MODELING OF THE METEOROLOGY DROUGHT SOUTH OF MADAGASCAR, APPROACH BY ARIMA PROCESS

Jean Eugène RANDRIANANTENAINA [1*], Adolphe Andriamanga RATIARISON [2], Pierre Ruphin FATIANY[3], Jacques Chrysologue RATSIMAVO[4]

[1] *Master of conferences, Department of physics, Faculty of Sciences, Toliara, Madagascar*
[2] *Emeritus Professor, Department of physics, DYACO laboratory, Antananarivo, Madagascar*
[3] *Professor, Department of physics, Faculty of Sciences, Toliara, Madagascar*
[4] *Doctor of physics, Department of physics, Faculty of Sciences, Toliara, Madagascar*

ABSTRACT

The ARIMA (Autoregressive Integrated Moving Average) was applied to the series of SPI -1. The results of the evaluation of the model showed good agreement between observations and forecasts , as was also confirmed by the values of some performance indices and the results seem to be better for better SPI Serial and this may be due to increase the filter length reduces the effective noise .

Keywords: *Drought, Standardized Precipitation Index, autocorrelation function, partial autocorrelation function, ARIMA*

---

1. INTRODUCTION

Climatic models (*stochastic or deterministic)* are used for various purposes, from the study of time dynamics and climatic systems for the projections of the future climate. The goal of this section is to propose a statistical model based on sets of SPI-1 and to study the evolution of the climate in the south of Madagascar. The index SPI was used as an indicator of drought for forecasting because of its advantages over other drought indices. The capacity of the ARIMA Model in predicting drought has been studied using the Box and Jenkins method.

2. METHODOLOGY

2.1. ARIMA time series modeling

ARIMA models allow three types of time processes to be combined: autoregressive processes *(AR),* moving average processes *(MA)* and integrated processes *(I).* In the general case, an ARIMA model *(p, d, q)* is a combination of these three types of process. The *p, d, q and* designing respectively the order of the autoregressive process, the order of integration and the order of the moving average. It is by the Box & Jenkins method to build a model restoring as best as possible the behavior of a time series following three stages: identification, estimation and diagnosis.

2.1.1. Autoregressive processes

A process (X_t) is said to be autoregressive of order p, *AR (p),* if the present observation $X_t$is generated by a weighted average of the past observations up to the p- *th*period in the following form:

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

That we can also write in a condensed form:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t \qquad (1)$$

α_iAre real coefficients fixed with the condition?$\alpha_p \neq 0$

ε_tIs a white noise ie they $\varepsilon_t$are *iid*according to a law $\boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2)$

2.1.2. Moving average processes

In a (X_t) moving average process of order *q*, each observationX_t is generated by a weighted average of hazards up to the *q- th*period in the past.

$$\forall t \in \mathbb{Z}: \; X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \;, où\left(\theta_1, \dots, \theta_q\right) \in \mathbb{R}^q et \theta_q \neq 0$$

Generally, we use the following notation:

$$X_t = \Theta(L)\varepsilon_t, \;\; où \Theta(L) = I + \sum_{i=1}^{q} \theta_i L^i$$

An MA process is always stationary.

2.1.3. ARMA processes (p, q)

ARMA models are representative of processes generated by a combination of past values and past errors.

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

We can also write the ARMA model *(p, q)* in the form:

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

Where *L* is the shift operator and $\varepsilon_t$is the white noise process.

2.1.4. ARIMA processes (p, *d, q)*

A process $X_t$is called ARIMA *(p, d, and q), where p, d and q* are positive or zero if the process $(1-L)^d X_t$is a stationary ARMA process (p, *q)*.ARIMA processes are useful for processes that have positive and slowly decreasing correlations because this property of autocorrelations can be a sign of a trend in the series.

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

The ARIMA process (0, 1, and 0) is called Random Walk Model. It is often used to analyze the efficiency of financial markets.

3. RESULTS

3.1. Drought analysis by SPI modeling

The SPI was developed with the aim of defining and monitoring droughts. The global climate change in recent years is likely to increase the frequency of droughts. Although much of the time we live is brief and short-lived, drought is a more gradual phenomenon, affecting an area slowly and tightening its grip over time. In severe cases, drought can last for many years, and can have devastating effects on agriculture and water supply. It is very difficult to determine when a drought begins or ends. A drought can be short, lasting a few months, or it can persist for years before weather conditions return to normal. Drought forecasts play an important role in mitigating the effects of drought on water resource systems. Traditionally, statistical models have been used for hydrological forecasting of droughts as a function of time in the series methods. One of the basic shortcomings in mitigating the effects of drought is the inability to predict drought conditions reasonably well in advance by either a few months or a few seasons.

3.2. Development of an ARIMA model for the SPI_1 time series

The frequency of drought episodes was calculated using the Standard Precipitation Index (SPI). The following figure shows a sample to calculate for the time series of SPI_1.

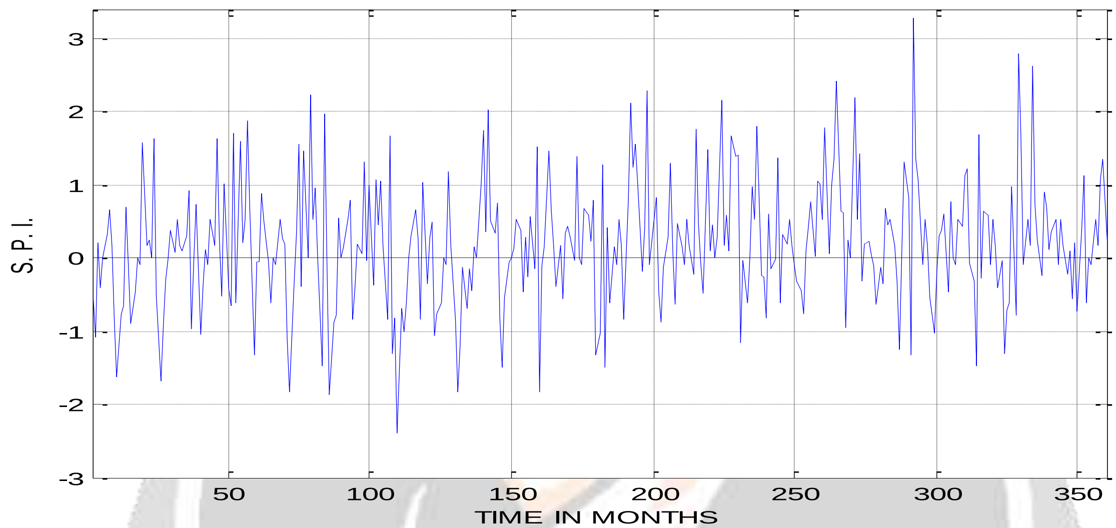

Fig-1:*SPI-1 time series*

The development of the time series model consists of three phases: identification, estimation, and diagnostic verification *(Box and Jenkins, 1970).* The identification stage involves transforming the data *(if necessary)* to improve the normality, the stationary of the time series and the determination of the general form of the model to be estimated. During the estimation, the parameters of the model are calculated. Finally, diagnostic tests of the model are made to reveal the shortcomings of possible model and help in choosing the best model.

Model Identification: The autocorrelation and the correlation part are measures of the association between current and past sets of values; they indicate the values of past series most useful for forecasting future values. The following figures estimate the autocorrelation function (FAC) and the partial autocorrelation function (FACP) for the time series SPI_1.

The function of autocorrelation *(Figure 2)* is sinking in as sinusoidal with peaks important the first two delays. The first two values are important to the partial autocorrelation function *(Figure 3),* which indicates that the process is modeled as ARIMA.
The ARIMA model has been identified by examining the graphs FAC and FACP Series SPI_1. This indicates that an ARIMA model with $p = 1 - 2$ and $q = 0$ is possible.

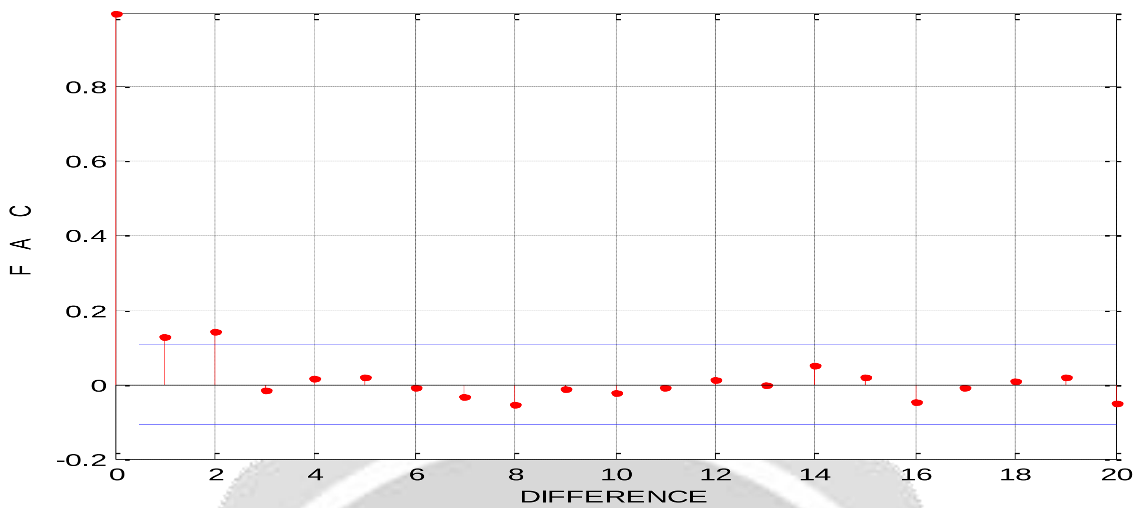

Fig- 2:*Correlogram of SPI_1*

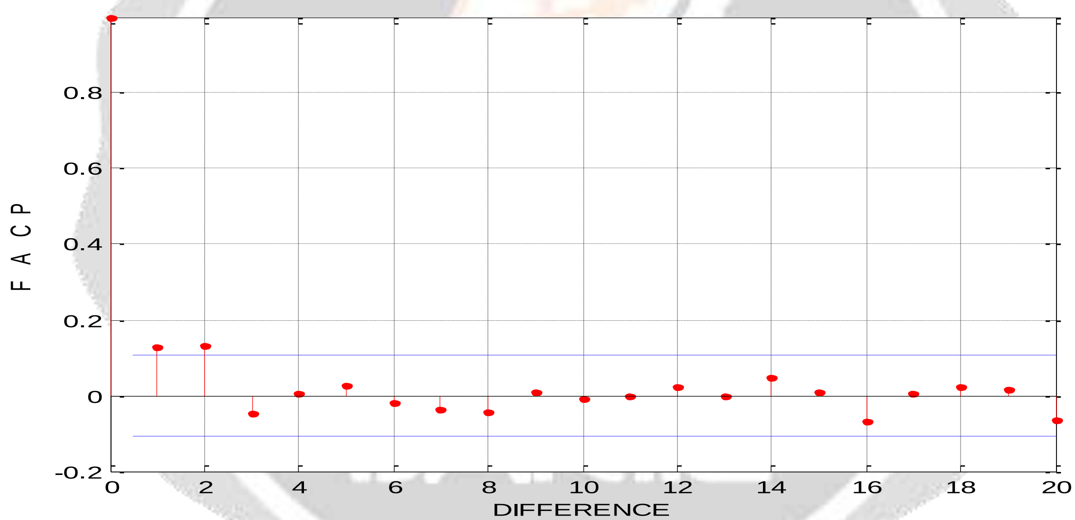

Fig-3:*Correlogram partial SPI_1*

Specification and Estimation of the parameters of an ARIMA(2,0,0) model: We specify then estimate the ARIMA(2,0,0) model for the SPI_1 series. This model has a degree of non-seasonal differentiation and two delays of the AR model. By default, the distribution of innovation is Gaussian with a constant variance.

Table-1: *Evolution of* ARIMA(2,0,0)*parameters after statistical test*

Setting	Value	Standard error	t-test
Constant	0.13038	0.0462274	2.82042
AR {1}	0.122446	0.0510369	2.39916
AR {2}	0.144064	0.0526619	2.73563
Variance	0.730902	0.0428182	17.0699

The two AR coefficients *(AR {1} and AR {2})* are significant at the 0.05 threshold level.

Checking diagnosis : As mentioned previously, a model has been selected, namely ARIMA(2,0,0). The model has been identified and the parameters were estimated, then the verification of the model concerns the control model residuals to see if they contain any systematic trend can be still removed to improve the model chosen. For a good forecast model, the residuals after adjusting the model should be white noise. This is done by examining the autocorrelation and the autocorrelation partial of residues of various kinds. For this purpose, the different correlations up to 20 offsets have been calculated. Also the histogram and the normal probability curve of the residuals were established to check whether the residue came from distribution or not normal.

The test of Ljung - Box, which is commonly used in modeling ARIMA, was applied to tailings AR models equipped. The Ljung -Box test is a type of statistical test of whether one of a group of autocorrelations in a time series is different from zero. Instead of testing randomness at each separate offset, it tests for " global" randomness based on a number of delays .

Checking the Quality of the fit of the ARIMA (2,0,0)model : We deduce the residuals from the fitted model and verify that the residuals are normally distributed and not correlated.

The function of autocorrelation residual *(RACF)* and the function of autocorrelation partial residual *(RPACF)* must be calculated to determine if the residuals are of white noise. If part of RACF or part of RPACF is significantly different from zero, this may indicate that the present model is insufficient. The ACF and PACF of the ARIMA(2,0,0) model residues are presented in Figures *4 and 5* respectively. As shown in Figures *4 and 5,* most of the RACF and RPACF values are within the confidence limits , except that very few individual correlations appear large with respect to the confidence limits , which should cause the 20 delays . The figures do not indicate a significant correlation between the residues. The residue histogram for SPI_1 is shown in *Figure 7.* The histogram shows that the residues are normally distributed. This means that the residue will be white noise. The cumulative distribution graph for residual data normally appears as a straight line when plotted on normal probability paper, as shown in *Figure 6.* The figure shows that the normal probability figure of the residuals appears fairly linear so the assumptions of normality of the residuals hold.

All results of the test of Ljung - Box retain the hypothesis zero that a series of residues does not exhibit autocorrelation (that is $h = 0$ to say, and$p_value = 0,9872$).

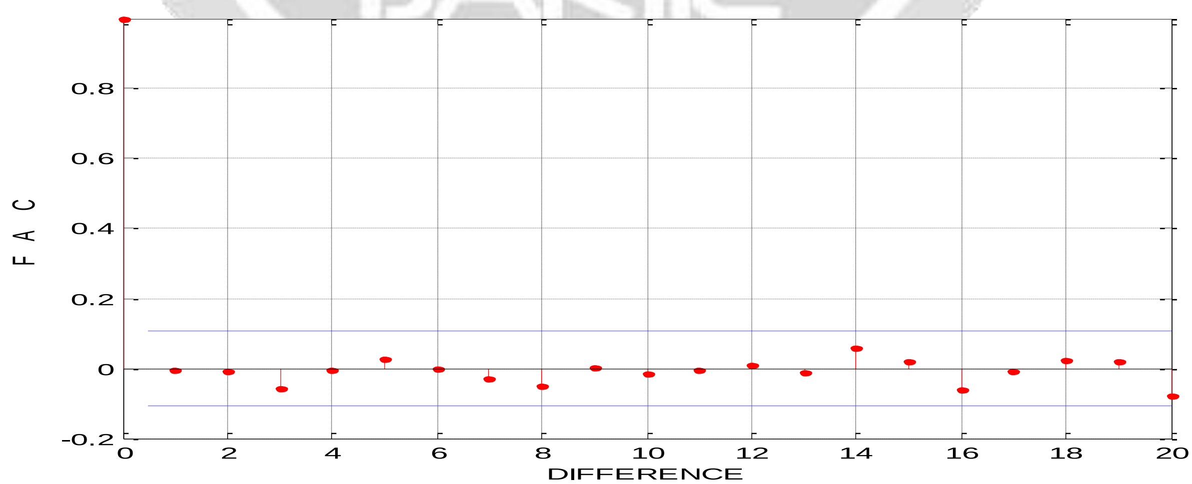

Fig-4:*Correlogram residues SPI_1*

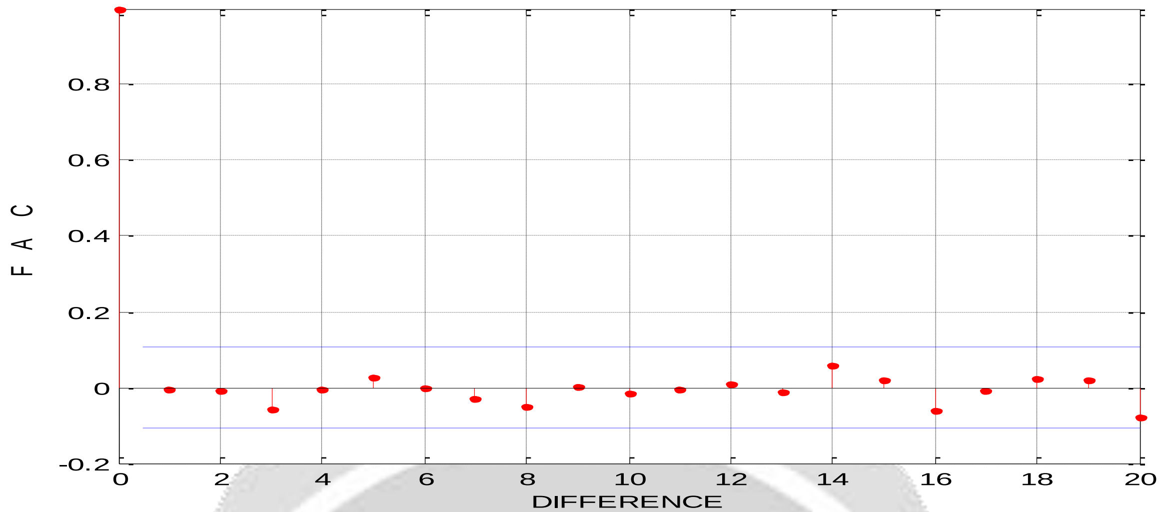

Fig-5: *Partial correlogram residues of SPI_1*

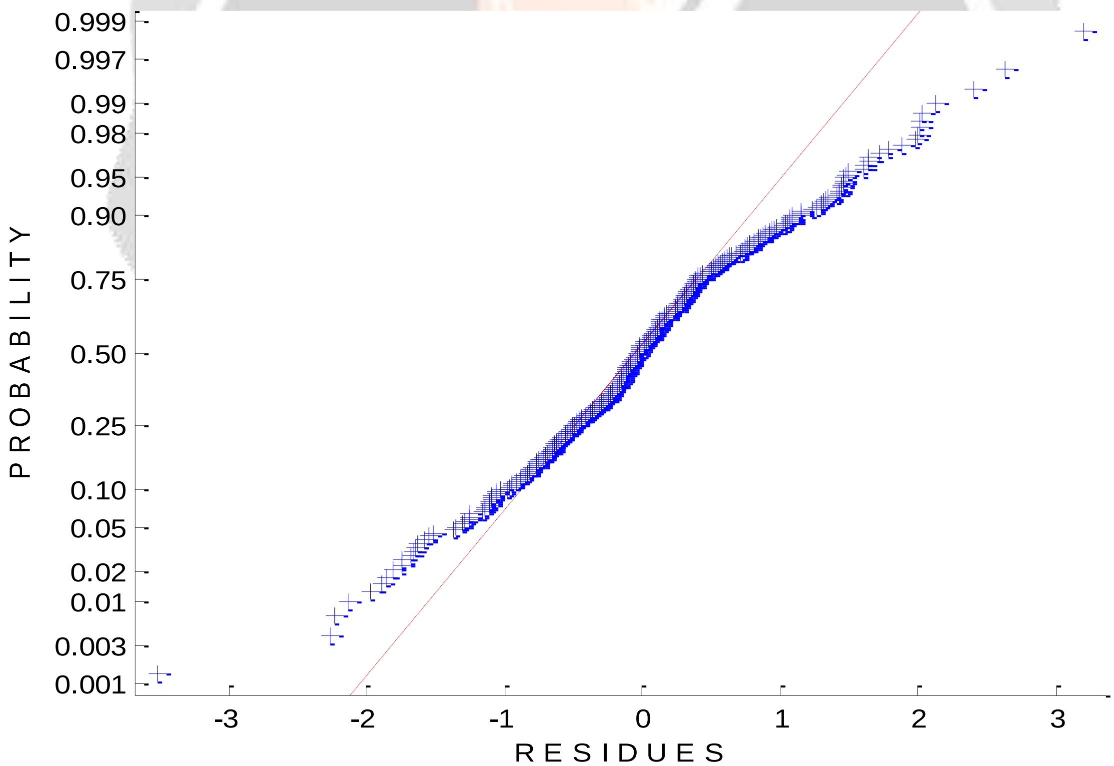

Fig-6: *Normal probability of residues for ARIMA (2 , 0.0)*

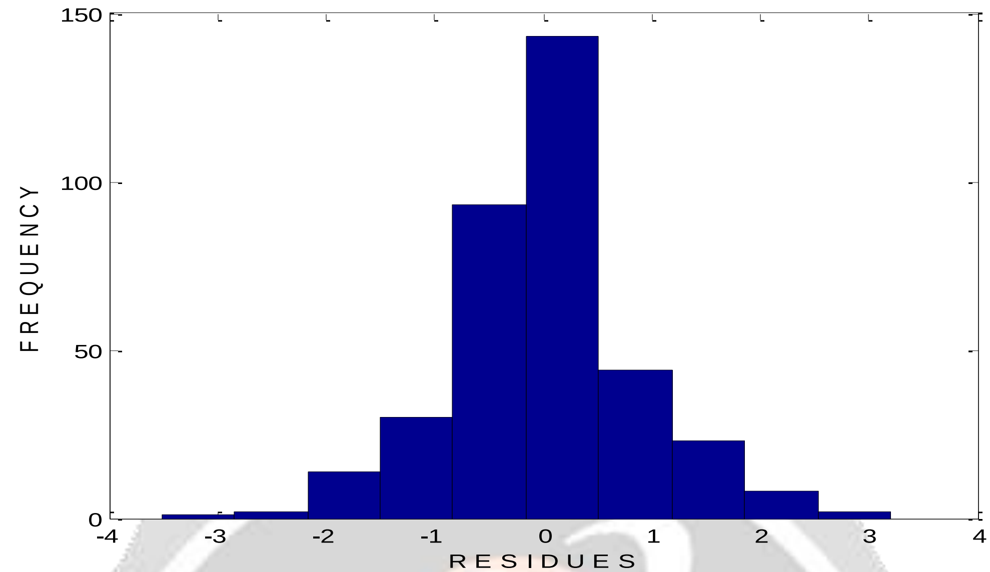

Fig-7: *Residue histogram for ARIMA (2, 0.0)*

Estimation of an ARIMA (2,0,0) model
The model to be estimated is :

$$y_t = 0{,}13038 + 0{,}122446 y_{t-1} + 0{,}144064 y_{t-2} + \varepsilon_t,$$

Where $\varepsilon_t$is normally distributed with a standard deviation of.
The signs of the estimated AR coefficients correspond to the AR coefficients on the right equation of the model. In delay operator in polynomial notation, the adjusted model is:

$$(1 - 0{,}122446\mathrm{L} - 0{,}144064\mathrm{L}^2)(1 - \mathrm{L})\mathrm{y}_t = \varepsilon_t,$$

With the sign of inverse AR coefficients.

Estimated droughts of the selected model:ARIMA models are mainly developed to predict the corresponding variable. There are two groups of forecasts, namely the sample period forecast and the sample forecast. The first group is used to develop confidence in the model and the second to generate real forecasts for use in planning. The ARIMA model can be used to produce two groups of forecasts.

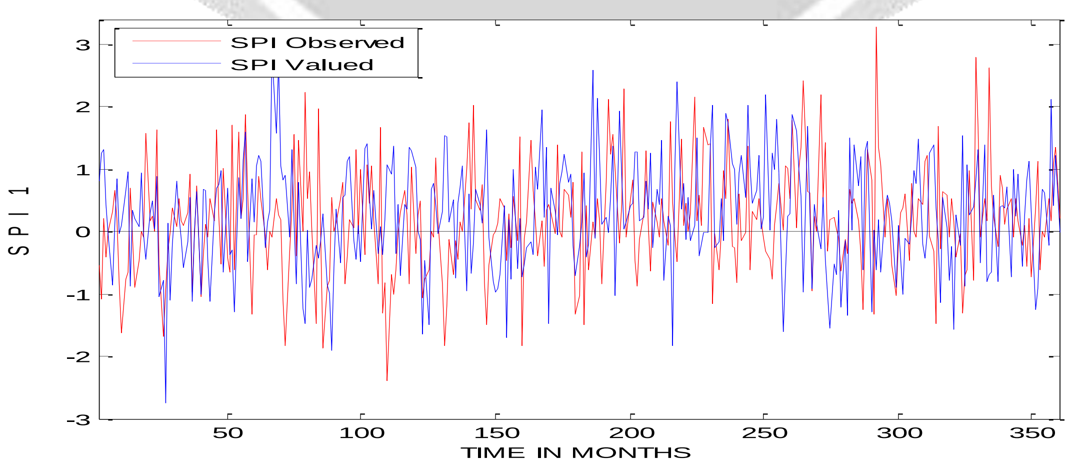

Fig-8: *Comparison of SPI calculated with SPI estimated for ARIMA (2 , 0.0)*

The estimation was made by using the best models from historical data. Results of the ARIMA (2, 0, and 0) model estimation are shown in *Figures 8*, and *Figure 9*. *Figure* 9 is a zoom window for the 336 months at the end of the SPI time series, and this zoom window was taken from *Figure 8.*

Table-2: *Statistical property for ARIMA* (2,0,0)

Model	Average SPI observed	Estimated average SPI	SPI standard deviation observed	Estimated SPI standard deviation	RMSE
ARIMA	0,180	0.0028	0.872	0.856	0.249

It is clearly observed that the predicted values of the SPI follow the values calculated in close collaboration. To evaluate the model, the basic statistical properties were compared between the observed and estimated data. The results, as shown in *Figure 9*, show that the predicted values preserve the statistical properties underlying the observed series.

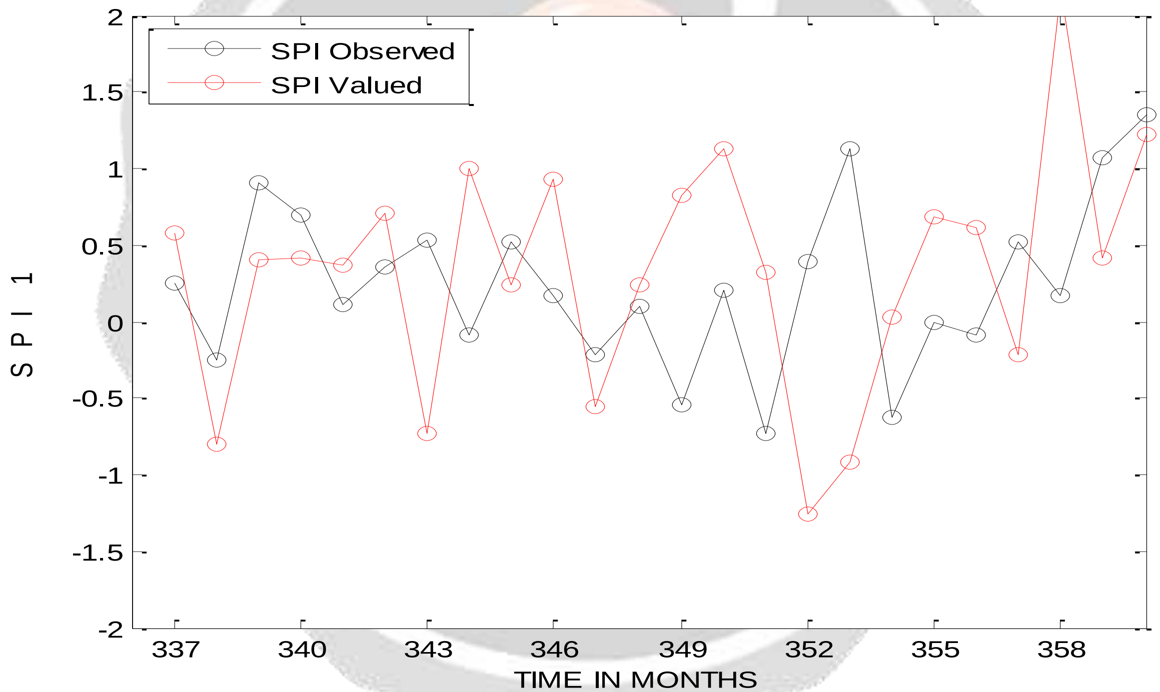

Fig-9: *Comparison of SPI calculated with SPI estimated for ARIMA (2, 0.0) (From 336 months to the end of the time series)*

4. CONCLUSION

In this study, the SPI index was used as an indicator of drought and for predicting drought because of its many advantages over other indices. This study investigated the ability of ARIMA models to predict droughts using the methods of Box and Jenkins. The validation of the forecast models was carried out by comparing the SPI values calculated on the observed precipitation and the corresponding forecasts. The results showed a fairly good agreement between the observations and the forecasts, as was also confirmed by the values of certain performance indices.

5. ACKNOWLEDGEMENT

Our deep gratitude to RATIARISON AdolpheAndriamanga, Emeritus Professor at the Faculty of Sciences ofthe University of Antananarivo, Madagascar and his team, for allowing us to conduct our investigations in theirDYACO laboratory.

6. REFERENCES

[1] Baillie , R., and T. Bollerslev . "Prediction in Dynamic Models with Time-Dependent Conditional Variances." Journal of Econometrics. Flight. 52, 1992, pp. 91-113.
[2] Bollerslev , T. "Generalized Autoregressive Conditional Heteroskedasti city ." Journal of Econometrics. Flight. 31, 1996, pp. 307-327.
[3] Bollerslev , T. "A Conditionally Heteroskedastic Time Series Model for Speculative P rices and Rates of Return." The Review Economics and Statistics. Flight. 69, 1987, pp. 542-547.
[4] Box, GEP, GM J enkins, and GC Reinsel . Time Series Analysis: Forecasting and Control 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1994.
[5] Enders, W. Applied Econometric Time Series. Hoboken, NJ: John Wiley & Sons, 1995.
[6] Engle, RF "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdo m Inflation." Econometrica . Flight. 50, 1982, pp. 987-1007.
[7] Hamilton, JD Time Series Analysis. Princeton, NJ: Princeton University Press, 1994.