UNIVERSITY OF ANTANANARIVO
FACULTY OF THE SCIENCES
DEPARTMENT OF PHYSICS
LABORATORY OF RHEOLOGY OF THE SUSPENSIONS

“FORECASTING OF THE RAINFALL AND THE DISCHARGE OF THE NAMORONA RIVER IN VOHIPARARA”

"FORECASTING OF THE RAINFALL AND THE DISCHARGE OF THE NAMORONA RIVER IN VOHIPARARA

A.T.RABENJA, Adolphe RATIARISON, Olivier RAKOTOVAZAHA (L.P.S) Physics Department, University of Madagascar. E-mail:andritefy2@yahoo.fr

Introduction:

The time series prediction corresponds regularly to the analysis of observations occasional in the time. In this article, the monthly time step has been chosen for discharge and rainfall data.

The study of the forecasting done early uses the classic statistical methods, to know the forecasting by adjustment of law of distribution of the extreme values, and the frequencies study based on the Fourrier analysis.

In this article, the combination of the self regressive methods and moving average has been used in order to forecast the evolution of the rainfall and the discharge. The first, ARMA(p,q) model, the second, the ARIMA(p,d,q) model integrates the data, and the ARIMA(p,d,q)(P,D,Q) process that takes in account the seasonal variations.

The purpose of this paper is to know which of these three models is the more adapted to the forecasting of the rainfall and discharge data.

Methodology:

Study area:

Our site of study is the Vohiparara station. It is find in 21°14' Latitude South - 47°23' Longitude East. This station controls a drainage basin of 445km2. The middle altitude of the basin is 1250m

A scale had been implanted in 1929 for the study of discharges to hydroelectric use. The ORSTOM scale has been installed in November 1951, downstream the first, to 400m upstream of the first falls. Remove in 1960, on the occasion of the road repairing, it has been reinstalled in October 1960, then in January 1970 after the passage of the JANE cyclone. The period of study spreads of 1952 à1989.

Soil is constituted of a pedestal “migmato-gneissique”, sometimes covered of clay “latéritique”,

On the West part of the drainage basin, there are some savannas, prairies and of the replanted forests, and also of numerous rice fields.

In the part is, one finds some forests secondary, with, in part, of the primary forest shreds (Joël Danloux, Luc Ferry, Pierre Chaperon, Paris 1993).

Data:

The rainfall data in millimeter, and of discharge in cubic meter per second have been gotten in the meteorological station, at the hydrological service of Ampandrianomby in Antananarivo. The data spread on about thirty years.

ARMA(p,q), ARIMA(p,d,q), and SARIMA(p,d,q)(P,D,Q) models:

The ARMA model is a mixture of the self regressive models and moving average.

A process (X_t) is an ARMA(p,q) process if it is stationary, that means their white noises ε_t and ε_{t-k} are independent, for all k, for all t, as:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \text{ , for all t.}$$

These processes are stationary under some conditions, and can be writing as:

$$\Phi(L) X_t = \theta(L) \varepsilon_t$$

$$\text{Where } \Phi(L) = II - \Phi_1 L - \ldots - \Phi_p L^p$$

$$\text{And } \theta(L) = II + \theta_1 L + \ldots + \theta_p L^p$$

II is the matrix identity, L represents the delay operator, to the sense where $LX_t = X_{t-1}$, and with the convention $L^p = LoL^{p-1}$, either: $L_p X_t = X_{t-p}$, the serie (Yt), as $Y_t = L_p X_t$ is then the serie (X_t) retarded of p periods.

In the same way, a non stationary process is integrated of order 1, differentiated until it becomes a stationary process; (X_t) non stationary will be said integrated of order 1 if the process (Y_t) defines by: $Y_t = X_t - X_{t-1} = (1-L)X_t$ is stationary.

The process (Xt) can be written like this:

$$\Pi(L)X_t = \Phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t$$

Will be the ARIMA(p,d,q process), ε_t is a white noise.

For the real data, d will be noted d = 1, 2 or 3 (to the maximum). It means that (Yt) defines like difference of d order of the process (Xt). $\Pi(L)$ is the product of delay operators L.

SARIMA models it is a model ARMA integrated and taking into account the seasonal part of the data (Arthur Charpentier,2000).

<u>Estimate of the models parameters:</u>

For the model of p and q order, i for 1 to p and j for 1 to q, the parameters $[\Phi_i]$ and $[\theta_j]$ will be determine by the maximum verisimilitude method. The probability of apparition of the elements is raised if the verisimilitudes function $V(\Phi,\theta,\sigma)$ is maximum, that means:

$$V(\Phi,\theta,\sigma) = P(\Phi,\theta,\sigma)$$

The variables valuers Φ and θ that would return the probable forecasting values are φ and ϑ as : $p(X_t \mid \varphi,\vartheta) \succ p(X_t \mid \phi,\theta)$ What comes back to maximize its logarithm, either:

$$\frac{\partial \ln(p(X \mid \phi,\theta,\sigma))}{\partial \phi}\Big|_{\phi=\varphi} = 0, \frac{\partial \ln(p(X \mid \phi,\theta,\sigma))}{\partial \theta}\Big|_{\theta=\vartheta} = 0, \text{ and } \frac{\partial (\ln(p(X \mid \phi,\theta,\sigma))}{\partial \sigma}\Big|_{\sigma=\sigma^*}$$

The verisimilitude valuers X_t^* and ε_t^* are the solutions of these equations, the initial values X_t^* and ε_t^*, as well as the previous values of X_t and. ε_t are supposed known.

Then, ln(V) is maximal if $\sum_{t=1}^{T} \left\{ \varepsilon_t \left(\frac{\phi,\theta}{\varepsilon_t^*, X_t^*, X_t} \right) \right\}$ is absolutely minimal.

All these models of forecastings, to know ARMA(p,q),ARIMA(p,d,q) and SARIMA(p,d,q)(P,D,Q), are based on the DURBIN-LEVINSON algorithm:

$$\hat{X}_{t+h} = \sum_{k=1}^{T} \alpha_k X_{t+1-k}$$

The α_k parameters is choose in order to minimize the prediction error.(Arthur Charpentier,2000).

Results:

Results of the discharges forecasting:

ARMA(2,12) forecasting of the monthly discharge of 1952 to 1979:

The monthly data of discharge spread from January 1952 to December 1979. The forecasting with the ARMA(2,12) method had begun the January 1980 and have been finished in December 1980.

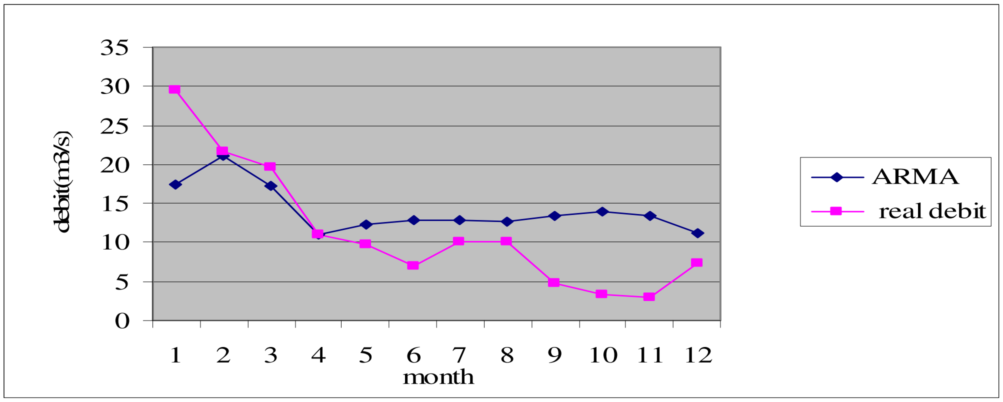

Figure1: Comparison of the ARMA(2,12) forecasting and the real value of the monthly discharge.

For the ARMA forecasting of the monthly discharge of the year 1980, the months of January, June, September, October, presents a deviation.

ARIMA(2,1,12) forecasting of the monthly discharge of 1952 to 1979:

The forecasting with the ARIMA model has been made on the same period that with ARMA.

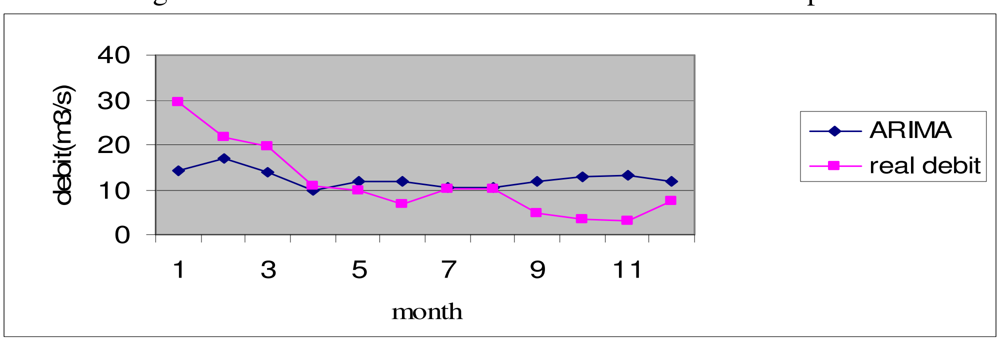

Figure 2: Comparison of the ARIMA(2, 1, 12) forecasting with the real value of discharge.

At the time of the use of the ARIMA(2, 1, 12) forecasting, that means when the process was integrate once, a reduction of difference between the forecasting and the real value of monthly discharge has been observed.

SARIMA(2,1,12)(1,1,1) forecasting of the monthly discharge of 1952 to 1979

The forecasting with the SARIMA method has been made on the same period that the two previous methods.

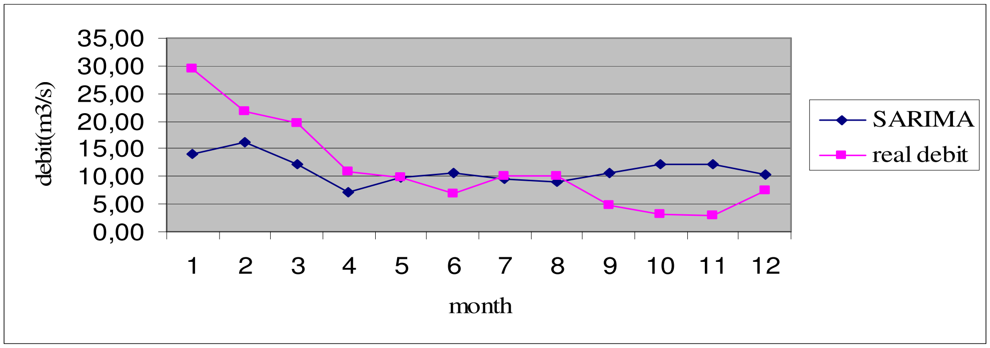

Figure 3: Comparison of the SARIMA (2, 1, 12) (1, 1, 1) forecasting with the real value of discharge.
While introducing the seasonal components, an improvement has been noticed for the forecasting.

Mistakes of forecasting of the three models(ARMA,ARIMA, and SARIMA) for the monthly discharge data:

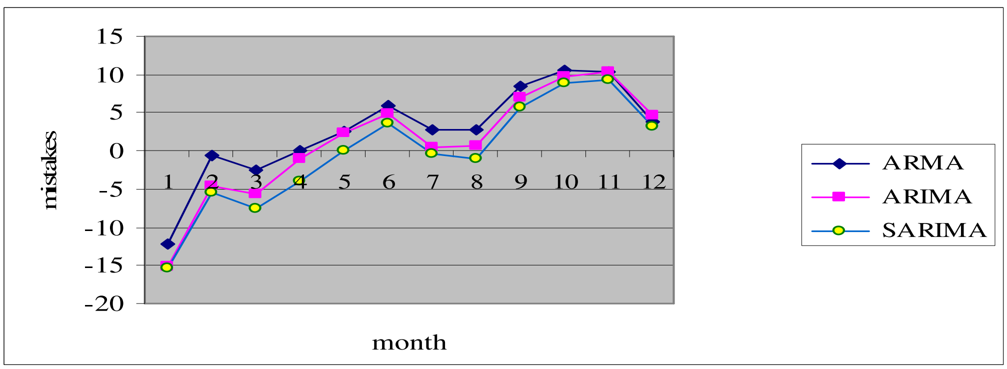

Figure 4: Mistake of three models ARMA, ARIMA and SARIMA forecasting of the monthly discharge data.

For the first four months of forecasting that means from January to April, the method ARMA (2,1) estimates the real values of discharge better, but for the other months, of May to December. The forecasting with seasonal components SARIMA (2, 1,12) (1, 1, 1) is the best model.

Results of rainfall forecasting:
ARMA(24,12) forecasting :

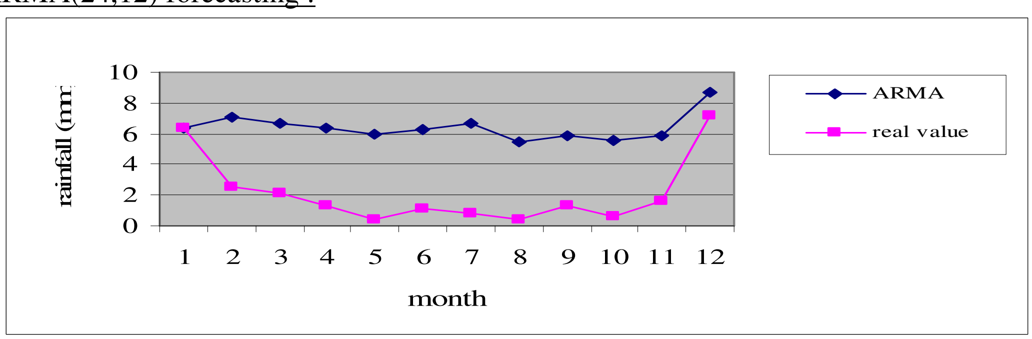

Figure 5: Values forecast by ARMA(24,12) and values real of the monthly rains.
A gap is noticed enter the forecasting with the ARMA(24,12) model and the real values, this distance is constant safe for the first and the last month of forecasting.

ARIMA(24,1,12) forecasting:

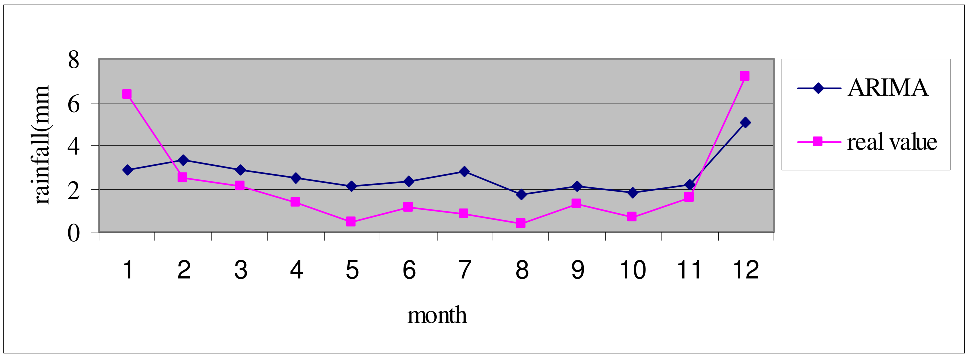

Figure 6: Values foreseen by ARIMA (24,1,12) and values real of the monthly rains.
The ARIMA(24,1,12) model estimates the values of rain better. On this figure, the difference between the ARIMA(24,1,12) forecasting and the real value is less important than the one observed with the model ARMA.

SARIMA(24,1,12)(1,1,1) forecasting:

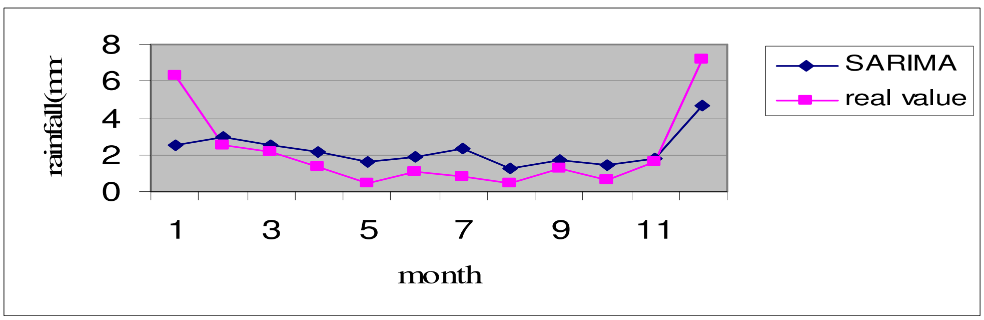

Figure 7: Values forecasting by SARIMA(24,1,12)(1 ,1,1) and real values of rains.
On this figure, the difference between the forecasting values and the real values are minimal.

Mistakes of forecastings of the monthly rainfall data by the three methods ARMA, ARIMA and SARIMA:

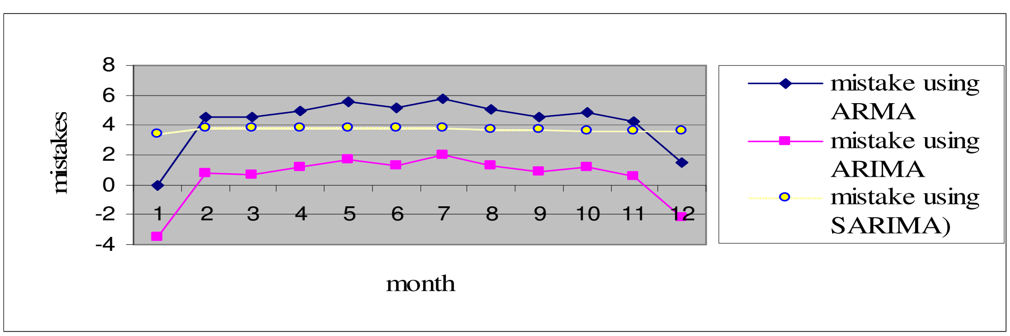

Figure 8: Mistakes of forecasting of the three ARMA, ARIMA and SARIMA models.
On this figure, it is the ARIMA model that presents the mistake the more close to zero, however, the committed mistake as using the SARIMA model is therefore constant while multiplying by a factor; It is the SARIMA model that is the best.

Discussion and Conclusion:

During this study, ARMA(2,12),ARIMA(2,1,12), and SARIMA(2,1,12)(1,1,1) are the pattern fixed to foresee the monthly discharge data.

With the ARMA(2,12) model, the mistakes of forecasting are big enough. It is due that the data must be integrated. With the ARIMA(2,1,12)model which integrates the data once, the forecasting improved, nevertheless, it remains always a some difference. The figure illustrating the mistakes committed while using the three models watch that the SARIMA(2,1,12)(1,1,1) model that is the best model to foresee the monthly discharge data. It's explained that the data must be integrated, and they also present a marked seasonal part that requires the hold in account of the seasonal components. It is the case of the SARIMA(2,1,12 model).

For the rainfall data, the p and q parameters used are not the same that those used at the time of the forecasting of the discharge data. It is due in fact that the parameters used changes in functions of the data to foresee. The choice of these parameters is made while using the method of maximum verisimilitude that gives the parameters for which the forecasting value is the likeliest. Model them used are: ARMA(24,12), ARIMA(24,1,12) and SARIMA(24,1,12)(1,1,1).

Among these models talked, SARIMA estimates best the real values for the rain forecasting, because on the discharge data, it is necessary to integrate the data first once, and the rainfall data present a seasonal part that divides the year in two, of November to April, the strong rains periods and May to October the dryness periods.

In conclusion, it is the SARIMA model that is the best model to forecast the rainfall and the discharge.

Litterature cited:

[1]: "Annuaire hydrologique de la France Outre-mer année 1954", author: office of the scientific and technical research overseas (ORSTOM), page 462-466; Edition Paris 1993 and CD-ROM version 2005.

[2] "Element d'hydrologie de surface", author,: J. P. Laborde, Doctor of science in hydrology. Edition 2000.

[3] "Fleuve et rivière de Madagascar", author,: Joël Danloux, Luc Ferry, Pierre Chaperon. I.R.D edition Paris 1993, CD-ROM version 2005.

[4] "Série temporelle, théorie et application, Volume 1, Introduction aux processus en temps discrèt, modèle ARIMA et méthode de Box & Jenkins", author: Arthur Charpentier, DESS Mathématique of the decision Edition 2000.