IEC
Institute of Electronics
and Computer

Journal of Artificial Intelligence and Systems, 2024, 6, 135-145
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

Modelling Precipitation as a Function of Several Variables in Northern Madagascar Using the ANFIS Network

RABENIAINA Anjara Davio Ulrick[a,*], RAKOTOVAO Niry Arinavalona[b], Ambeondahy[c], Hary Jean[d]

Laboratory of Atmosphere, Climate and Ocean Dynamics, University of Antananarivo, Science and Technology, Antananarivo, Madagascar
[a]davioulrick@gmail.com; [b]rakniry@gmail.com; [c]Ambeondahyp@gmail.com; [d]jjeanhary@yahoo.com
*Corresponding Author: RABENIAINA Anjara Davio Ulrick, Email: davioulrick@gmail.com

How to cite this paper: RABENIAINA Anjara Davio Ulrick, RAKOTOVAO Niry Arinavalona, Ambeondahy, Hary Jean (2024). Modelling Precipitation as a Function of Several Variables in Northern Madagascar Using the ANFIS Network. Journal of Artificial Intelligence and Systems, 6, 135–145.
https://doi.org/10.33969/AIS.2024060110.

Received: March 20, 2024
Accepted: May 7, 2024
Published: May 10, 2024

Copyright © 2024 by author(s) and Institute of Electronics and Computer.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

This article presents a study on the modeling of precipitation in northwest Madagascar as a function of several variables. The method is based on the use of the Granger test to assess the causal relationship between precipitation, temperature, atmospheric pressure, wind at 925 hPa and Outgoing Longwave Radiation (OLR). The model is developed using the multiple linear regression method, followed by the use of the Sobol index to identify the most influential of these variables. Forecasts are based on the ANFIS model. The results show that precipitation in this area (Sofia, Diana and Sava regions) is influenced by pressure, wind and OLR. The explanatory model obtained highlights the relationship between precipitation and these climatic variables, showing a 7 mm decrease in precipitation according to the ANFIS model developed.

Keywords

Precipitation, Granger causality test, Modeling, Multiple linear regression, ANFIS.

1. Introduction

Precipitation modeling is a fundamental challenge in the fields of climatology and meteorology, given its vital importance for water resource management, weather prediction and many other essential applications. However, it is crucial to recognize that precipitation is the result of a complex combination of various climatic factors.

In Madagascar, where agriculture is the backbone of the economy and most of the population depends on crops for their livelihood, rainfall is of vital importance. Rainfall, which generally occurs between November and April, conditions not only agricultural yields, but also food security and the well-being of local communities.

The aim of this article is therefore to present a detailed analysis aimed at identifying the predominant climatic variables associated with rainfall in this region. The first part of the article presents the study area and the data used, followed by an explanation of the methodologies employed to arrive at the results obtained.

2. Methods

2.1. Study area

The study region is located in the north of Madagascar and is delineated as shown in figure 1 below. The regions concerned are Sofia, Diana and Sava.

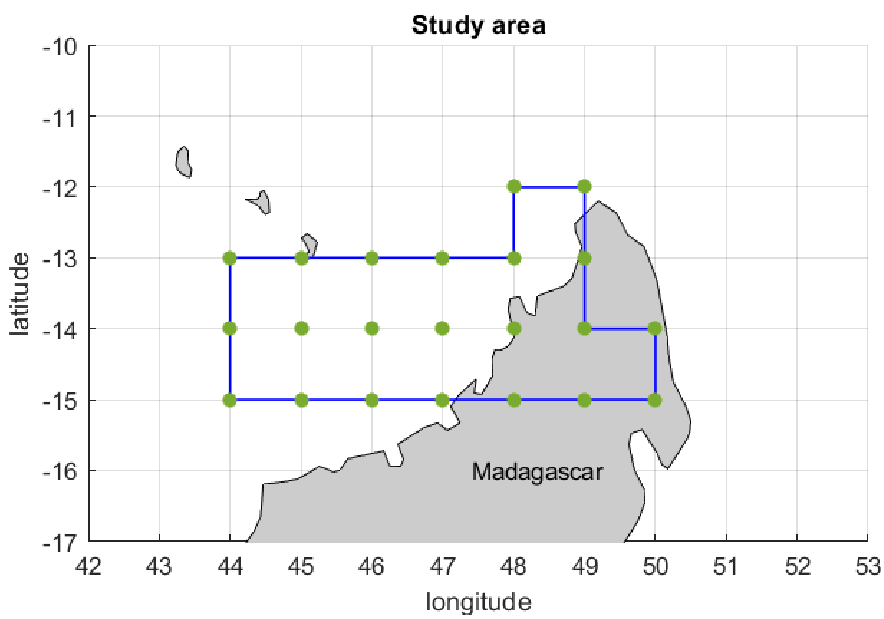

Figure 1. Numbers of IoT devices linked over time

2.2. Database

The data used to carry out this study are wind field at 925 hPa, temperature at 2m altitude, and precipitation data supplied by ECMWF (European Center for Medium-Range Weather Forecasts) with a grid resolution of 1° x 1°, and covering the period from 1979 to 2017.

The OLR (Outgoing Longwave Radiation) data are sourced from NOAA (National Oceanic and Atmospheric Administration) with a grid resolution of 2.5° x 2.5°, covering the same period.

2.3. Method of Interpolation

In this study, we used interpolation, a mathematical technique aimed at estimating the intermediate points between well-known data. To do this, we looked for the best polynomial to pass through a set of given points, which enabled us to extrapolate intermediate values. The most basic interpolation is linear interpolation, where two points are connected by a straight line. However, to obtain more accurate results and smooth the data, we opted for polynomial interpolation using Lagrange's method [1].

In the two-dimensional context we are interested in, the two-variable interpolation function must satisfy the following equation 1:

$$f(x,y) \approx \hat{f}(x,y) = \sum_{j=0}^{m}\sum_{i=0}^{n} f(x_i, y_j)U_i(x)V_j(y) \qquad (1)$$

Where $U_i(x) = \prod_{\substack{k=0 \\ k\neq i}}^{n} \frac{(x-x_k)}{(x_i-x_k)}, V_j(x) = \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{(y-y_k)}{(y_j-y_k)}$ are the Lagrange polynomials. x_i and y_j represent the geographic coordinates of a given point, while $f(x_i, y_j)$ corresponds to the known value of the parameter at that specific point. Our objective is to deduce the unknown value $f(x,y)$ for a point with coordinates (x,y) specified.

2.4. Causality in the Sense of Granger

Granger-causality is a statistical concept used to assess whether there is a causal relationship between two time series. The Granger-causal time series provides additional predictive information for the target time series. The following hypotheses are tested by the F-test [2]:

- If the F-test statistic > the critical value (C), the null hypothesis is rejected, indicating that series y causes series X. Then Y causes X.

- If the F test statistic < the critical value (C), the null hypothesis is not rejected, suggesting that series y does not cause series x. Therefore, Y does not cause X.

2.5. Multiple Linear Regression

Multiple regression generalizes the concept of simple regression. It makes it possible to predict a response variable Y using not a single predictor X, but rather m predictors, designated as X_1; X_2;…; X_m. By combining these independent variables, we seek to establish a linear relationship that best explains the variations in the dependent variable Y [3]. The general equation (2) of a multiple linear regression is written as:

$$Y = \beta_j X_j + \varepsilon \qquad (2)$$

Where

β_j Unknown coefficients of the model.

X_j Vectors of the different explanatory (independent) variables.

Y Vector of the dependent variable.

2.6. Sensitivity Analysis with the Sobol Method

Global sensitivity analysis methods, such as the Sobol method, look at the effect of uncertain parameters over their entire range of variation in relation to the model output. They take into account the probability distribution of each parameter and examine the impact of all the parameters simultaneously. In the case of the Sobol method, it provides sensitivity indices of various orders to assess the response of the output to uncertain parameters [4].

- Total Effect Index (ST):

The total effect index measures the total contribution of each input variable and all its interactions to the output variance. It is calculated as follows:

$$ST = \frac{Var(f(x))}{Var(Y)}$$

Where

$Var(f(x))$ is the variance of the output $f(x)$, where X is a set of all input variables.

$Var(Y)$ is the total variance of the output Y (the variance of $f(x)$ when all input variables are considered).

- First-order index (S1, S2, …, Sn)

First-order indices measure the individual contribution of each input variable to the variance of the output, ignoring interactions with other input variables. The first-order indices for each variable are calculated as follows:

$$S_i = \frac{Var(f({}^{X}/_{X_i}))}{Var(Y)}$$

Where

$Var(f({}^{X}/_{X_i}))$ is the variance of the output $f(X)$ when the variable X_i is removed from the set X.

$Var(Y)$ is the total variance of the output Y.

- Interaction index (S_{ij})

Interaction indices measure the joint contribution of two or more input variables, including their interactions, to the variance of the output. Interaction indices between two variables, for example X_i and X_j, are calculated as follows:

$$S_{ij} = \frac{Var(f(X/X_i, X/X_j))}{Var(Y)}$$

2.7. Hybrid Network: ANFIS

ANFIS (Adaptive Neuro-Fuzzy Inference System) is an inference system that integrates the learning abilities of artificial neural networks with the knowledge representation of fuzzy inference systems to address complex modeling and prediction tasks [5]. The ANFIS network consists of five main layers [6]:

- The first layer corresponds to the input nodes that receive the input data.
- The second layer consists of membership nodes, which apply relevance functions to the input data.
- The third layer is formed by rule nodes that combine the relevance functions to generate fuzzy rules.
- The fourth layer consists of consequence nodes that calculate the contribution of each rule to the overall system output.
- The fifth layer is the output node, which combines all the rule contributions to produce the final output of the system.

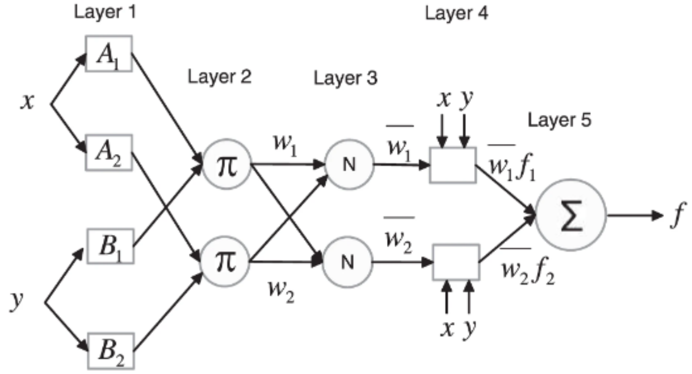

Figure 2. Architecture of ANFIS [6]

3. Results

3.1. Results of Granger causality tests

Granger causality tests reveal that wind, atmospheric pressure, and OLR (Outgoing Longwave Radiation) are causal factors for precipitation in Northern Madagascar. However, temperature does not seem to play a causal role in precipitation formation in this geographic area.

3.2. Multiple linear regression

In the studied region, the amount of precipitation will be assessed using sea-level pressure, wind field, and OLR. The corresponding relationship is given by the following equation, which is obtained from a multiple linear regression analysis:

$$\begin{aligned} y = &-7705,7322x_3 - 1,0976685x_3^2 - 11,6303x_2x_3 - 9,3779399e^{(-5)}x_2x_3^2 \\ &+2122,1235x_1 + 15,852975x_1x_3 + 0,0010874155x_1x_3^2 + 3,1092272x_1x_2 \\ &+0,011532962x_1x_2x_3 - 4,2816204x_1^2 - 0,0081424108x_1^2x_3 - 0,0030764795x_1^2x_2 \\ &+0,0021589579x_1^3 \end{aligned}$$

The graphs in figure 3 highlight a significant correlation between precipitation and the three measured variables (pressure, wind, and OLR), with a regression coefficient of 0.98. Thus, the model used satisfactorily explains the variations in precipitation in this area.

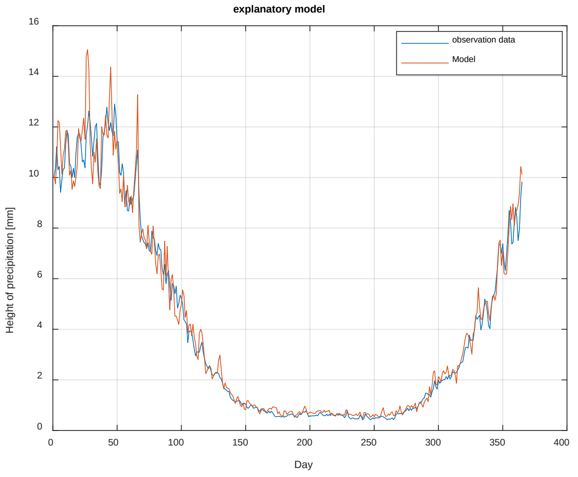

Figure 3. Explanatory model from multiple linear regression

3.2. Sobol Indices Analysis

Table 1 summarizes the results of the Sobol indices, highlighting the significant contribution of variables such as wind, pressure, and OLR, as well as their interactions, to the total variance of precipitation. The results particularly underscore the role of atmospheric pressure as a major factor, as well as the importance of complex interactions between these climatic variables. Additionally, the OLR variable (S3) also contributes significantly, albeit negatively, to precipitation variance, indicating an inverse relationship.

Table 1: Results of the Sobol index calculation

ST (Main Index)	**5329864.358870**
S1 (wind)	**-0.291441**
S2 (atmospheric pressure)	**354.677152**
S3 (OLR)	**-1285.667972**
S12 (wind-pressure)	**5329509.973160**
S13 (wind-olr)	**5331150.318283**
S23 (pressure-OLR)	**5330795.349690**

3.3. ANFIS Modeling

3.3.1. Model Architecture

Figure 4 below illustrates the ANFIS model used to model precipitation as daily climatological averages. The model is characterized by four input variables, one output variable and 17 rules. Each model input variable is associated with 17 fuzzy sets.

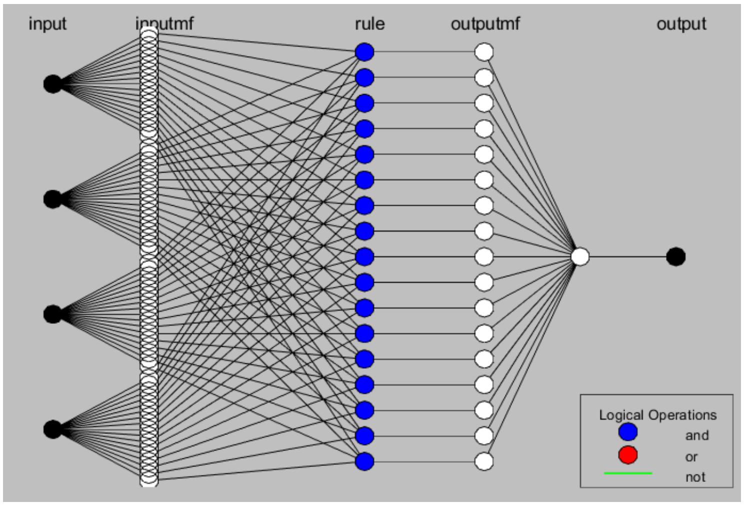

Figure 4. Model architecture

3.3.2. Fuzzy rules

After 200 iterations, the total number of fuzzy rules acquired is 17. Here are a few examples:

- *1.If (Φ_1 is A1) and (Φ_2 is B1) and (Φ_3 is C1) and (Φ_4 is D1) then (ω is M1)*
- *2.If (Φ_1 is A2) and (Φ_2 is B2) and (Φ_3 is C2) and (Φ_4 is D2) then (ω is M2)*

  ....

- *3.If (Φ_1 is A16) and (Φ_2 is B16) and (Φ_3 is C16) and (Φ_4 is D16) then (ω is M16)*
- *4.If (Φ_1 is A17) and (Φ_2 is B17) and (Φ_3 is C17) and (Φ_4 is D17) then (ω is M17)*

3.3.3. Model validation

The figure 5 shows both the model curve and the initial series, with an RMSE of 0.22. Furthermore, the coefficient of the regression line between the model output and the initial series (figure 6) is 0.99. These results clearly indicate that the model obtained is excellent.

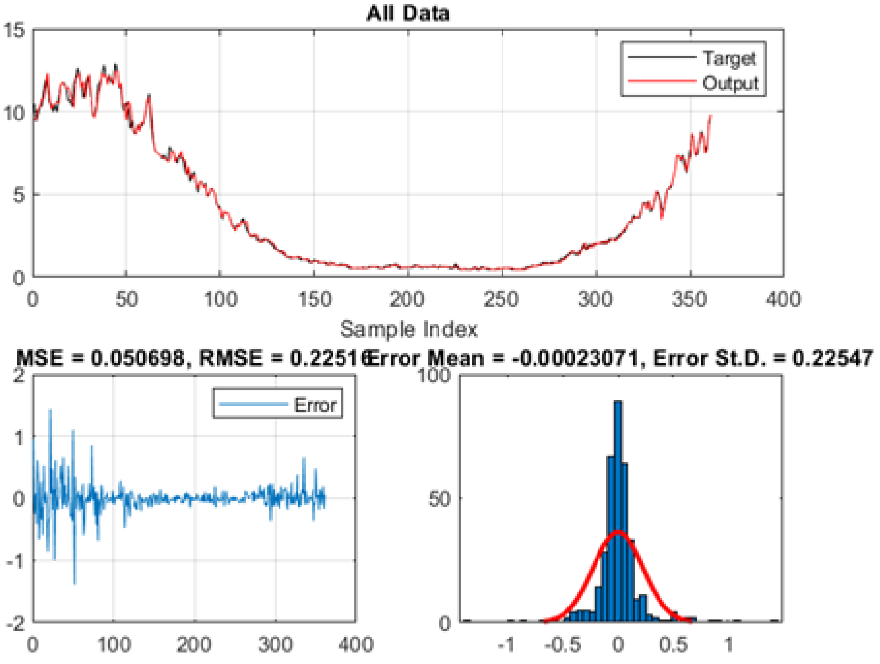

Figure 5. Model curve from ANFIS

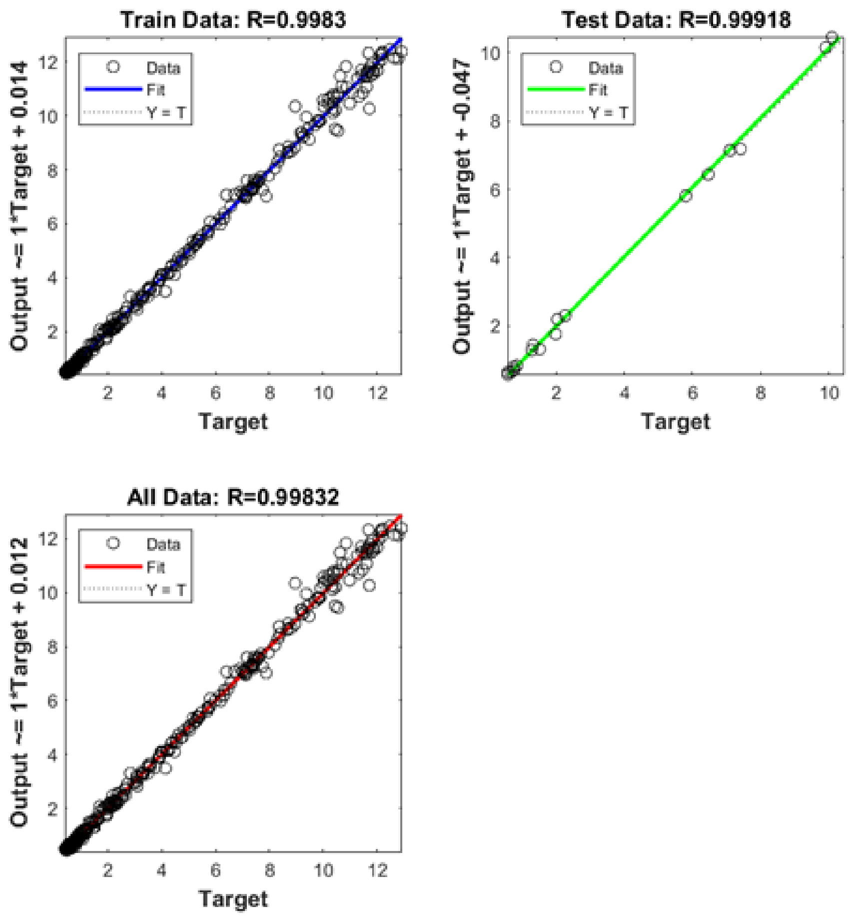

Figure 6. Regression line

3.3.4. Model prediction

The figure 7 presents the forecast of daily climatological average precipitation, which shows almost overlapping curves with the observation curves. According to this forecast, a decrease of approximately 7 mm in precipitation height is anticipated. This result highlights a negative correlation between precipitation and the variables of wind, pressure, and OLR in the studied region. In other words, when these variables increase or decrease, the amount of precipitation tends to decrease proportionally. This underscores the significant impact of these climatic factors on precipitation amount, despite the negative correlation, thus highlighting the complexity of local climate relationships in our study region.

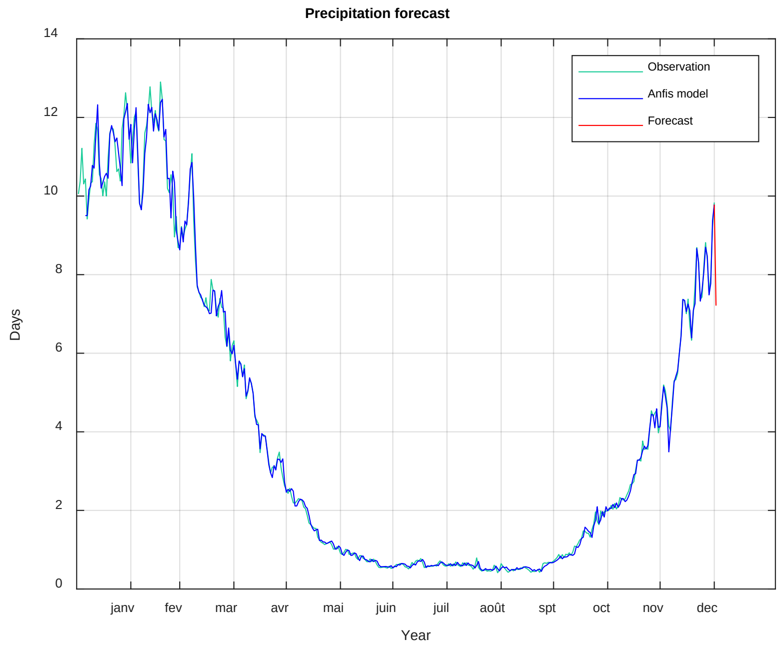

Figure 7. Precipitation forecasts from the ANFIS model

4. Conclusion

In conclusion, this article has revealed that precipitation in the Sofia, Diana and Sava region is influenced by atmospheric pressure at sea level, wind at 925 hPa and OLR. The explanatory model derived from multiple linear regression clearly determines the relationship between precipitation and these climatic variables, highlighting different types of correlation. Atmospheric pressure plays a major role in contributing to the variation in precipitation. Furthermore, the ANFIS model with four inputs, seventeen rules and one output showed remarkable performance with low RMSE values, enabling a reliable prediction for the future. This forecast indicates an anticipated decrease of around 7 mm in precipitation height.

References

[1] Salim Ali : Études de la variabilité climatique et de l'impact des modes ENSO et IOD au Nord de Madagascar et aux Comores. 2018

[2] Rajaonarisoa : Caractérisation de l'évolution et analyse de l'influence des paramètres climatiques température, humidité et pression atmosphérique sur la précipitation annuelle au sud de Madagascar. 2022.

[3] Rousson, Valentin : Régression linéaire multiple. In : Statistique appliquée aux Sciences de la vie. Springer, 2013, S. 219–258

[4] KHABBAZ, Mohamed ; BENHAMOU, Brahim ; LIMAM, Karim ; HAMDI, Hassan ; BENNOUNA, Amin : ÉTUDE DE SENSIBILITÉ ET PARAMÉTRIQUE COMPLETE D'UN ÉCHANGEUR AIR-SOL POUR LE RAFRAICISSEMENT DE L'AIR SOUS LE CLIMAT CHAUD ET

SEMI-ARIDE DE MARRAKECH. (2016)

[5] D.J. Cho, S. Kim, and K. Lee, "An adaptive-network-based fuzzy inference system," IEEE Transactions on Fuzzy Systems, vol. 6, no. 2, pp. 210-221, May 1998.

[6] Jang, J.-S. R. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665-685.