0$, $\varphi_q \neq 0$ and if $\mathcal{A}$ and $\mathcal{B}$ have no common root. Without this, it is always possible to find a formulation ARMA(p', q') equivalent with $p' \leqslant p$ and $q' \leqslant q$ generating (X_t).

2.1. Autocorrelation and Correlograms Functions

The autocorrelation function of a process $(X_t, t \in \mathbb{Z})$ with average $E(X_t) = m$, denoted $\rho(k)$ or ρ_k, is defined by $\forall k \in \mathbb{Z}$:

$$\rho(k) = \rho_k = \frac{\gamma(k)}{\gamma(0)}, \tag{2.3}$$

where $\rho(k) \in [-1, 1]$, and $\gamma(k) = \gamma_k$ denotes the autocovariance function,

$$\forall \in \mathbb{Z}, \gamma(k) = \gamma_k = E[(X_t - m)(X_{t-k} - m)].$$

The partial autocorrelation [8] of order k denotes the correlation between X_t and X_{t-k} obtained when the influence of the variables X_{t-i} with $i < k$ was removed.

The graph of an autocovariance function (resp. Autocorrelation) is called a variogram (resp. correlogram). Similarly, we define a partial correlogram as the graph of the partial autocorrelation function (resp. correlogram).

An AR (p) has simple correlogram characterized by a geometric decrease in its terms and partial correlogram characterized by its first p terms different from 0.

	AR(p)	MA(q)
$\rho(h)$	decreases exponentially to 0 with h	0 if $\|h\| > q$ and non-zero if $h = q$
$r(h)$	0 if $h > p$ and non-zero if $h = p$	-
$\rho^i(h)$	0 if $h > p$ and non-zero if $h = p$	decreases exponentially to 0 with h

where $\rho^i(h)$ is the inverse auto-correlation of order h and is defined by $\rho^i_X(h) = \frac{\gamma^i_X(h)}{\gamma^i_X(0)}$ and $r(h)$.

Inverse autocorrelations [8] a MA(q) has the same properties as the autocorrelations of an AR(q).

Proposition 2.1. (cf. Jonathan et al. [5], Brockwell et al. [7] and Stocker [9]) The centered stationary process (X_t) is generated by minimal modeling AR(p) if and only if $\rho(p) \neq 0$ and $\rho(h) = 0$ for any $h > p$.

Proposition 2.2. (cf. Brockwell et al. [7]) The centered stationary process (X_t) is generated by a minimal MA(q) model if and only if $\rho(q) \neq 0$ and $\rho(h) = 0$ for any $|h| > q$.

There are other places in the rest of the paper where you need to make the same correction In practice, an ARMA process is often presumed under the following conditions:

1. the process is stationary in the visual analysis:
   - no trend,
   - no seasonality,
   - constant variance.
2. empirical correlation function is:
   - to decay too slow,
   - without periodic peaks.

Proposition 2.3. (cf. Brockwell et al. [7]) Either the process (X_t) generated by the stationary modeling minimal $ARMA(p, q)$ $\mathcal{A}(L)X_t = \mathcal{B}(L)\varepsilon_t$ where (ε_t) is a white noise variance $\sigma^2 > 0$. So, for $\lambda \in \mathbb{T} = [-\pi, \pi]$, its spectral density is given by $f_X(\lambda) = \frac{\sigma^2|\mathcal{B}(e^{-i\lambda})|^2}{2\pi|\mathcal{A}(e^{-i\lambda})|^2}$.

The stationary character of (X_t) is implicitly related to the fact that the polynomial $\mathcal{A}$ does not vanish on the unit circle, thus guaranteeing the existence of $f_X(\lambda)$ on over all the torus $\mathbb{T}$.

2.2. Estimation of ARMA Model

Parameter estimation of ARMA(p, q) where p and q commands are assumed to be known can be achieved by various methods in the time domain:

- Ordinary Least Squares (model without MA components, $q = 0$). In this case, there are the Yule Walker equations. Replacing theoretical

autocorrelations by their estimators, one can find the MCO estimators of the model parameters by solving the Yule Walker equations.

- Maximum Likelihood approach (Box and Jenkins 1970) [8].

- Exact Maximum Likelihood (Newbold 1974, Harvey and Philips 1979, Harvey 1981) [8].

We will present here briefly the approach of the estimate by maximum likelihood. This maximization is performed by using nonlinear optimization algorithms such as Newton-Raphson or the simplex method that we will not explain it in this chapter. Here it suffices to show how the writing of the likelihood maximization program to estimate the parameters of an ARMA (p, q).

2.3. Stationarity and Causality of the Process

Definition 2.4. We say that the process $(Y_t)_{t\in\mathcal{T}}$ ($\mathcal{T} = \mathbb{N}$ *or* $\mathbb{Z}$) is strictly stationary (or strongly stationary) if the law of $\{Y_{t_1}, \ldots, Y_{t_n}\}$ is the same as the law of $\{Y_{t_{1+\tau}}, \ldots, Y_{t_{n+\tau}}\}$ for all $(t_1, \ldots, t_n)$ with $t_i \in \mathbb{T}$ for $i = 1, \ldots, n$ and for any $\tau \in \mathcal{T}$ with $t_{i+\tau} \in \mathcal{T}$.

Thus, a random process is strictly stationary if all these statistical characteristics, that means all those moments are invariant for any change in the origin of time. But the stationary in strict sense is too restrictive, and this condition is relaxed by defining the stationary of second order.

Definition 2.5. A process $(Y_t)_{t\in\mathcal{T}}$, is called second-order stationary (or weakly stationary) if $(Y_t)_{t\in\mathcal{T}}$, is 2nd order and if the first two moments are time-invariant:

1. $E(Y_t) = m = constant \ \forall \ t \in \mathcal{T}$

2. $Var(Y_t) = \sigma^2 = \gamma(0) < \infty$

3. $Cov(Y_t, Y_{t-h}) = E(Y_t Y_{t-h}) - E(Y_t)E(Y_{t-h}) = \gamma(h) \ \forall \ t \in \mathcal{T}, \forall \ h \in \mathcal{T}$

In short, a process Y_t is called second-order stationary if its mean, its variance and its covariance are independent of time and if variance is non infinite. A such a process is without trend in mean and without trend in variance.

Example 2.6. The best known example of stationary process is white noise process (denoted BB or White Noise). A White Noise is a series of real random variable ε_t, $t \in \mathcal{T}$ such that: $E(\varepsilon_t) = 0 \ \forall \ t \in \mathcal{T}$,

$$\text{and} \ \ \gamma(h) = E(\varepsilon_t \varepsilon_{t-h}) = \begin{cases} \sigma^2 \ if \ h & = & 0 \\ 0 \quad h & \neq & 0 \ if \ not \end{cases}$$

Definition 2.7. An AR(p) process is called causal when there is a series of numbers α_k as $k \in \mathbb{Z}$,

$$\sum_{k \in \mathbb{Z}} \mid \alpha_k \mid < \infty$$

and

$$X_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{tk}.$$

By this definition, we can see any moving average process is causal.

3. Selection Criterion

The autocorrelation functions and partial autocorrelation allow us to determine the order of an autoregressive or moving average model. Let us look for the model from the statistical criterion.

3.1. Candidate Models Search

For simplicity assume that research is done among not what is a sidestep ARMA process, it allows us to put aside seasonal patterns. What is a fine ARMA process is that of the unknown couple (p^*, q^*) such that $(p^*, q^*) < (p_{max}, q_{max})$. In other words, it is assumed that real orders are respectively less than two orders p_{max} and q_{max} that one focuses on. In practice, a problem arises when the orders p_{max} and q_{max} are chosen too small to find the best model. Generally we examine correlograms representing the autocorrelations and partial autocorrelations estimated in order to set these maximum limits. Setting maximum limits p_{max} and q_{max} gives rise to a family of $(p_{max} + 1) \times (q_{max} + 1)$ candidate models, as shown in Table 1.

The search for an optimal model in the sense of a certain criterion will therefore be in this family. If the procedure results in the choice of a model

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

Table 1: Candidate models search

belonging to the last column or the last row then it is prudent to redefine the new values for $p_{\max}$ or $q_{\max}$ higher than those initially chosen.

Let l_i the log-likelihood of the ith pattern, T the size of the working sample and k_i the number of parameters, the selection criterion is written in general as follows:

$$c_i(T,k_i) = \frac{-2l_i}{T} + \frac{k_i g(T)}{T}, \tag{3.1}$$

where $g(T)$ is a positive function of many parameters to estimate and therefore denotes the magnitude of the penalty.

This minimization problem reflects the tradeoff between increasing likelihood and increasing complexity of the model.

3.2. AIC and BIC Criterion for Autoregressive Process

The Akaike Information Criterion (AIC) generates a function that estimates the quality of the fit. Recall that if the number of parameters increases, the variance $\hat{\sigma}_\varepsilon^2$ decreases. In order not to end up with an over-parameterization of the model, we add a factor that will make a compromise between the number of parameters and minimum variance. In the following, we will take a model $A(p)$ and considers $\hat{\sigma}_\varepsilon^2$ using maximum likelihood for many positive values of p . This method could also be used for a model $MA(q)$. The AIC consists of calculating

$$AIC(p) = log(\hat{\sigma}_\varepsilon^2) + 2\frac{p}{T}.$$

Using this criterion, we remark that if p is the obtained parameter of the minimization and p is the parameter of the real model, it has the following property: $P(\hat{p} \geqslant p) \longrightarrow 1$ when $T \longrightarrow +\infty$, [8]. The criterion therefore tends to select a larger number of parameters than the real model, which leads us to a small error term $\hat{\sigma}_\varepsilon^2$. If one wishes to have a better choice of order p, there is the Bayesian

information criterion (BIC) that uses a higher penalty. The BIC selects the parameter p which minimizes the following quantity $BIC(p) = log(\hat{\sigma}_\varepsilon^2) + \frac{p}{T} log(T)$.

3.3. AIC and BIC for ARMA Model

In estimation, it is a little more expensive to deduct the p and q order for a ARMA(p,q) process because to optimize the model it is necessary to minimize a function of two variables. The AIC and BIC criteria for a ARMA(p,q) process are written as:

$$AIC(p,q) = log(\hat{\sigma}_\varepsilon^2) + 2\frac{(p+q)}{T}\ ,\ BIC(p,q) = log(\hat{\sigma}_\varepsilon^2) + \frac{(p+q)}{T} log(T).$$

To minimize these functions, one method is to make two iterative loops on p and q to test all pairs (p,q) p until some limits $< P$ and $q < Q$. Inside these loops, first compute the estimators $\hat{\Phi}$, $\hat{\theta}$ and using for example the least squares or maximum likelihood, $\hat{\sigma}_\varepsilon^2$ we calculate the AIC and BIC criteria for these various levels and there is the minimum of these quantities. It was therefore the values $\hat{p}$ and $\hat{q}$ that minimize the AIC or BIC. Then we calculate efficient estimators of the model parameters ARMA $(\hat{p}, \hat{q})$ using the maximum likelihood method. If several models are competing, we choose the pair (p,q) which minimizes statistics $AIC(p,q) = log(\hat{\sigma}_\varepsilon^2) + 2\frac{(p+q)}{T}$ or $BIC(p,q) = log(\hat{\sigma}_\varepsilon^2) + \frac{(p+q)}{T} log(T)$, cf. Peter et al. [6].

4. Estimation, Validation Testing and Forecasting ARMA

The modeling procedure of Box and Jenkins comprises the following steps, [8]:

1. Stationarity and seasonal adjustment
2. Identification
3. Estimate
4. Validation and Test
5. Forecast

4.1. Stationarity Tests of the iqa Series

To verify the stationarity there are several methods: visual examination in the series, the calculations of the mean and variance of the subsets of the series

and equality tests, visual analysis of the decrease in the function of autocorrelation and unit root tests (Dickey-Fuller, Phillips-Perron Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test ...). We use the unit root test in the current research.

	Statistic	p-value
Phillips-Perron Unit Root Test	-11.0836	0.01
Augmented Dickey-Fuller Test	-4.8911	0.01
KPSS	0.5337	0.03408

We observe that for the Dickey-Fuller Augmented, the Phillips-Perron and just like the KPSS test, the p-value less than 0.05 so the variable log(iqa) is stationary.

4.2. Model Identification of iqa: ARMA(2,1)

In some literature the analysis of univariate time series as that of our research on the Index of air quality in the city of Dakar in the Box and Jenkins boils down to three steps: identification, estimation, validation. The initial phase of identification is often difficult in practice. It is not clear at all to often find the right model to suit the timeline series considered. It is in this phase that takes place the great work operations research series good adjustment model.

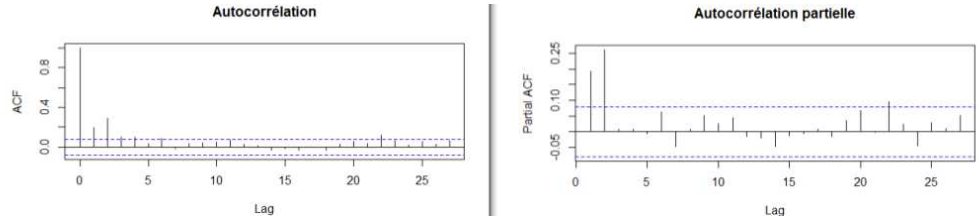


Figure 1: ACF and PACF for series log(iqa).

By observing Figure 1, there is seen that the autocorrelation and partial autocorrelation decay slowly and take time before they significantly nil.

For this reason we rule out possible to retain an AR or MA model. Estimates the AIC and BIC criteria for an $ARMA(p,q)$ $p = 1,\ldots,5$ and $q = 1,\ldots,5$. Looking couple (p,q) optimal (minimum) for our model. Tables 2 and 3 give us those values.

Operational research by the AIC seems to direct us towards the operation (the model) ARMA(4,4) followed by ARMA(5,4) and ARMA (2,1). But

p \ q	1	2	3	4	5
1	805.07	790.47	779.93	774.89	770.37
2	**762.64** *third candidate optimum(2,1)*	764.57	766.4	768.4	770.22
3	764.56	766.64	768.5	769.84	772.33
4	766.39	768.2	764.29	**762.25** *first candidate optimum(4,4)*	764.04
5	768.38	770.36	772.42	**762.52** *second candidate optimum(4,5)*	764.45

Table 2: Criterion AIC series iqa for $(p, q) \in [1, ..., 5]^2$

the third optimal choice (minimum) which would be the lowest AIC operation ARMA(2,1). We choose to estimate three parameters, eight or even nine to estimate. Selecting a model of nine or eight parameters maximizes the likelihood of the model. This model could give us a better fit in our sample, but could be completely wrong to make predictions. So we choose to estimate three parameters instead of eight or nine. To check if this choice fits well to our data, we also have the opportunity to compare the two variance estimators residues.

p \ q	1	2	3	4	5
1	825.0663	815.4707	809.9266	809.8786	810.3665
2	**787.6319** *optimum(2,1)*	800.9503	826.8575	810.1573	822.2392
3	794.5862	801.3645	808.2271	814.8847	822.1193
4	801.3752	808.201	811.2327	827.0412	819.5427
5	808.3771	814.8691	815.7096	819.4184	825.8988

Table 3: BIC Criterion for series $log(iqa)$ with $(p, q) \in [1, ..., 5]^2$

The BIC of the table 3 shows that the pair (2,1) gives us the minimum. We retain finally fit for our operation ARMA(2,1).

4.3. Estimation and Simulation of IQA

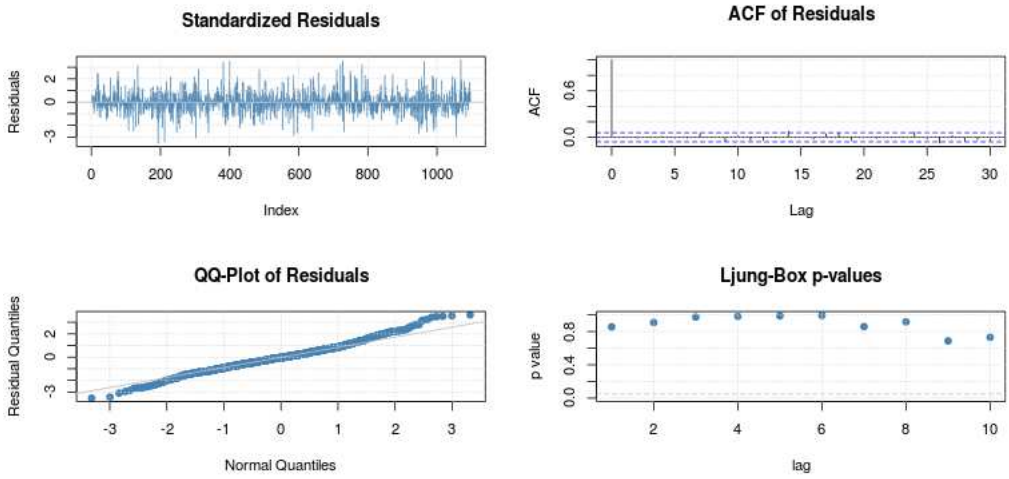
In this section we estimate and simulation of $log(IQA)$ by an ARMA(2,1) under the R software. With Farma library in R, the simulation gives us:
Title: ARIMA Modelling
Call: armaFit(formula = arma(2, 1), data = log(iqa))
Moments: Skewness=0.2539 Kurtosis= 0.9756.
Skewness measures the asymmetry of the distribution process and kurtosis measures the excess kurtosis: it is equal to $3 - \widehat{KU}$. The estimate of kurtosis and

skewness are respectively given by:

$$\widehat{KU} = \frac{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - \bar{u})^4}{(\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - \bar{u})^2)^2} \quad \text{and} \quad \widehat{SK} = \frac{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - \bar{u})^3}{(\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - \bar{u})^2)^{3/2}}.$$

The R software provides us the coefficients of the estimation in the following table:

```
           Estimate  Std. Error  t value Pr(>|t|)
ar1         1.59923     0.03428    46.65   <2e-16 ***
ar2        -0.60384     0.03298   -18.31   <2e-16 ***
ma1        -0.93293     0.01862   -50.11   <2e-16 ***
intercept   4.04585     0.14121    28.65   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
sigma^2 estimated as: 0.9162
log likelihood:       -376.32
AIC Criterion:        762.64
```



We note the model estimation seems good, because the value of our estimator is very close to the true value.

QQ-PLOT test of normality is a graphical method: The point cloud is formed by (quantile N(0.1), reduced empirical quantile of $\hat{u}_t$), assuming H0 the cloud is on the straight line y = x). We confirm that there is a normality.

4.3.1. Adequacies Tests

With R software, Box.test gives us the following result:
Box-Pierce test
Data: predict$residuals
X-squared = 6383.219, df = 10, p-value = 0,6742
The p-value is large ($0.6742 > 0.05$) so the residuals are not correlated.

And the Kolmogorov-Smirnov lillie test with the function () provides us:
Lilliefors (Kolmogorov-Smirnov) normality test
data: predict$residuals
D = 0.1125, p-value < 0.3221
We can therefore say the residues data follow a Gaussian distribution.

4.4. Forecast of iqa by ARMA(2,1)

In this section we predict log(IQA) for a 31-day period in January 2013, we compare these results with real measurements taken at stations during the same period.

	1erJan2013	2jan2013	3jan2013	4jan2013	5jan2013	6jan2013	7jan2013	8jan2013
logiqa Prdit	4.7178	4.5378	4.4269	4.3582	4.3152	4.2880	4.2705	4.2589
Residuals	0.3409	0.4096	0.4388	0.4535	0.4620	0.4676	0.4716	0.4748

9jan2013	10jan2013	11jan2013	12jan2013	13jan2013	14jan2013	15jan2013	16jan2013
4.2509	4.2451	4.2407	4.2372	4.2341	4.2314	4.2290	4.2266
0.4776	0.4800	0.4823	0.4844	0.4864	0.4883	0.4901	0.4919

17jan2013	18jan2013	19jan2013	20jan2013	21jan2013	22jan2013	23jan2013	24jan2013
4.2244	4.2222	4.2201	4.2180	4.2159	4.2139	4.2119	4.2099
0.4936	0.4953	0.4970	0.4986	0.5001	0.5016	0.5031	0.5045

25jan2013	26jan2013	27jan2013	28jan2013	29jan2013	30jan2013	31jan2013
4.2080	4.2061	4.2042	4.2023	4.2004	4.1986	4.1968
0.5059	0.5073	0.5086	0.5099	0.5111	0.5124	0.5135

4.5. Measuring Quality Prediction

In our initial sample, $(X_1, \ldots, X_T)$ only consider, $T_1 = [(1-\varepsilon)T]$ observations with $\varepsilon > 0$. The $L = T - [(1-\varepsilon)T]$ either are the $L = T - T_1$ will be predict by the model. We can then consider several criteria:

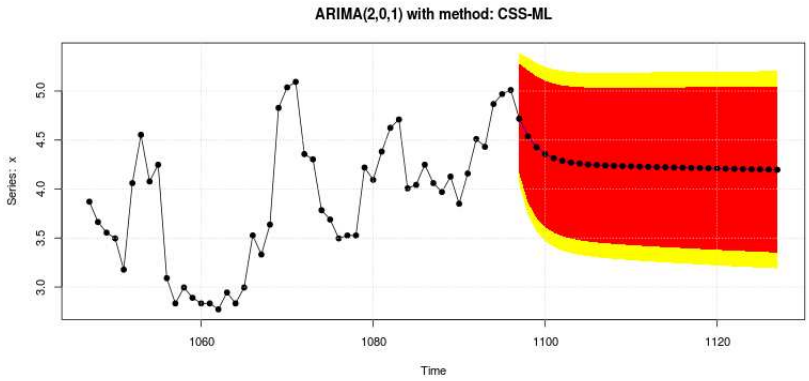


Figure 2: Forecasts and confidence interval to 90% (red) and 95%(yellow).

- Mean Absolute Percentage Error

$$MAPE = \frac{1}{L}\sum_{i=1}^{L}\left|\frac{X_{T_1+i}-\hat{X}_{T_1+i}/T_1}{X_{T+i}}\right|$$

- Mean Square Error

$$MSE = \left(\sum_{i=1}^{L}\frac{(X_{T_1+i}-\hat{X}_{T_1+i}/T_1)^2}{L}\right)^{1/2}$$

Here L is the number of fitted points. We get under the R software, MSE = 4.687186 and MAPE = 4.624215. In conclusion we can say ARMA(2,1) is a good model for our fit and taking into account the information criterion corrected above, the pair (2,1) is a good choice of optimal pair (p, q) that minimizes the criterion. Indeed MSE gives greater weight to larger deviations (which could result from outliers) and MAPE gives less overall weight to a large deviation if the time series value is large. MSE give us the averages of the squared deviations and MAPE averages the absolute percent errors. Although the concept of MAPE sounds very simple and convincing, it has major drawbacks in practical application [10].

It cannot be used if there are zero values (which sometimes happens for example in demand data) because there would be a division by zero. For forecasts

which are too low the percentage error cannot exceed 100%, but for forecasts which are too high there is no upper limit to the percentage error. When MAPE is used to compare the accuracy of prediction methods it is biased in that it will systematically select a method whose forecasts are too low. This little-known but serious issue can be overcome by using an accuracy measure based on the ratio of the predicted to actual value (called the Accuracy Ratio), this approach leads to superior statistical properties and leads to predictions which can be interpreted in terms of the geometric mean [10]. Mean square error (MSE) can also be utilized in the same fashion. Squaring the forecast errors eliminates the possibility of offsetting negative numbers, since none of the results can be negative.

5. Conclusion and Perspectives

This research leads us to choose the ARMA(2,1) to model and predict the index of the air quality in Dakar. The model can be used to make short-term predictions of IQA in Dakar by the air quality management center (CGQA). It provides a tool that the quality monitoring center air Dakar could use to benefit public health. An interesting avenue of research work would be to couple an urban sprawl model with a simulation model of emissions of pollutants from road transport. Indeed, there is evidence that as the city grows in extent, the more origin-destination trips made daily by motorists is growing, resulting in increased traffic. Such a model could be used to assess the impact of new areas of development plans and cities around Dakar on transport and the impact of that traffic on people and plants.

References

[1] A. Fiordaliso, *Systèmes Flous et Prévision de Séries Temporelles*, Hermès Science Publication, Paris (1999).

[2] Y. Aragon, Séies temporelles avec R. Méthodes et cas, Springer, Paris (2011).

[3] A. Schuster, *On the Periodicity of Sun-Spots*, Royal Society of London, London (1906).

[4] G. Colletaz, *Les Critères de sélection*, Notes de Cours Master 1, ESA, Novembre (2007).

[5] Jonathan D. Cryer and Kung-Sik Chan, *Time Series Analysis with Applications in R*, 2nd Ed., Springer (2008), 191-245.

[6] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods*, Springer (2009).

[7] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, 2nd Ed., Springer-Verlag, New York (1991).

[8] P.J. Brockwell and R.A. Davis, *Introduction to Time Series and Forecasting*, 2nd Ed., Springer Texts in Statistics, New York (2002), 83-421.

[9] T. Stocker, On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors, *Statist. Papers*, **48** (2007), 81-93.

[10] C. Tofallis, A better measure of relative prediction accuracy for model selection and model estimation, *J. of the Operational Research Society*, **66**, No 8 (2015), 1352-1362.