PREDICTION OF WIND SERIES IN THE NORTHERN PART OF MADAGASCAR

Andy TAMO[1], J. Eugène RANDRIANANTENAINA[2], I. Parker RAMAHAZOSOA[3], Adolphe A. RATIARISON[4]

[1]PhD student, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Toliara, Madagascar
2 Maître de conférences, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Toliara Madagascar
3 Maître de conférences, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo Madagascar
4 Professeur Emerite, Atmosphere, Climate and Oceans Dynamics Laboratory (DyACO), Physics and Applications, Sciences and Technologies, University of Antananarivo Madagascar

ABSTRACT

This research work consists in studying a phase in advance, allowing effective assistance to all those who have to make decisions regarding the planning and implementation of projects based on the wind in the Northern part of Madagascar. This part of the island is swept by a south-eastern Alizé wind regime known as "Varatraza". This wind blows violently and continues. With this in mind, we made a forecast of average monthly wind speeds based on daily low-level wind data from 1979 to 2017. Northern Madagascar is a high wind potential area with an average wind speed close to 9 m/s.

Keyword: *Varatraza,Wind speed, Northern part of Madagascar, Statistical Analysis, Prediction, ARIMA*

1. INTRODUCTION

Time series forecasting has been a long-standing issue. There are applications in many areas, such as economics, meteorology, energy, medicine, etc…

As an essential meteorological and climatological variable, wind is a source of energy called "renewable" but also of damage and it intervenes in countless physical phenomena or human activities [1].

Theoretically, time series prediction requires modelling the system that generated the series data. By having a system of mathematical and deterministic equations and knowing the initial conditions, it would be possible to predict the evolution of the system [2]. The general objective of this article, carried out on the northern part of Madagascar, is to predict the monthly changes in average wind speeds.

2. METHODOLOGIES

2.1. Databases

We used wind data at 950 hPa from the European Center ECMWF [3]. These data have a spatial discretization of 0.5°x0.5° in latitude and longitude and a temporal depth covering the period 1979-2017. Once these parameters have been extracted, the database used consists of 14245 daily wind states. The data are in the form of a table with n rows and p columns, which is stored as an X matrix of n p size.

2.2. Study area

The study area is the extreme north of Madagascar which is between (Fig-1):

• 12° South and 13.5° South latitude

• 48.5° East and 50° East longitude

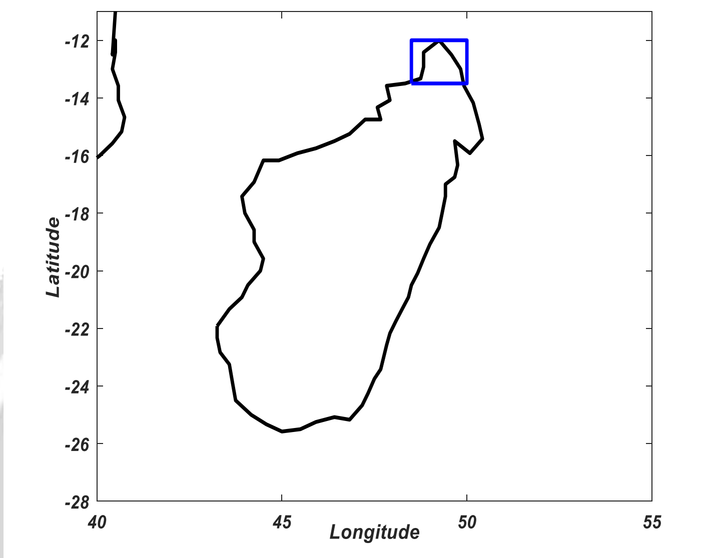

Fig-1: Representation of study area

2.3. Time series modelling [4]

Objective: To model and predict the future evolution of the time series from those observed.

2.3.1. ARIMA model

The class of ARIMA models [Box and Jenkins, 1976] was introduced to reconstruct the behaviour of processes subjected to disturbances over time and thus modify the values of the time series of observations. ARIMA models combine three types of time processes: Autoregressive Processes (AR-Autoregressive), Integrated Processes (I-Integrated), and Moving Average (MA-Moving Average).

The contribution of each of them is specified by the ARIMA rating (p, d, q), where p is the order of the AR (p) autoregressive process, d is the degree of integration of a I(d) process, and q is the order of the moving average MA(q).

2.3.2. ARIMA processes (p, d, q)

Let X_t be a series that is not stationary and has no seasonality. An ARIMA process (p, d, q) of the X_t series is a process of the following form:

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

With $$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots \dots \dots \dots \; \varphi_p L^P$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots \dots \dots \dots \; \theta_q L^q$$

And $\varepsilon_t \backsim BB(0, \sigma_\varepsilon^2)$, L is the delay operator, "d" is the degree of integration for the X_t series to become stationary ($d \geq 0$)and ($\varphi_1, \varphi_2 \ldots \ldots \varphi_p$)and $(\theta_1, \theta_2 \ldots \ldots \theta_q)$are the coefficients to be estimated.

2.3.3. SARIMA processes (p, d, q) (P, D, Q).

These processes are a generalization of the ARIMA models (p, d, q), containing a seasonal part.

$$(1 - L^s)\varphi(L)\phi(L^s)(1 - L)^d X_t = \theta(L)\vartheta(L^s)\varepsilon_t + \varphi_0$$

with

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots \ldots \ldots \ldots \varphi_p L^P$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots \ldots \ldots \ldots \theta_q L^q$$

$$\phi(L^s) = 1 - \phi_1 L^s - \phi_2 L^{2s} \ldots \ldots$$

$$\vartheta(L^s) = 1 - \vartheta_1 L^s - \vartheta_2 L^{2s} \ldots \ldots$$

2.3.4. Box and Jenkins methodology

Three steps are necessary for the construction of the models: an identification phase, an estimation phase, a validation phase [5]. The Box-Jenkins method allows us to determine the order (p, d, q) of an ARIMA process. To determine "p" and "q" it is necessary to trace the PACF (partial correlogram) and the ACF (correlogram) of the stationary ARMA process obtained after differentiations if necessary. These two graphs give us respectively the maximum value of p and q noted: $p_{max}$and q_{max} [6]. At the end of these three phases, once the best ARMA model, that is to say the one with the lowest BIC (Baisian information critérion) is determined (Schwarz, 1978), this model is used for forecasting purposes in the «h» horizon.

3. RESULTS

3.1. Monthly wind speeds

For the 39 years of study, the evolution of the monthly average speeds is represented by the figure 2. For modelling, we took the 82%, or 384 months from 1979 to 2010 (In blue). The remaining 84 months from 2011 to 2017 will be used for model validation (In green).

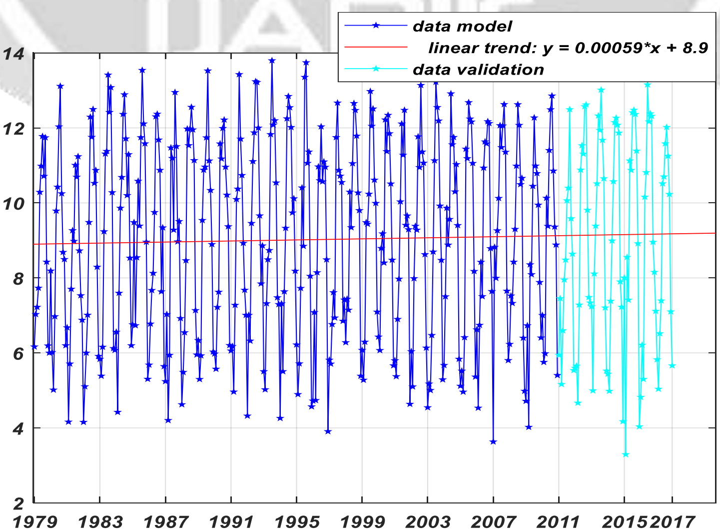

Fig-2: Evolution of monthly wind speeds in Northern Madagascar from 1979 to 2017(m/s)

Before starting the modelling, we have to check our series if it satisfies both stationary and non-seasonal conditions.

3.2. Statistical tests

Based on the various stationary and trend tests, we can conclude that our monthly wind time series in Northern Madagascar is stationary and does not show a significant trend therefore there is no differentiation to run (Tab-1). Thus in our SARIMA model, the parameter d=0.

Tab-1: Summary of Statistical Tests

Test	KPSS	Dickey-Fuller	Mann-Kendall
Hypothesis	H=0	H=1	H=0
p_value	0.1	0.04	0.6
alpha	0.05	0.05	0.05
Results	Stationary	Stationary	No tendency

3.3. Estimation of seasonality

The correlation shows the existence of a cyclic element with in the series (Fig-3). We have noticed significant peaks at the few offsets (as at shift k = 12 months, k = 24 months, k = 36 months, etc.) that signal us despite our series being stationary, it also has a seasonality (12 months) that we must remove before modelling the series.

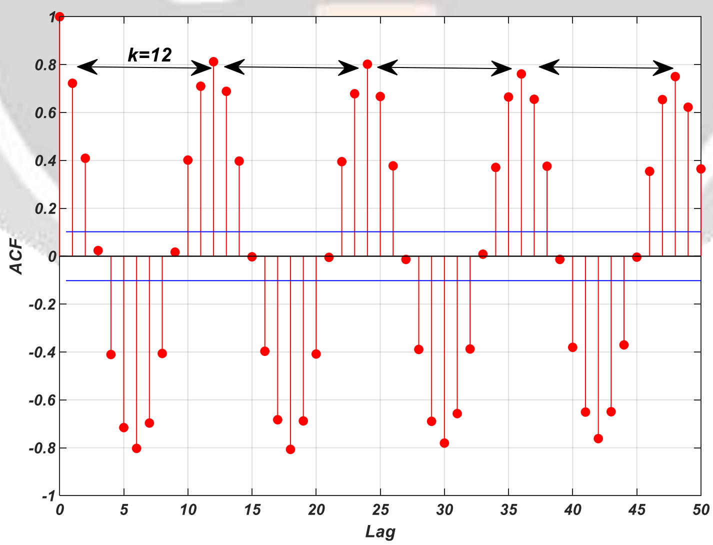

Fig-3: Initial series autocorrelogram

3.4. Seasonal adjustment of series by first seasonal differentiation

Since we have to work on a stationary and also non-seasonal series, we made a first seasonal difference (D = 1) to remove the seasonality because at this stage it is already stationary (d = 0). And after the differentiation [seasonal (D=1)], we can clearly see that our series is both stationary and not seasonal (Fig-4).

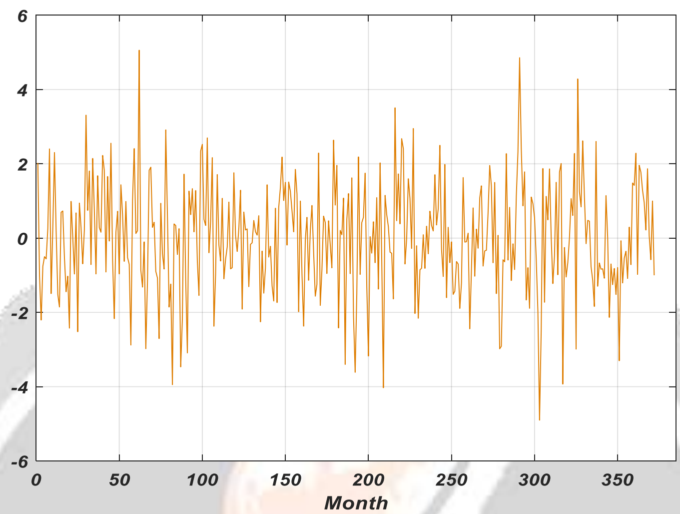

Fig-4: Seasonally adjusted series

3.5. Correlograms analysis

As we know that this series has a seasonality of 12 months (S=12) and we notice that the ACF and the PACF have peaks that go out of the confidence interval at offset k=12 then in this case the parameters P=Q=1, so we still have to find p and q. Now our model can be written as follows:

$$SARIMA\ (p,\ d,\ q)\ (P,\ D,\ Q)_S = SARIMA\ (p,\ 0,\ q)\ (1,\ 1,\ 1)_{12}$$

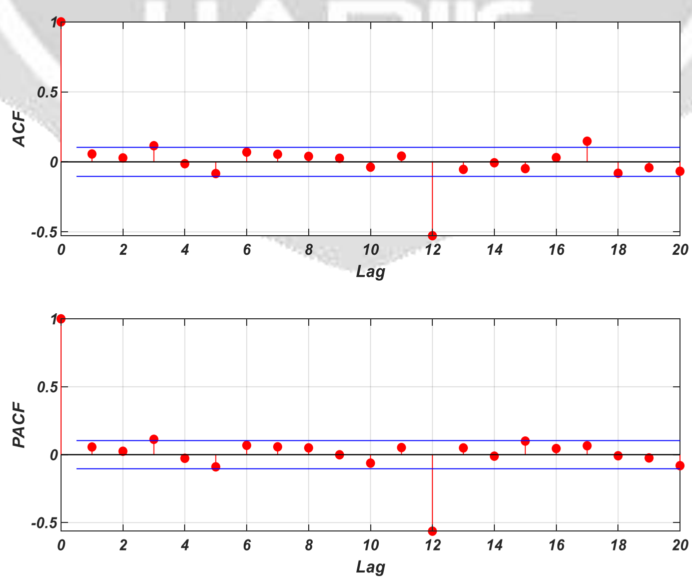

Fig-5: Autocorrelogram and partial autocorrelogram of the seasonally adjusted monthly wind time series.

To estimate the parameters, we will take the last peak that comes out of the confidence interval (horizontal line in blue) for each autocorrelogram and that we will consider as a reference of the maximum value.

By viewing the PACF (Fig-5), the last peak that comes out of the confidence interval has a 15 offset so the p value varies from 1 to 15 ($1 \le p \le 15$), and according to the ACF(Fig-5) the q value varies from 1 to 17 ($1 \le q \le 17$). We'll consider the peaks coming out of the confidence interval.

3.6. Identification of the model.

Based on previous results, the values of the selected parameters are:

- d=0
- D=1
- P=1
- Q=1

It remains for us to combine the significant values of p and q, that is, the peaks that come out of the confidence interval, then find the best combination (Tab-2).

The table presents the value of BIC for the candidates of the model to be selected. The model that gives the minimum Bayesian Information Criterion (BIC) is that of the SARIMA model $(3, 0, 3)(1, 1, 1)_{12}$ with a BIC of 1212.1.

Tab-2: Comparison of the BIC value of each model

SARIMA Model	BIC($x10^3$)	SARIMA Model	BIC($x10^3$)
$(3, 0, 3)(1, 1, 1)_{12}$	1.2121	$(12, 0, 3)(1, 1, 1)_{12}$	1.2500
$(3, 0, 5)(1, 1, 1)_{12}$	1.2220	$(12, 0, 5)(1, 1, 1)_{12}$	1.2472
$(3, 0, 12)(1, 1, 1)_{12}$	1.2312	$(12, 0, 12)(1, 1, 1)_{12}$	1.2802
$(3, 0, 17)(1, 1, 1)_{12}$	1.2557	$(12, 0, 17)(1, 1, 1)_{12}$	NaN
$(5, 0, 3)(1, 1, 1)_{12}$	1.2210	$(15, 0, 3)(1, 1, 1)_{12}$	1.2631
$(5, 0, 5)(1, 1, 1)_{12}$	1.2300	$(15, 0, 5)(1, 1, 1)_{12}$	1.2753
$(5, 0, 12)(1, 1, 1)_{12}$	NaN	$(15, 0, 12)(1, 1, 1)_{12}$	NaN
$(5, 0, 17)(1, 1, 1)_{12}$	1.2727	$(15, 0, 17)(1, 1, 1)_{12}$	NaN

3.7. SARIMA $(3, 0, 3)(1, 1, 1)_{12}$ Model Quality Check

The residual autocorrelation function (RACF) and the residual partial autocorrelation function (RPACF) must be calculated to determine if the residues are white noise.

The RACF and RPACF values are with in confidence limits (Fig-6.). The figures do not indicate any significant correlation between residues.

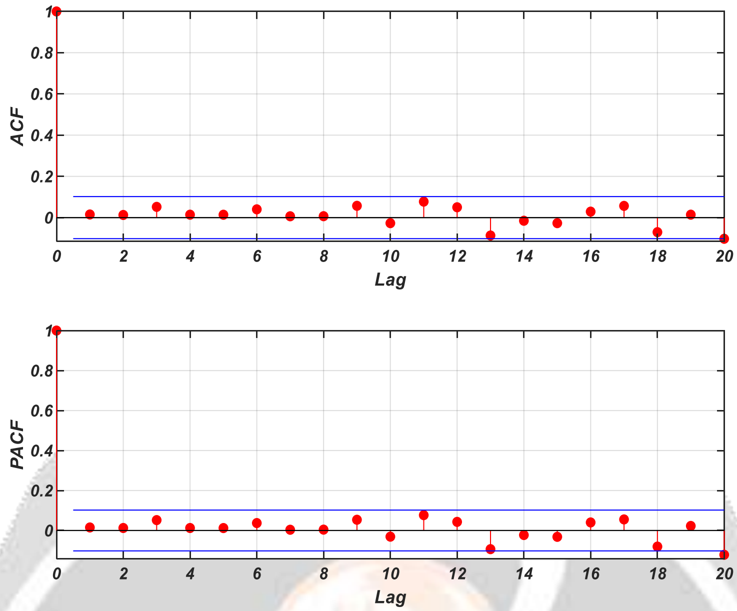

Fig-6: Autocorrelogram and Partial Autocorrelogram of residues.

The graph of the cumulative distribution for residual data normally appears as a straight line when it is carried on paper of normal probability (Fig-7), so the assumptions of normality of residues hold.

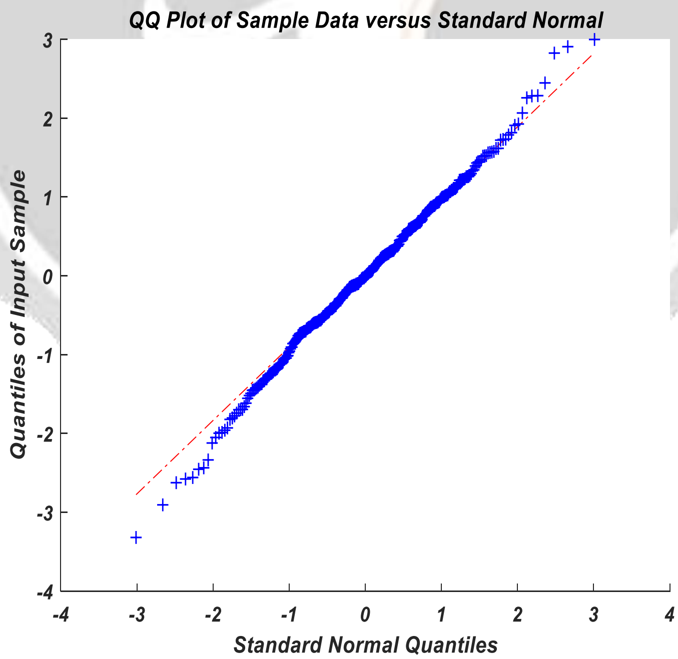

Fig-7: The quantile-quantile plot (QQ-plot) of the residues for SARIMA (3, 0, 3) $(1, 1, 1)_{12}$

The estimation of the density of the nucleus shows no obvious violation of the normality hypothesis (Fig-8). The residues are normally distributed, we can accept the normality of our residues and this means that the residues are white noises.

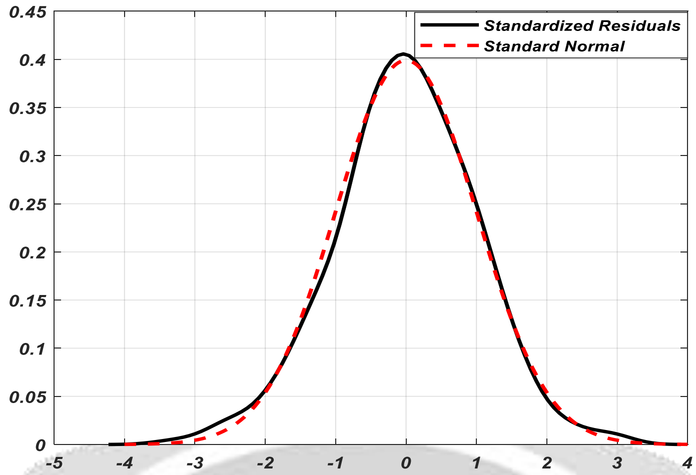

Fig-8: Normality of SARIMA residues (3, 0, 3) $(1, 1, 1)_{12}$ monthly wind

The results of the Ljung-Box test retain the null hypothesis according to which a series of residues does not show autocorrelation (i.e., h=0 and p_value is 0.601). We can conclude that this is a good model.

To confirm the homoscedasticity of our residues, we simply need to trace the ACF and the PACF of the residues squarely (Fig-9). According to the figure there is no peak that comes out of the confidence interval (right in blue), our residues are indeed homoscedastic.

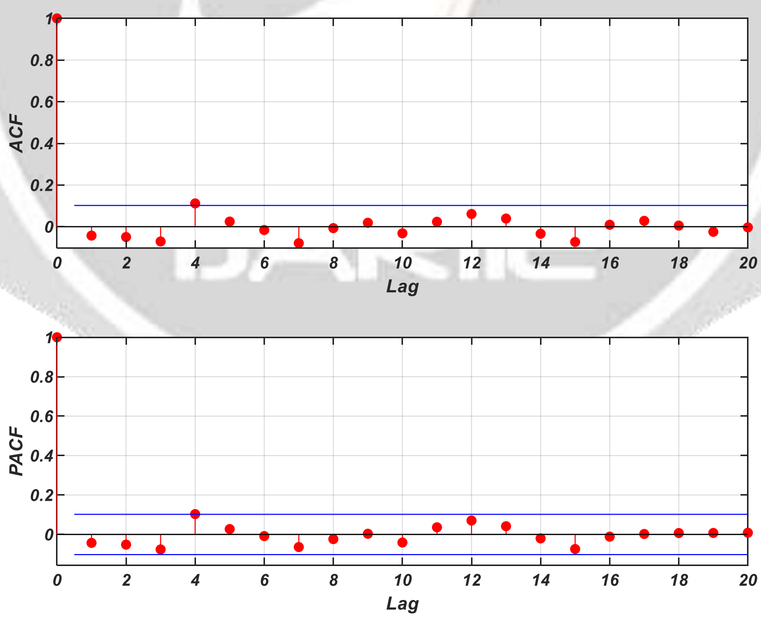

Fig-9: Autocorrelogram and Partial Autocorrelogram of the squared residuals.

3.8. Validation of the SARIMA model (3, 0, 3) $(1, 1, 1)_{12}$

The various tests we performed above on the residues summarize us that:

- Our residues are not self-correlated
- They follow a normal law
- They are homoscedastic

Our residues meet these three conditions well so they behave like white noises. This helps us validate our SARIMA model (3, 0, 1) $(1, 1, 1)_{12}$ of the monthly wind.

3.9. Equation of the SARIMA model (3, 0, 3) $(1, 1, 1)_{12}$

The mathematical prediction equation of the SARIMA model (3, 0, 3) $(1, 1, 1)_{12}$ is given by:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - \Phi_{12} L^{12}) Z_t = (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3)(1 + \Theta_{12} L^{12})\varepsilon_t + \text{cte}$$

is given by:
with
L is the delay operator, $Lx_t = x_{t-1} \quad \forall\, t \in \mathbb{Z}$

$Z_t = \nabla^0 \nabla_{12}^1 X_t = (1 - L^{12}) X_t = X_t - X_{t-12}$,X_t initial series

$\varepsilon_t \epsilon N(0, \sigma^2)$ is a white noises

and the estimated regression coefficients:
$\phi_1 = -0.0419\theta_1 = 0.1076\phi_{12} = -0.1010$
$\phi_2 = -0.5113\theta_2 = 0.5416 \quad \phi_{12} = -0.8394$
$\phi_3 = 0.1686\theta_3 = 0.0021\text{cte} = 0.0118$

3.10. Monthly wind prediction of the selected model (3, 0.3) $(1, 1.1)_{12}$

After calculating the average absolute error percentage (Mean Absolute Percentage Error or M.A.P.E), we were able to find the value M.A.P.E = 0.02% ($M.A.P.E < 10$), our forecast using the model is excellent. The figure-10 represents the model's prediction for the future months (from 469th to 528th months) which is a forecast to 2021.

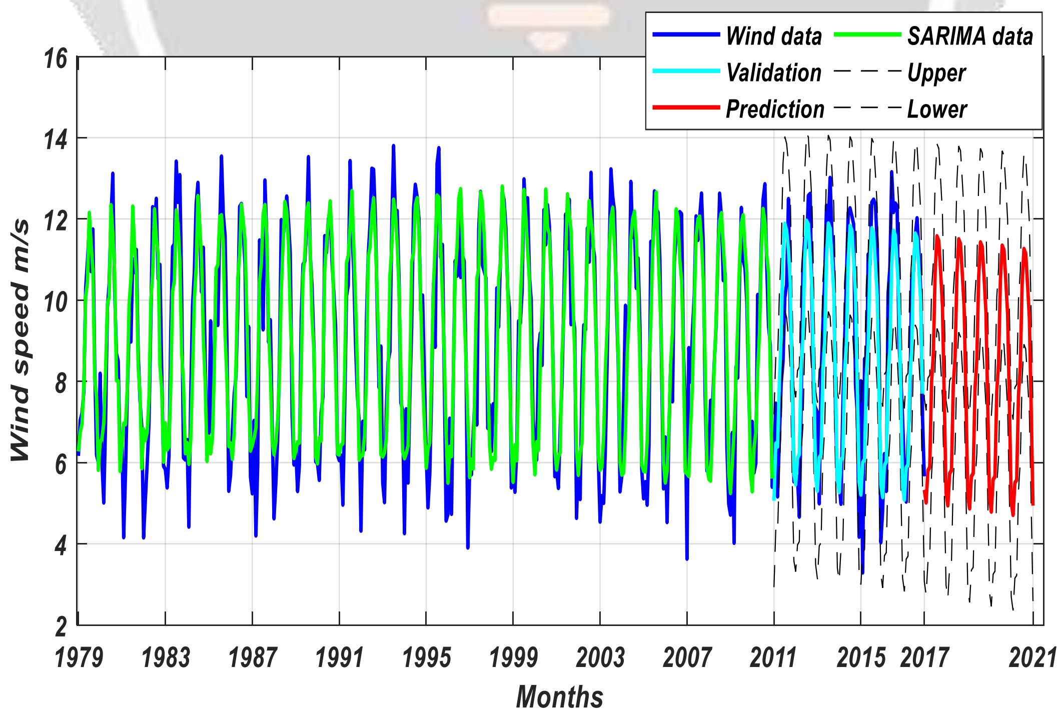

Fig-10: Adjustment of the SARIMA model (3, 0.3) $(1, 1.1)_{12}$ to the monthly wind series X_t.

Tab-3: Future values of monthly wind speeds in Northern Madagascar in m/s

Years	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	5.8	7.9	8.6	10.6	11.5	11.4	10.6	9.8	7.6	5.2	4.9	5.7
2019	5.7	7.8	8.5	10.6	11.4	11.3	10.6	9.7	7.5	5.1	4.8	5.6
2020	5.7	7.7	8.5	10.5	11.3	11.2	10.5	9.6	7.4	5.0	4.7	5.0
2021	4.7	5.5	5.6	7.7	8.4	10.4	11.3	11.1	10.4	9.6	7.3	4.9

4. CONCLUSION

This work provides statistical modelling of the monthly wind speeds in Northern Madagascar and predicting future values until 2021. In order to optimize the exploitation of wind energy, certain measures must be taken into account. A poor choice of parameters may harm a wind plant or other uses. For this reason, a preliminary study phase is of great interest since it will be possible to identify the conditions and constraints to be taken into account when carrying out a wind project.

The statistical model that generates the wind phenomenon in this region was proposed. The monthly wind speeds predicted in this region are regular with an average of 8.5 m/s and the trend is increasing. The study of wind speeds in the north of Madagascar reveals speed values, which may favour exploitation for the production of energy from the wind.

5. REFERENCES

[1]. Pierre Carrega, (2008). Le vent: Importance, mésures, modélisation et tribulations, Bulletin de la société géographique de Liège, 51, pp.17-29

[2]. D.K. Kidmo, N. Djongyang, S.Y. Doka, and D, (2014). Raidandi ‘Assessment of Wind Energy Potential for Small Scale Water Pumping Systems in the North Region of Cameroon’, International Journal of Basic and Applied Sciences, Vol3, N°1, pp. 38 – 46

[3]. European Centre for Medium-Range Weather Forecasts, www.ecmwf.int, (June 2018).

[4]. BELAID SABRINA, (2010). Modélisation des données de l’irradiation solaire

[5]. Dominique DESBOIS, (2005). Une introduction à la méthodologie de Box et Jenkins, Revue MODULAD

[6]. AYIAGNIGNI GBAMBIE Mohamed Mustapha, (2012). Analyse statistique de données de vent côtier au Nord du Golfe de Guinée, p 10