



Available online at
<https://jurnalteknik.unisla.ac.id/index.php/CVL>
<https://doi.org/10.30736/cvl.v2i2>



Rainfall Forecasting Using Historical Rainfall and Humidity Data in Wonorejo Watershed, Indonesia

Nastasia F. Margini¹, M. Bagus Ansori²

^{1,2} Lecturer at Department of Civil Engineering, Institut Teknologi Sepuluh Nopember, Surabaya
60111, Indonesia.

Email : ^{1*}nastasia@its.ac.id. ^{2*}bagusansori@its.ac.id

ARTICLE INFO

Article History :

Article entry : 11-03-2026
Article revised : 08-04-2026
Article received : 05-05-2026

Keywords :

Rainfall forecasting,
Decomposition method,
Regression analysis,
Nonlinear Regression, Water
resource management,
Wonorejo Watershed

IEEE Style in citing this article :

N. Margini and M. B. Ansori,
"Rainfall Forecasting Using
Historical Rainfall and Humidity
Data in Wonorejo Watershed,
Indonesia ", CVL, vol. 11, no. 1,
pp. 123–133, Mar. 2026.

ABSTRACT

Rainfall prediction plays a crucial role in the design, planning, and management of water resource systems. This study addresses the importance of accurate rainfall forecasting to optimize the operational efficiency of the Wonorejo Reservoir in Tulungagung Regency, Indonesia. By improving rainfall predictions, it becomes possible to estimate available water volumes more precisely and enhance reservoir utilization beyond current practices. This research employs three analytical approaches—Decomposition, Multiple Linear Regression, and Nonlinear Regression—to forecast rainfall in the Wonorejo Watershed. The models are developed using monthly total rainfall and average air humidity data collected from January 1998 to December 2018. Model performance is evaluated using the Root Mean Square Error (RMSE). The Nonlinear Regression method yields the lowest RMSE (72.71), followed by Decomposition (73.78) and Multiple Linear Regression (103.57). Based on these results, the Nonlinear Regression model is identified as the most suitable approach for forecasting rainfall in the Wonorejo Basin for the subsequent 24 months. This study recommends the application of the Nonlinear Regression method for rainfall forecasting in this specific case study area.

1. Introduction

Accurate rainfall forecasting is a fundamental component of water resource management, directly influencing irrigation optimization, water allocation strategies, and flood prevention infrastructure [1], [2]. However, the inherent complexity of atmospheric dynamics results in unique spatial and temporal rainfall characteristics that remain challenging to predict [3]. These variations are driven by regional



Copyright © 2026 Nastasia F. Margini, at al. This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Allows readers to read, download, copy, distribute, print, search, or link to the full texts of its articles and allow readers to use them for any other lawful purpose.

geographic factors, including mountainous terrain and island orientations, which lead to non-uniform distribution patterns. Given the stochastic nature of precipitation series, hydrological planning requires high-precision modeling and long-term historical data to enhance simulation reliability [1], [4]-[6]. Current research paradigms have shifted from traditional physical models toward integrating meteorological satellite data, fuzzy logic techniques, and weather radar for advanced quantitative precipitation estimation [2], [7], [8].

Rainfall prediction is a complex task, and various techniques have been used to predict rainfall accurately. Machine learning-based techniques, physical-based models, satellite-based techniques, and ensemble-based techniques are some of the recent advancements in rainfall prediction. Statistical techniques are also widely used for rainfall prediction. These techniques use historical rainfall data to estimate future rainfall. Some common statistical techniques include time series analysis, autoregressive models, and regression analysis [1], [9], [10].

The relationship between rainfall and weather elements are incorporated in the hydrological cycle. Several previous studies on weather elements that influence rainfall have been conducted. Fadholi studied the relationship between air humidity and temperature on rainfall using Regression Analysis [11]. In this research, it was concluded that rainfall forecasting in Ternate obtained good results when only using air humidity predictors with a regression model. Apart from that, Hossain et al also gave suggestions that the rain forecasting they carried out using the Non-Linear Regression Method obtained satisfactory results for the Australian Capital Territory [12]. Based on several previous studies on the application of rainfall forecasting which have shown a high level of reliability in several areas, this research will compare several methods for forecasting rainfall in the Wonorejo watershed as an effort to increase the usefulness of the Wonorejo Reservoir, Indonesia.

The Wonorejo watershed located in Tulungagung Regency, holds a crucial role as a primary supplier of raw water to the Brantas river, providing irrigation covering an area of 1470 hectares, a 6.02 MW hydropower plant, and flood control for the Tulungagung area and its vicinity. However, according to research of Indarto, findings indicate a notable decline in the quantity of water flow (discharge) in the Wonorejo river over the past 10 years [13]. This reduction encompasses monthly and daily flows as well as extreme flows. The decrease is also evident in the count and peaks of high and low flows. Over time, the current age of the Wonorejo dam is 22 years, during which an effective storage decrease of 10% from the original capacity has been observed. Consequently, it is essential to conduct a study on rainfall forecasting using historical rainfall and humidity data to predict future discharge magnitudes in the Wonorejo watershed. This discharge prediction study can also serve as a reference for relevant authorities involved in the operation and maintenance of the Wonorejo reservoir [14].

In this research, predictive modeling of the amount of rainfall in the next two years was carried out on humidity as an influencing variable. This research took the research location in Tulungagung Regency (Figure 1). This research is intended to be able to predict the amount of rainfall in the future at the Wonorejo Reservoir, which currently has not been done properly in order to provide an overview of the inflow discharge that can be utilized. So that the Wonorejo Reservoir can get maximum benefits from the available water. The model error validation values will be analyzed using Non-linear Regression, Model Decomposition, Regression Model, and Winter model.

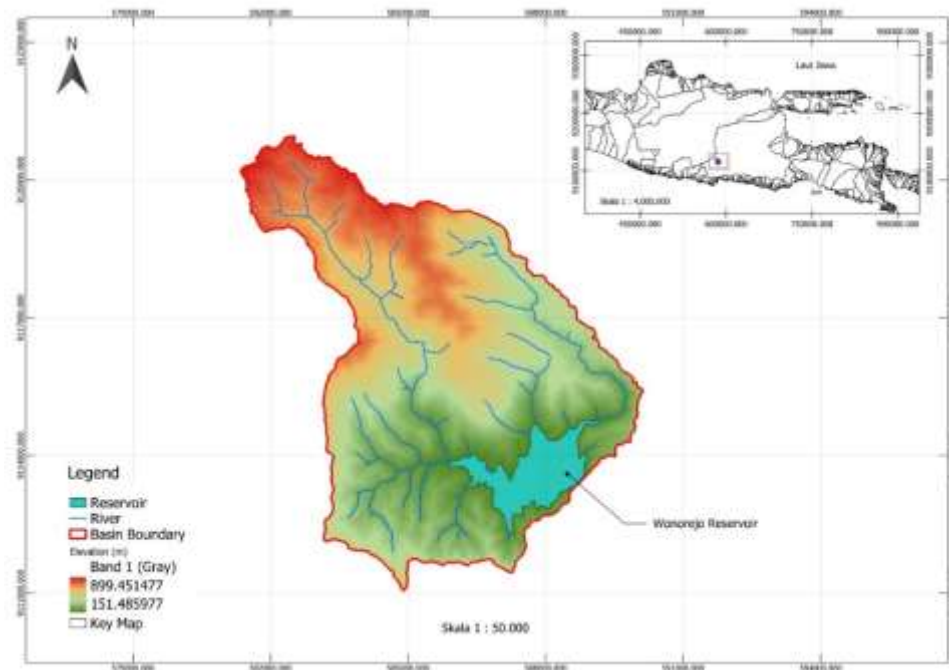


Figure 1. Map of Wonorejo Watershed, East Java, Indonesia (source : author's analysis)

2. Research Method

2.1 Data Collection and Preparation

Data were obtained from meteorological stations in various places within the Wonorejo Reservoir catchment area with the assistance of BBWS Brantas. The features of the data collected include rainfall and air humidity data. The length of data used is 21 years from January 1998 – December 2018. There are three (3) rain stations with the names Bendungan Station, Sumber Pandan Station, and Bagong Station which are located around the Wonorejo Reservoir location, but the data availability is incomplete. So the Bendungan Rain Station was chosen as the input data in this forecasting analysis. In the process of developing the model the rain data and air humidity data are divided into two (2), namely training and testing. The training data starts from January 1998 – December 2016 with 228 rainfall data and 228 air humidity data. Meanwhile, test data was used from January 2017 to December 2018. This data was used to form a regression equation. Monthly air humidity data in 2008 was used to predict total rainfall. Meanwhile, total monthly rainfall data in 2008 was used as a comparison in verifying the predicted results of total monthly rainfall.

The research methodology (Figure 2) follows these sequential steps:

1. Problem identification, objective formulation, and data acquisition
2. Exploratory data analysis and statistical testing of rainfall and humidity series
3. Rainfall prediction using the Decomposition method
4. Humidity prediction using the Decomposition method
5. Causal regression analysis incorporating lagged variables
6. Nonlinear Regression analysis employing humidity as predictor
7. Model evaluation using RMSE and comparative analysis

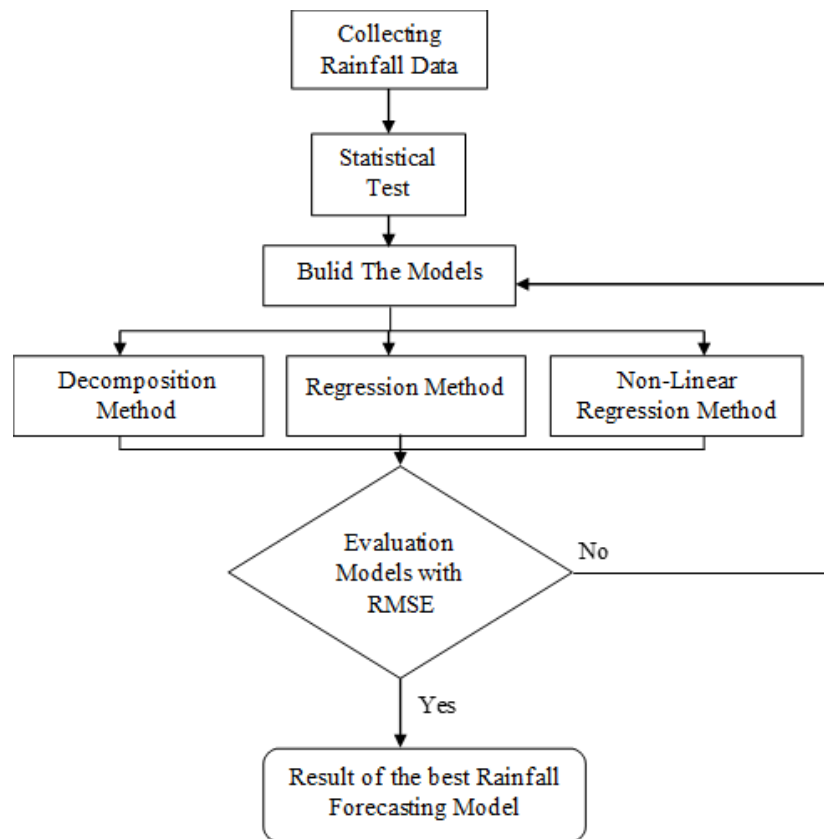


Figure 2. Flow chart research (source : author's analysis)

- **Decomposition Method**

The decomposition method is a method used if the data we have has the following four types of patterns, namely trends, cyclical, seasonal, and irregular [15]. In bulk data forecasting using air humidity data, both of which are seasonal data types with characteristics having factors related to periodic fluctuations with a fixed period. So that it is expected to form an appropriate model.

The decomposition method assumes that the data is structured as follows:

$$\begin{aligned} \text{data} &= \text{pattern} + \text{error} \\ &= f(\text{trend, cycle, seasonal}) + \text{error} \end{aligned} \quad (1)$$

Besides the pattern component, there are also random error components, whereas the general mathematical components of the decomposition model are:

$$X_t = f(I_t, T_t, C_t, E_t) \quad (2)$$

where:

- X_t = actual data in period t
- I_t = seasonal index in period t
- T_t = the trend component in the period t
- C_t = cycle component in period t
- E_t = seasonal component in period t

the decomposition method can be divided into 2, namely the simple average Decomposition method which assumes the additive and multiplicative models. However, this paper uses an additive model because the data types are only seasonal without a trend. The formula used is as follows:

$$X_t = (I_t + T_t + C_t) + E_t \quad (3)$$

- **Decomposition Method**

Regression prediction methods used are Linear Regression and Non-Linear Regression. For linear Regression, this is done by forming an equation used to simulate monthly total rainfall predictions using more than one independent variable. While, Nonlinear regression is a form of regression analysis in

which observational data are modeled by a function, which is a nonlinear combination of the model parameters and depends on one or more independent variables.

The general equation for this method is as follows:

$$Y = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 X_{2i} + e \quad (4)$$

where:

Y = response

X_k = k^{th} predictor

β_k = k^{th} population regression coefficient

e = error

In this paper the independent variable or predictor is air humidity, while the dependent variable or response variable is monthly rainfall. The process of making predictions there are two stages, the first stage is making a regression equation for each month based on the same month for 19 years from 1998 - 2016 and the second stage is predicting the total lunar month by assigning the value of the predictor variable (predictor) to the regression equation created. In inputting predictor data on each regression equation the time difference (time lag) of 1 (one) month to 12 months and predictor is used.

2.2 Model Evaluation: Root Mean Square Error

This method is used to determine the magnitude of deviations that occur between the predicted total rainfall value compared to the actual total rainfall that occurred over a period of 24 months or two (2) years. From this value can be analyzed the prediction of total rainfall with predictors of air humidity and then look for large or small deviation values [16]. It should be noted that for the validity of the forecast results the greater the RMSE value, the farther the estimated total monthly rainfall value of the actual total rainfall and the smaller the RMSE value the better the rainfall prediction. Because the level of errors that can be minimized can increase the level of accuracy of forecasts. From each analysis method used, then the RMSE value is searched and compared as a comparison ratio of the analysis results with the Winter's RMSE method. the smallest results that determine the analysis that is close to the actual conditions. The formulation is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - O_i)^2} \quad (5)$$

where:

N = Number of samples

f = Forecasts

O = Observed values

3. Results and Discussions

3.1 Exploratory Data Analysis

In an analysis of upcoming data forecasting, the first thing to do is to explore the available data. This is done to find out what method is most appropriate for the analysis process. From the available data, rainfall and humidity data in Wonorejo Watershed, Tulungagung Regency, starting from January 1998 to December 2018, obtained a large standard deviation value of 140.51, indicating that rainfall in the Tulungagung Regency, especially Wonorejo Watershed is very diverse. January-February-March is the peak of the rainy season while the peak of the dry season occurs in July-August-September.

Rainfall values are in the average range of 211.3 mm. Whereas the range of humidity data for the years 1998-2018 was 79% with maximum humidity occurring in February. Figures 1 and 2 provide information that the humidity has the same grid line pattern as the rainfall line diagram that is decreasing in the dry season (June-July-August) and rising again in the rainy month (November-December-January). Thus, it is clear there are similarities in the distribution pattern of rainfall and humidity. From Figures 3 and 4 also shown the diversity of rainfall and humidity between months that are not homogeneous.

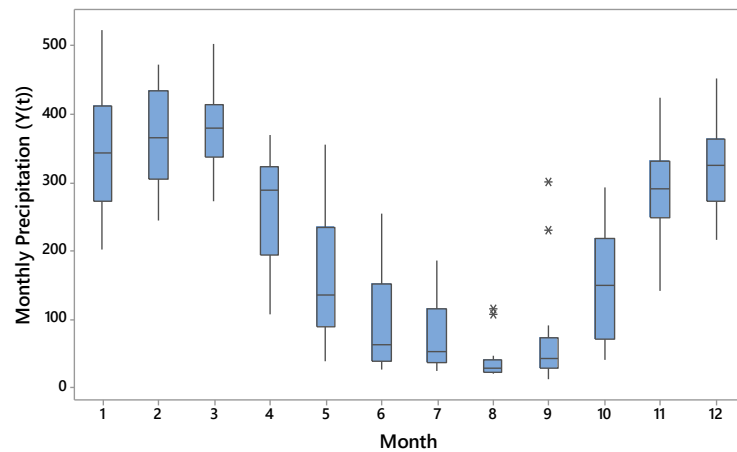


Figure 3. Box diagram of monthly rainfall line for 1998-2018 (source : author's analysis)

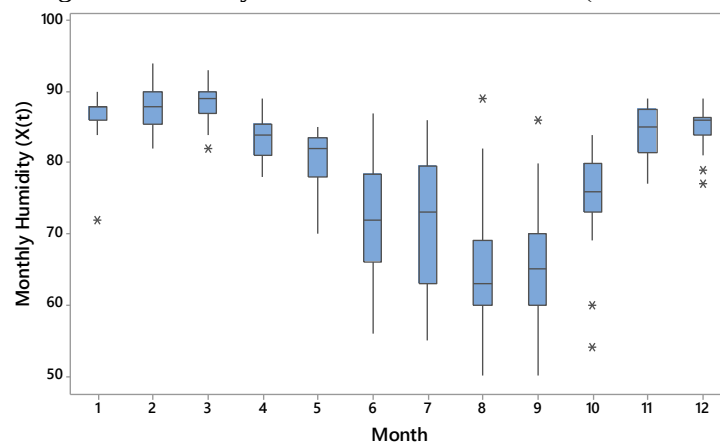


Figure 4. Box Diagram of monthly humidity line for 1998-2018 (source : author's analysis)

In January, the rainfall line box diagram has a greater bandwidth than the other months while the largest box humidity in July means that rainfall and humidity are more diverse than other months. According to Kartasaputra, high humidity values cause a great potential for clouds to form and large amounts of rainfall. This can be seen in Figure 5 namely the higher the humidity the air will also increase the rainfall. The correlation value between rainfall and humidity is 0.72 which means that between rainfall and humidity it has an interrelated relationship. This is also supported by the correlation p-value of 0.816.

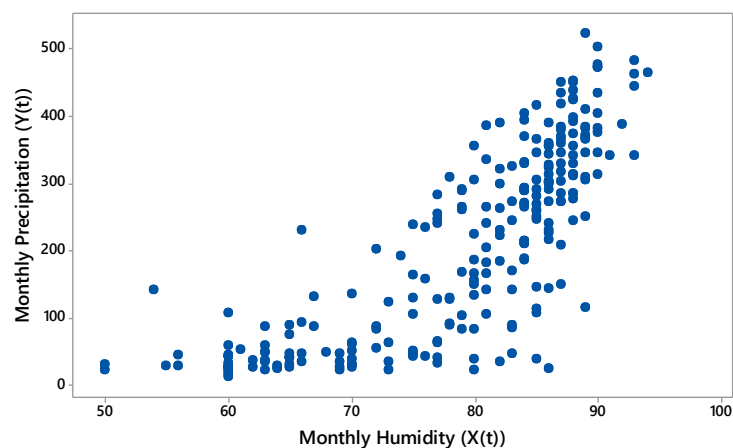


Figure 5. Scatterplot of Monthly Rainfall and Humidity (source : author's analysis)

This shows that humidity and monthly rainfall have a strong correlation. So, it can be used as a builder of future rainfall prediction models. If you look in more detail about the distribution pattern of monthly rainfall data and humidity data has a rhythmic pattern as shown in Figure 6 below:

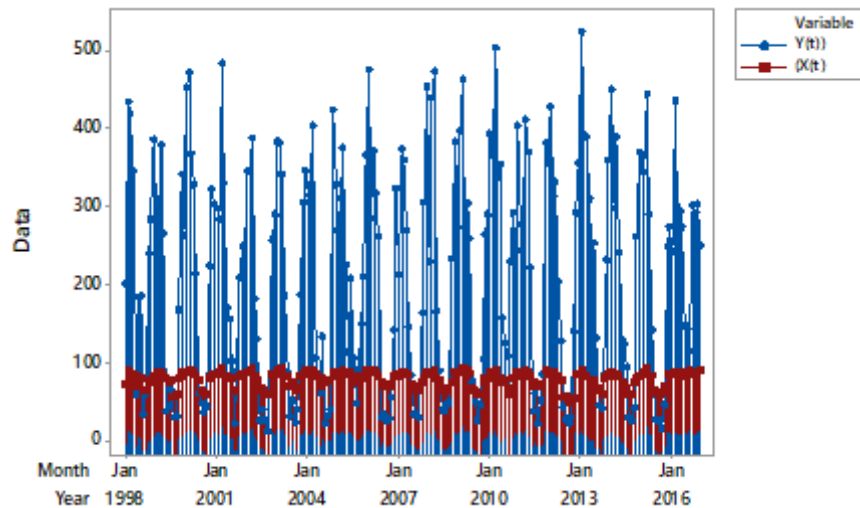


Figure 6. Time Series of Monthly Rainfall $Y(t)$ And Humidity $X(t)$ (source : author's analysis)

From Figure 6, the highest rainfall value was found in January 2013 with 522.3 mm and the lowest in September 2002 reaching 12.3 mm. While the highest humidity in March 2015 was 93% and the lowest in September 2012 reached 50%.

3.2 Model Development and Comparison

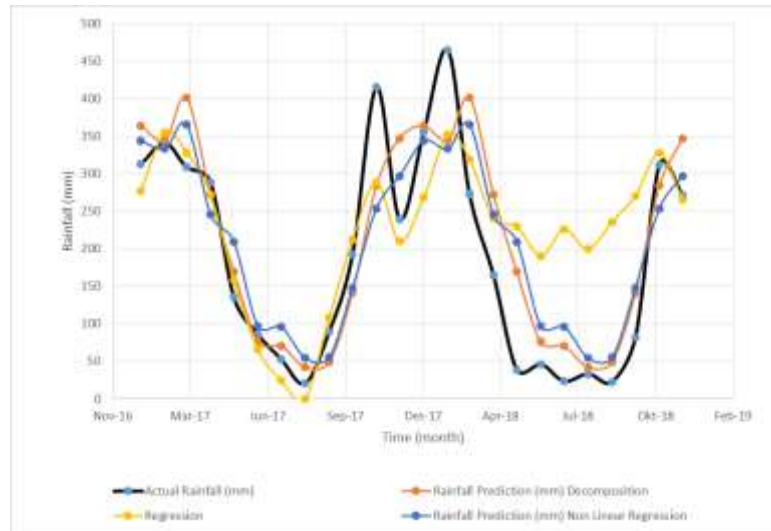
The rainfall and humidity relationship model that was built uses several methods to find methods that are more suitable for the data. From the rainfall and humidity data of the Wonorejo watershed, Tulungagung Regency was analyzed using the Decomposition Method, Regression Method, and Non-Linear Regression Method. From the four methods we get the results as shown in Table 1 and visualized in Figure 7.

Table 1. Rainfall Prediction Results using various Methods

Month	Actual Rainfall (mm)	Rainfall Prediction (mm)		
		Decomposition	Regression	Non-Linear Regression
Jan-17	313	364	277	344
Feb-17	341	344	355	333
Mar-17	309	401	328	366
Apr-17	289	272	273	246
Mei-17	136	169	158	210
Jun-17	86	76	66	97
Jul-17	53	71	25	96
Agu-17	21	42	0	54
Sep-17	90	49	109	55
Okt-17	192	142	212	148
Nov-17	415	284	289	254
Des-17	240	347	210	297
Jan-18	356	364	269	344
Feb-18	465	344	352	333
Mar-18	273	401	320	366
Apr-18	164	272	240	246
Mei-18	39	169	229	210
Jun-18	47	76	190	97
Jul-18	23	71	226	96

Month	Actual Rainfall (mm)	Rainfall Prediction (mm)		
		Decomposition	Regression	Non-Linear Regression
Agu-18	33	42	200	54
Sep-18	23	49	236	55
Okt-18	82	142	271	148
Nov-18	312	284	328	254
Des-18	271	347	265	297

(source : author's analysis)

**Figure 7.** Scatterplot of Actual Rainfall VS Model Rainfall (source : author's analysis)

The next model built is a causal regression and non-linear regression model. Both models use variable humidity as a determining factor, in addition to the rainfall data itself. In the regression model there are several variables that have been shown to have the greatest influence on monthly rainfall events, namely $(X_{(t)}, Y_{(t-1)}, Y_{(t-12)}, \text{ and } X_{(t-1)})$. Then the Regression Model equation as appropriate, as follows:

$$Y_t = -255.2 + 7.899X_t + 0.3347Y_{(t-1)} + 0.3007Y_{(t-12)} - 3.683X_{(t-1)} \quad (6)$$

where,

Y_t = Prediction of rainfall

X = humidity

$Y_{(t-1)}$ = Rainfall of $(t - 1)$ month

$Y_{(t-12)}$ = Rainfall of $(t - 12)$ month

$X_{(t-1)}$ = Humidity of $(t - 1)$ month

The equation obtained from the Non-Linear Regression Model is as follows:

$$Y(t) = \exp(-0.759342 + 0.0749715 \times X_{(t)}) \quad (7)$$

From the equation, it is obtained a pattern approaching the existing condition (as shown in Figure 8. There is a similarity of the pattern of the model formed between the Non-Linear Regression Model and the Decomposition Model, both of them cannot represent the amount of rainfall at the 2 highest peaks, but can describe the pattern well and almost coincide, especially in the descending part of the graph.

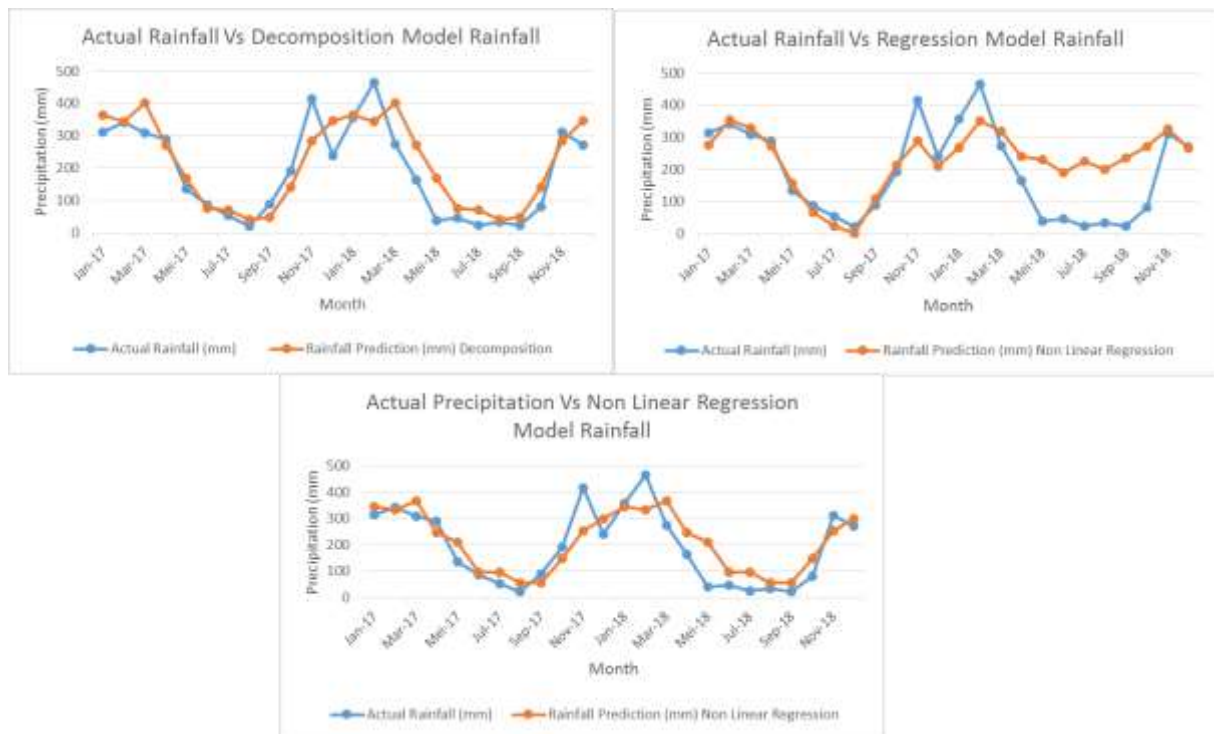


Figure 8. Line chart of actual rainfall and model rainfall on all three models (source : author's analysis)

3.3 Model Validation and Comparison

After building the four models, validation is then performed to see which method is able to better predict the predicted rainfall results. Comprehensive validation metrics were computed for each modeling approach, as presented in Table 2. The Nonlinear Regression method achieved the lowest RMSE (72.71), followed closely by Decomposition (73.78), with Multiple Linear Regression yielding the highest RMSE (103.57). These results establish Nonlinear Regression as the superior method for this application. These results mention that the Non-Linear Regression Method ranks number one for methods with smaller error values than other methods. Decomposition is in order number two with an error validation value that approaches the Non-Linear Regression value and followed by the Regression Method and Winter Method as in table 2.

Table 2. Validation Number of all model method

Error Validation Method	Rainfall Prediction		
	Causal Regression	Regression Non-Linear	Decomposition
MSE	10726.29	5286.58	5006.66
RMSE	103.57	72.71	73.78
MAPE	156.10	73.73	56.96
SMAPE	24365.75	5435.60	3244.95
MAE	75.92	58.94	56.11
Ratio of RMSE with Causal Regression Model	1.00	0.70	0.68

(source : author's analysis)

After building the three models, validation was then carried out to assess which method was more appropriate for predicting rainfall prediction results. These results indicate that the Non-Linear Regression Method is ranked number one for the method whose error value is smaller than other methods with a value of 72.71. Decomposition is in second place with a validation error value close to the Non-Linear Regression value of 73.78 and continued with the Regression Method with a value of 103.57. Figure 9 shows that in this study the ratio of RMSE to Non-Linear Regression is the smallest.

This is also supported by [18] which states that Non-Linear Regression is a simple but accurate enough way to predict hydrological conditions. This is also in line with [19] that humidity and rainfall have a relationship especially when used to predict monthly total rainfall using nonlinear regression.

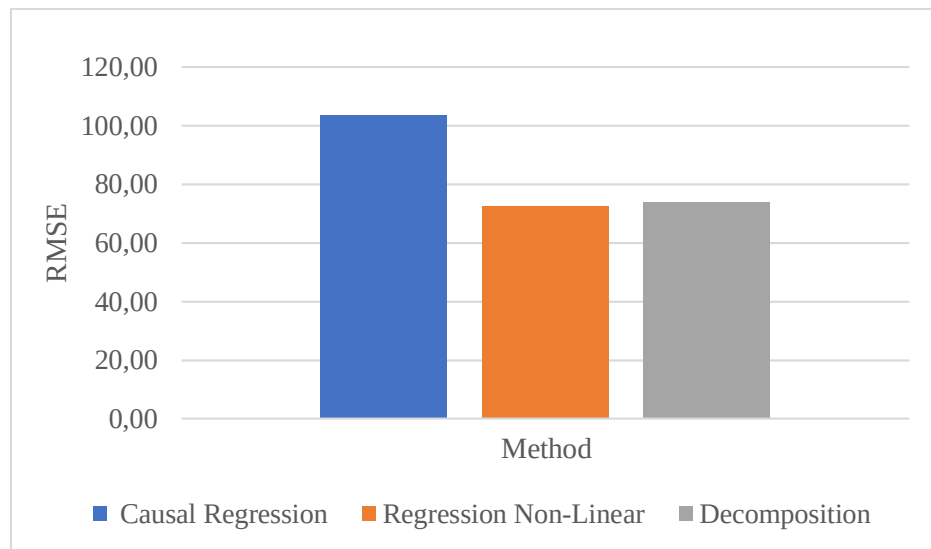


Figure 9. Bar Chart of Ratio RMSE of Decomposition, Regression, Non-Linear Regression (source : author's analysis)

The substantial error reduction achieved by Nonlinear Regression (30% improvement over Linear Regression) underscores the importance of capturing nonlinear dynamics in rainfall-humidity relationships. However, the remaining prediction errors suggest that additional meteorological variables may further enhance model performance. Temperature, wind direction, and wind speed represent potentially valuable predictors for future model development.

4. Conclusion

In this study, the capabilities of Decomposition, Regression and Non-Linear Regression models for predicting the rainfall were evaluated. Different types of input were considered as model inputs. The results suggest that the choice of input to the models has a significant impact on forecasting accuracy. The model that has the smallest error validation value is Non-linear Regression, followed by Model Decomposition, Regression Model. For each number validation errors can still be minimized. This error rate is still relatively large due to the influence of factors other than humidity which also influence rainfall. Other data that may be included such as Temperature, Wind direction, and wind speed. Related to the future can be carried out further studies on the relationship of the components mentioned above to the prediction of rainfall.

Acknowledgements

The authors express their appreciation to Bapak Dedy Dwi Prastyo and Bapak Suhartono, from the Sepuluh Nopember Institute of Technology for their guidance, BBWS Brantas and BMKG Surabaya for their assistance in data collection.

References

- [1] D. C. Siregar, "Simulasi Prediksi Total Hujan Bulanan di Tanjungpinang (Studi Kasus Tahun 2017)," *J. Stat. dan Apl.*, vol. 2, no. 2, pp. 1–7, 2018, doi: 10.21009/jsa.02201.
- [2] S. S. Chinchorkar, G. R. Patel, and F. G. Sayyad, "Development of monsoon model for long range forecast rainfall explored for Anand (Gujarat-India)," *Int. J. Water Resour. Environ. Eng.*, vol. 4, no. 11, pp. 322–326, 2012, doi: 10.5897/IJWREE11.097.
- [3] E. G. Union, "Interactive comment on ' Downscaled Rainfall Prediction Model (DRPM) using a Unit Disaggregation Curve (UDC)' by S . Tantanee et al .," 2005.
- [4] A. Fadholi, A., Persamaan regresi prediksi curah hujan bulanan menggunakan data suhu dan

- kelembapan udara di Ternate, 13 (1) 2013, pp. 7–16 [Online]. Available: <https://ejournal.unisba.ac.id/index.php/statistika/article/view/1068>.
- [5] Hettiarachchi, P., Hall, M. J., Minns, A. W., & Hettiarachchi, A. P., The extrapolation of artificial neural networks for the modelling of rainfall-runoff relationships, *Journal of Hydroinformatics*, 07(4), 2005, pp. 291-296.
 - [6] Sloughter, J. M. L., Raftery, A. E., Gneiting, T., & Fraley, C., Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Monthly Weather Review*, 135(9), 2007, pp. 3209–3220. <https://doi.org/10.1175/MWR3441.1>
 - [7] Wijayarathne, D., Coulibaly, P., Boodoo, S., & Sills, D., Use of Radar Quantitative Precipitation Estimates (QPEs) for Improved Hydrological Model Calibration and Flood Forecasting, *Journal of Hydrometeorology*, 22(8), 2021, pp. 2033–2053. <https://doi.org/10.1175/JHM-D-20-0267.1>
 - [8] Huria, A., Kamboj, G., Kukreti, D., & Rawat, J., A Relative Analysis of Modern ML Methods for Rainfall Prediction, *Proceedings of the Advancement in Electronics & Communication Engineering*, 2022. <http://dx.doi.org/10.2139/ssrn.4159466>
 - [9] Wang, D., Huo, Z., Miao, P., Tian, X., Comparison of Machine Learning Models to Predict Lake Area in an Arid Area, *Remote Sensing*, 15(17), 2023, pp.4153. <https://doi.org/10.3390/rs15174153>
 - [10] Wang, J., Wong, R. K. W., Jun, M., Schumacher, C., Saravanan, R., & Sun, C., Statistical and machine learning methods applied to the prediction of different tropical rainfall types. *Environ. Res. Commun*, 3, 2021.
 - [11] Laddimath, R. S., Patil, N. S., and Hooli, S., Downscaling of Precipitation Data from GCM outputs using Artificial Neural Network for Bhima basin, *International Journal of Applied Environmental Sciences*, 10 (4), 2015, pp. 1493-1508.
 - [12] Bishara, A. J., Hittner, J. B., Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality, *Educ Psychol Meas*, 75(5), 2015, pp. 785-804. doi: 10.1177/0013164414557639.
 - [13] Santoso, S., *Panduan Lengkap Menguasai Statistik dengan SPSS 17*, PT Elex Media Komputindo, Jakarta, 2009.
 - [14] Margini, N. F., Damarnegara, S., Anwar, N., Yusop., Z, Water Allocation in Multi-Purpose and Multi-Year Reservoir using Ant Colony Optimization. *Sustainable Water Resources Management*, Vol. 10, 2024. doi : 10.1007/s40899-024-01093-4.
 - [15] Indarto, I. Review of the Application of Hydrological Change Indicators in the Wonorejo Watershed, *Journal of Agricultural Engineering*, 6 (3), 2018, , pp. 241-248.
 - [16] Wilks, D.S., *Statistical Methods in the Atmospheric Sciences*, Academic Press, San Diego, pp. 467, 1995.
 - [17] Kartasapoetra, A. G., et. all., *Klimatologi: Pengaruh Iklim Terhadap Tanah dan Tanaman*, Bumi Aksara, Jakarta, 2012.
 - [18] Rezaeianzadeh, M., Tabari, H., Arabi Yazdi, A., Isik, S., and Kalin, L., Flood flow forecasting using ANN, ANFIS and regression models, *Neural Comput. Appl.*, 25 (1), pp. 25–37, 2014, doi: 10.1007/s00521-013-1443-6.
 - [19] M. Rezaeianzadeh, H. Tabari, A. Arabi Yazdi, S. Isik, and L. Kalin, “Flood flow forecasting using ANN, ANFIS and regression models,” *Neural Comput. Appl.*, vol. 25, no. 1, pp. 25–37, 2014, doi: 10.1007/s00521-013-1443-6.