BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1185–1198
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1185-1198

SPATIAL INTERPOLATION OF RAINFALL DATA USING COKRIGING AND RECURRENT NEURAL NETWORKS FOR HYDROLOGICAL APPLICATIONS IN SURABAYA, INDONESIA

Danang Ariyanto [1*], A'yunin Sofro [2], Riskyana Dewi Intan Puspitasari [3], Riska Wahyu Romadhonia [4], Hernando Ombao [5]

[1,2]*Actuarial Science Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya*
[3]*Data Science Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya*
[4]*Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya*
Jln. Ketintang Gedung D1, Surabaya, 60231, Indonesia

[5] *Statistics Study Program, Computing, Electrical, and Mathematical Sciences and Engineering Division,*
King Abdullah University of Science and Technology
Thuwal, 23955 – 6900, Saudi Arabia

*Corresponding author's e-mail: * **danangariyanto@unesa.ac.id***

Article Info

Article History:

Received: 5th May 2025
Revised: 11th August 2025
Accepted: 30th September 2025
Available online: 26th January 2026

Keywords:

Interpolation;
Rainfall;
Recurrent Neural Network;
Spatial;
Surabaya.

ABSTRACT

Urban hydrological challenges, such as flooding and water resource management, require accurate rainfall data to support sustainable development. This study investigates the use of Recurrent Neural Networks (RNN) for spatial interpolation of monthly rainfall data across 31 districts in Surabaya, Indonesia, and compares its performance with the geostatistical method Cokriging. Elevation data were incorporated as an additional variable to account for geographical variability. The dataset was divided into training (26 locations) and testing (5 locations) subsets, with testing locations treated as missing data points to simulate real-world conditions. The results show that the RNN-based interpolation method achieved progressively lower Root Mean Square Error (RMSE) values from January (48.65) to April (13.78), indicating higher accuracy compared to the Cokriging method. These findings underscore the potential of RNN in addressing data gaps and spatial variability, offering robust solutions for hydrological applications in urban environments. This approach not only supports flood risk mitigation strategies but also contributes to optimizing drainage systems and water resource planning. Further research is recommended to incorporate additional environmental variables and extend the application to broader spatial and temporal contexts.

How to cite this article:

D. Ariyanto, A. Sofro, R. D. I. Puspitasari, R. W. Romadhonia, and H. Ombao, "SPATIAL INTERPOLATION OF RAINFALL DATA USING COKRIGING AND RECURRENT NEURAL NETWORKS FOR HYDROLOGICAL APPLICATIONS IN SURABAYA, INDONESIA", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1185-1198, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Rainfall is an important parameter in hydrology, climatology, and water resources management studies [1], [2]. Accurate and complete rainfall data is needed for various applications, including flood planning and management, irrigation, and regional spatial planning [3], [4]. In the city of Surabaya, one of the largest cities in Indonesia with a high level of urbanization, reliable rainfall data is vital to support sustainable development, disaster mitigation, and effective hydrological planning [5]. However, collecting rainfall data in the field often faces challenges, as rain gauges spread across various locations can experience damage, data loss, or operational disruptions due to extreme weather and other technical issues. These limitations lead to gaps in rainfall data, which hinder accurate hydrological analysis and decision-making processes [6], [7].

Missing data is a value that is not available on a particular object, which can be caused by data corruption, errors in recording values, damage to measuring instruments, or conditions that do not allow measurement [8]. Missing data categories are divided into three, namely Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) [9]. To address this issue, interpolation methods are often used to estimate values at unknown locations by leveraging surrounding data points [10]. Spatial interpolation methods, such as Inverse Distance Weight (IDW) and kriging, are used to estimate the value of an unknown location point by utilizing values from other points in the same area [11]. This method has been used widely, but is often unable to capture complex spatial patterns in rainfall data [12]. Rainfall interpolation is the process of estimating rainfall values in locations that do not have direct measurement data, based on data from known surrounding locations [13], [14].

Interpolation methods can generally be categorized into two main approaches, namely deterministic methods and geostatistical methods. Deterministic methods, such as Inverse Distance Weighting (IDW) and Radial Basis Function (RBF), use direct mathematical relationships between data points to generate new value estimates [15]. Meanwhile, geostatistical methods such as Ordinary Kriging (OK) and Universal Kriging (UK) utilize statistical models to consider spatial correlations in the data used [16].

Several studies have compared the effectiveness of interpolation methods in various applications. For example, a study in Portugal found that Empirical Bayesian Kriging Regression (EBKR) provided more accurate results in rainfall estimation compared to other methods [17]. Meanwhile, in Canada, the Spatio-Temporal Kriging (STK) method was superior to Thiessen Polygons (TP) and IDW in modeling maximum temperature [15].

With the advancement of computing technology, various new approaches have been developed to improve the accuracy of spatial interpolation. One of the prominent innovations is the combination of interpolation methods with machine learning, which allows for more complex and adaptive spatial modeling [16]. In addition, hybrid methods such as IDW-RBF Neural Network (IDW-RBFNN) have been introduced to overcome the limitations of conventional methods by combining the strengths of deterministic approaches and artificial neural networks [18].

In recent years, advancements in computing technology and machine learning algorithms have enabled more accurate data interpolation [19]. Artificial neural network (ANN) methods, which fall into the deep learning category, such as Deep Feedforward Neural Networks, Convolutional Neural Networks, and Recurrent Neural Networks, offer a more adaptive approach and are able to capture non-linear relationships in data [20]. Among these, Recurrent Neural Networks (RNN) are particularly effective in processing sequential data, such as rainfall time series, due to their ability to incorporate past information and model temporal dependencies [21]. This makes RNN a promising method for improving spatial interpolation accuracy, particularly in the hydrological domain, where precise rainfall data is crucial for flood modelling, watershed management, and water resource planning [22].

The city of Surabaya, with its predominantly lowland topography, is highly vulnerable to flooding, particularly during the rainy season. Accurate rainfall data plays a critical role in hydrological applications, such as flood risk mapping, drainage system optimization, and early warning systems. However, the lack of complete rainfall data often hampers the ability to develop reliable hydrological models [23], [24]. By utilizing historical rainfall data and applying an RNN for spatial interpolation, this study aims to provide more accurate rainfall estimates. These estimates can serve as a foundation for hydrological planning, enabling policymakers and urban planners to design more effective flood mitigation strategies and optimize water resource management in Surabaya.

2. RESEARCH METHODS

2.1 Data Source

Surabaya, one of Indonesia's largest metropolitan cities, was chosen as the study area due to its complex geographical characteristics and hydrological challenges. The city is predominantly a lowland area with some hilly regions in the southern part, making it particularly vulnerable to flooding during the rainy season. Additionally, the high level of urbanization and rapid infrastructure development often place significant pressure on its drainage systems and water resource management. With varied rainfall distribution across its 31 districts, Surabaya provides an ideal context for testing the effectiveness of the RNN-based spatial interpolation method in producing accurate rainfall estimates to support sustainable water management and flood risk mitigation [23], [5].

The dataset employed in this study was sourced from the Centre for Hydrometeorology and Remote Sensing (CHRS), accessible via https://chrsdata.eng.uci.edu. Monthly rainfall data (January 2024-April 2024) were gathered for 31 districts within Surabaya City, Indonesia, and supplemented with elevation data (in meters above sea level) to account for geographical variability. These data were systematically divided into two subsets: 26 locations for training the RNN model and 5 locations for testing, which were deliberately treated as missing data points to simulate real-world data gaps. The inclusion of elevation as an additional variable provides a more comprehensive perspective, capturing the influence of topographical variations on rainfall distribution. Surabaya's geographical characteristics, predominantly lowland interspersed with modest elevations in its southern regions, present a unique challenge for spatial interpolation. This structured approach aims to assess the effectiveness of RNN in providing accurate rainfall estimates and addressing critical gaps in hydrological.

2.2 Methods

Cokriging is a geostatistical interpolation technique that extends ordinary Kriging by incorporating one or more secondary variables to improve prediction accuracy. Unlike ordinary Kriging, which relies solely on the spatial autocorrelation of the primary variable, Cokriging also considers the cross-correlation between the primary and secondary variables. This makes it particularly useful in cases where the primary data are limited but correlated auxiliary data are available. In the context of rainfall estimation, elevation is often regarded as an important predictor, since topography plays a significant role in influencing rainfall patterns.

In this study, Cokriging was applied by combining rainfall observations as the primary variable and elevation as the secondary covariate. The interpolation was performed using the gstat package in R, which enables modelling of both variograms and cross-variograms as the basis for the estimation process. By integrating elevation into the interpolation, Cokriging was expected to provide a more reliable estimation of rainfall distribution in Surabaya compared to relying on rainfall data alone. This method served as the geostatistical benchmark for evaluating the performance of the RNN-based approach [25], [26].

Traditional spatial interpolation methods, such as IDW and Kriging, are widely used; they often struggle to capture complex non-linear spatial patterns, particularly in datasets with high variability or missing data [27]. Recent advancements in machine learning have introduced neural networks as a promising alternative for spatial interpolation. Unlike conventional methods, neural networks, particularly RNNs, excel in handling sequential and dependent data, making them suitable for datasets with temporal and spatial variability. By leveraging their ability to model non-linear relationships and incorporate contextual information through hidden states, RNN-based spatial interpolation offers a more adaptive and robust approach for estimating rainfall in areas without direct measurements. This study explores the use of RNN to address limitations in traditional methods, aiming to improve spatial interpolation accuracy for hydrological applications [4], [28].

Artificial neural networks are a form of simulation of the human brain's ability to learn using neurons and dendrites. The structure of an artificial neural network can change, namely in the form of changes in weight values. When training is performed on different inputs, the weight values change dynamically until a balanced value is reached. When this value is reached, it indicates that each input is connected to the expected output [29].

One method of artificial neural networks is the Feedforward Neural Network. Feedforward generally consists of an input layer, a hidden layer, and an output layer. The input layer is the place to enter information that will be processed by the artificial neural network. In many studies, this information is usually called features. This feature is the benchmark when making predictions. Features are passed through the hidden layers until they finally reach the output layer, where the result of the output layer is a prediction [30].

Feedforward neural networks have the assumption that the data is independent [31]. This assumption makes feedforward inappropriate for use in time series prediction cases such as sequence labelling. This is because context features play a big role in sequence labelling. Context features that are often used in the case of sequence labelling imply that their data is not independent. In other words, data depends on other data (dependent). Even so, feedforward can still be used for sequence labelling cases, but with limitations. For example, a study that combines context features (3 words before and 2 words after) with other features (current word). The context feature for this method is usually called context windows. The disadvantage of this method is that the context features are limited, and new parameters arise that must be thought about and looked for, namely, the number of words before and after. Therefore, an algorithm is needed that can accommodate the needs of time series prediction, such as sequence labelling [32], [33].

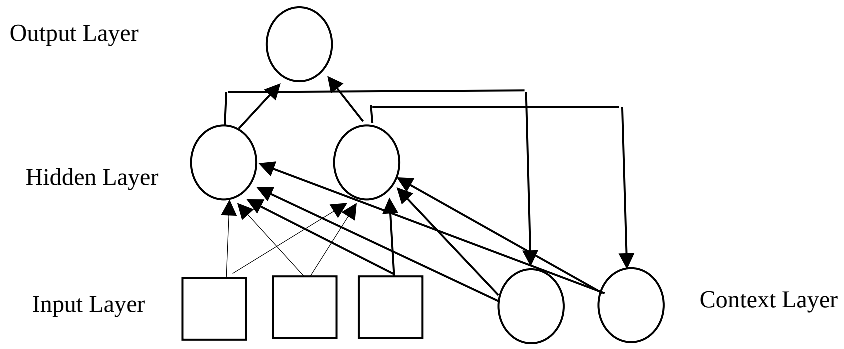


Figure 1. Elman RNN Network Topology

Recurrent Neural Networks (RNN) network architecture is like a feed-forward network architecture. The difference from the RNN method is that the RNN network architecture allows the flow to loop back to the previous layer, as shown in Fig. 1. The form shown in Fig. 1 is a simplification of the actual RNN form. If described, the complete form will be as shown in Fig. 2. The RNN network architecture consists of an input layer, a hidden layer, and an output layer. The number of timesteps is unlimited so that the RNN can meet the needs of sequence labelling, which allows complete input of sequential data directly into the RNN. Each hidden layer is associated with hidden layers included in the previous time step, which is the source of the benefits of the RNN. This hidden layer connection allows information to flow from previous data to the next data, so that the prediction process always considers past information. This is the contextual information required for sequence labelling [34].

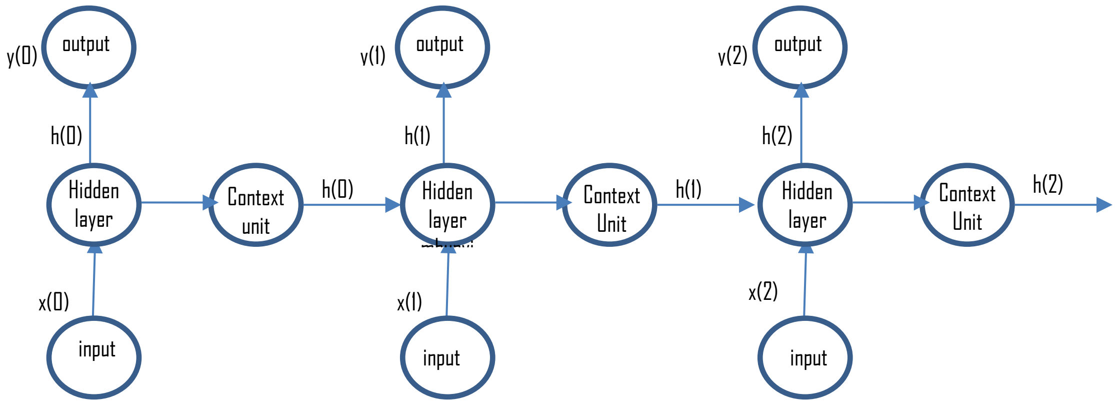


Figure 2. Unfolding Elman RNN

Fig. 2 illustrates the architecture of the Unfolding Elman Recurrent Neural Network, which is specifically designed to capture temporal dependencies in sequential data. This network comprises three primary layers: the input layer, the hidden layer, and the output layer. A distinctive feature of the Elman RNN is the inclusion of context units, which store information from previous time steps (hidden states) and integrate it with the current input [35]. This mechanism enables the network to “remember” patterns from prior data, making it particularly effective for processing sequential datasets, such as monthly rainfall data.

The unfolding process expands the network across multiple time steps, where each step connects the hidden layer's state from the previous step to the next, thereby generating more informed predictions. This architecture is particularly well-suited for hydrological modelling, as it allows for dynamic pattern recognition in rainfall data over time [36], [33]. The Elman RNN model consisted of an input layer with three predictor variables (latitude, longitude, elevation), one hidden layer with 5–25 neurons using the sigmoid activation function, and an output layer with a linear activation function. The model was trained with a learning rate of 0.01, a maximum of 1000 epochs, and an error tolerance of 0.0001. Training was conducted using the backpropagation through time algorithm with mean squared error as the loss function.

Learning with a Recurrent Neural Network is implemented by giving random initial values for all weights between the input-hidden layer and the hidden-output layer, learning rate, error tolerance, and maximum epoch. Each hidden layer unit z_j is added with input x_i multiplied by weight v_{ij} and combined with the context layer y_h multiplied by test weight u_{hj} and added with bias v_{0j} as follows in Eq. (1):

$$z_{inj} = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i v_{ij} + \sum_{h=1}^{m} \sum_{j=1}^{m} y_h u_{hj} + v_{0j} . \tag{1}$$

The activation function used is a binary sigmoid as in Eq. (2):

$$z_j = \frac{1}{1 + e^{-\left(\sum_{i=1}^{n} \sum_{j=1}^{m} x_i v_{ij} + \sum_{h=1}^{m} \sum_{j=1}^{m} y_h u_{hj} + v_{0j}\right)}}. \tag{2}$$

Then the output value will be sent to all output layers. The signal is sent to all neurons in the output layer and context units in the input layer. This step is done as many times as the number of hidden layers. Each output neuron sums the weighted input signals with Eq. (3):

$$y_k = w_{0k} + \sum_{j=1}^{m} z_j w_{jk} + \theta_k. \tag{3}$$

Then learning is carried out by improving the bias value by changing the weight value and changing the correlation value. This process is carried out continuously until it reaches the error tolerance. By implementing the RNN-based interpolation method, this study aims to produce reliable rainfall estimates, which are critical for hydrological applications such as flood risk modelling and water resource management in Surabaya.

Although this study employed a basic Elman RNN architecture, more advanced models such as LSTM and GRU were not explored. This choice was consistent with the relatively small size and short temporal span of the dataset, where a simple RNN was considered sufficient. Nevertheless, future research may examine LSTM or GRU to capture longer-term dependencies when larger datasets become available.

2.3 Data Preprocessing and Workflow

Prior to model implementation, the rainfall dataset was subjected to several preprocessing steps. First, missing values in the testing locations were deliberately introduced to simulate real-world conditions, while training locations contained complete data. Outlier detection was performed by examining rainfall distributions and removing values beyond three standard deviations from the monthly mean. All input variables (rainfall, latitude, longitude, and elevation) were normalized using min–max scaling to ensure consistent ranges for the neural network training process. For feature processing, elevation data (meters above sea level) were integrated as an additional predictor variable alongside spatial coordinates (latitude and longitude) and rainfall data. This integration allowed the model to capture the influence of topography on rainfall distribution.

The overall analysis workflow consisted of four main steps:

1. Data preparation: splitting rainfall data into training (26 locations) and testing (5 locations). Out of the 31 districts, 26 were randomly selected for training and 5 for testing, ensuring representation across different geographical areas of Surabaya. The testing locations were deliberately excluded from training to simulate missing data scenarios and to provide an unbiased evaluation of model performance.

2. Model development: implementing spatial interpolation using Cokriging and RNN. Cokriging was performed with the *gstat* package in R, while the RNN model was trained using the *keras* and *tensorflow* packages.
3. Model evaluation: predicted rainfall values at the testing locations were compared to actual observations. The Root Mean Square Error (RMSE) was calculated as the primary evaluation metric to assess model accuracy.
4. Result interpretation: RMSE values and interpolation maps from both methods were compared to evaluate the relative strengths of Cokriging and RNN for spatial rainfall estimation.

All analyses were performed using open-source software in the R programming language (version 4.3.1) within the RStudio environment. The following R packages were employed: *keras* and *tensorflow* for implementing the Recurrent Neural Network (RNN) models, *neuralnet* for network training and visualization, and *gstat* for geostatistical interpolation (Cokriging) [40], [41]. The experiments were run on a workstation equipped with an Intel Core i7 processor, 16 GB RAM, and a Windows 11 operating system [37], [38].

3. RESULTS AND DISCUSSION

3.1 Data Explanation

The city of Surabaya, as the capital of East Java Province, Indonesia, is located on the north coast of East Java Province, or precisely between 7° 9′- 7° 21′ South Latitude and 112° 36′ – 112° 54′ East Longitude. The city of Surabaya borders the Madura Strait to the north and east, Sidoarjo Regency to the south, and Gresik Regency to the west. The area of the City of Surabaya is 52,087 hectares, with a land area of 33,048 hectares or 63.45% and a sea area managed by the City Government of 19,039 hectares or 36.55%.

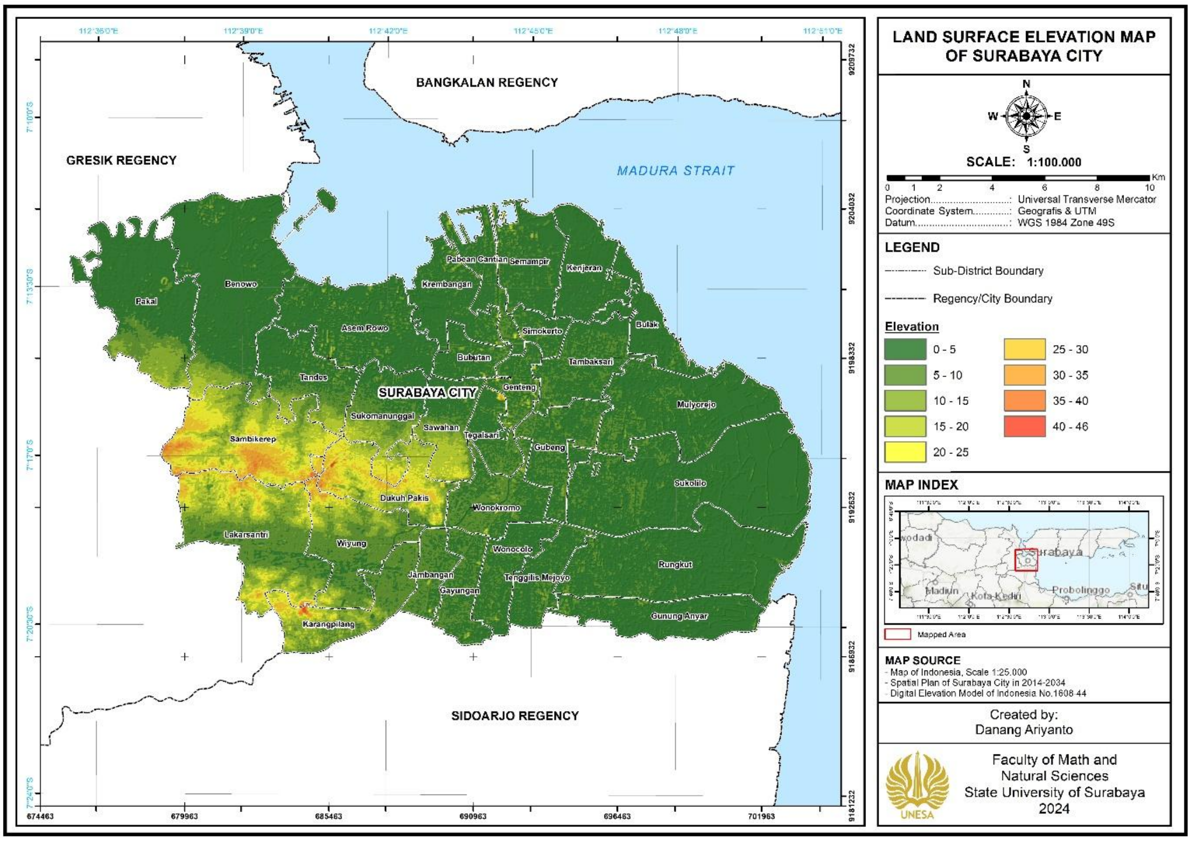


Figure 3. Land Surface Elevation Map of Surabaya City

Based on Fig. 3, the height in Surabaya is 0-46 meters above sea level (m asl). Topographically, 80% of Surabaya City is lowland, with a height of 3 - 6 meters above sea level (m above sea level), except in the southern part there are two sloping hills in the Lidah area (Lakarsantri District) and Gayungan with a height of 25 - 50 m above sea level. The distribution of monthly rainfall data in the City of Surabaya in 2024 is presented in Fig. 4. The data presented is in the form of rainfall values, average values, and maximum values of monthly rainfall in the City of Surabaya in 2024.

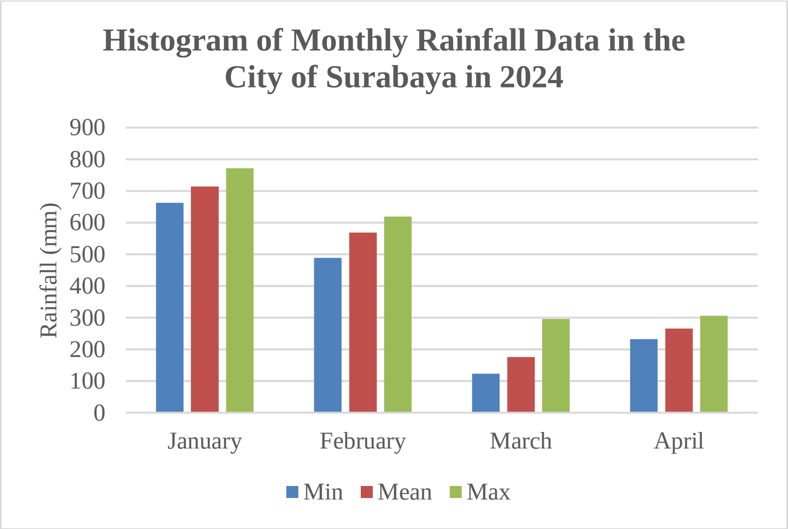


Figure 4. Histogram of Monthly Rainfall Data in the City of Surabaya in 2024

Based on Fig. 4, it can be concluded that in January-April 2024, the highest rainfall is in January and decreases until March, and begins to experience a slight increase in April. This fluctuation is in line with the pattern of the rainy season in the city of Surabaya, namely the dry season occurs in April-October, and the rainy season occurs in November-April.

3.2 Spatial Interpolation using Cokriging

The first step to perform interpolation using the cokriging method is to obtain the empirical semivariogram value and the matching process with the theoretical semivariogram by creating a crossvariogram plot for the Height and rainfall for January-April 2024 in Surabaya City. Based on the crossvariogram model formed, the cokriging analysis model for Surabaya City rainfall in January 2024 is as follows:

a. Height Model

$$\gamma(h) = 55.2\left(1 - exp\left(\frac{-3h^2}{6.795^2}\right)\right).$$

b. January Rainfall Model

$$\gamma(h) = \begin{cases} 1.104\left(1.5\left(\frac{h}{3.969}\right) - 0.5\left(\frac{h}{3.969}\right)^3\right), 0 < h \leq 3.969 \\ 1.104, \quad h > 3.969 \end{cases}.$$

c. Cokriging Model

$$\gamma(h) = \begin{cases} 1.098\left(1.5\left(\frac{h}{4.090}\right) - 0.5\left(\frac{h}{4.090}\right)^3\right), 0 < h \leq 4.090 \\ 1.098, \quad h > 4.090 \end{cases}.$$

After obtaining the best crossvariogram model, cokriging interpolation is then carried out. The results of the interpolation of rainfall for January-April 2024 using the cokriging method in Surabaya City can be seen in Fig. 5.

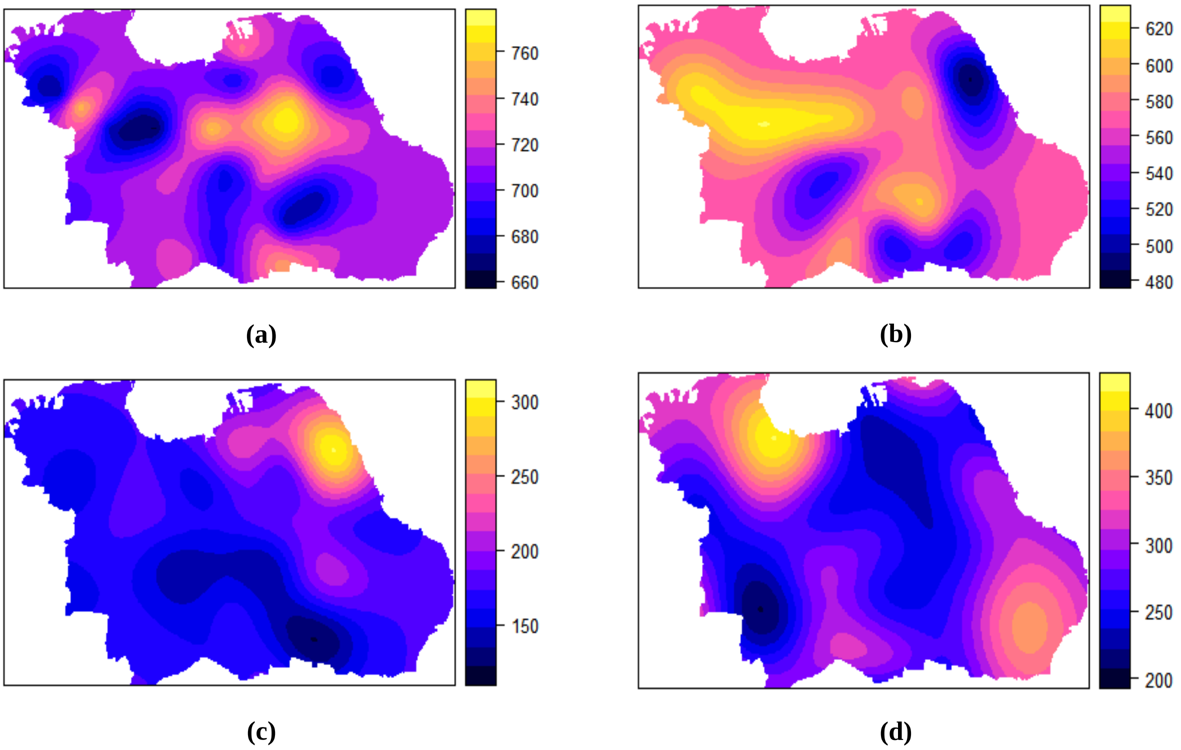


Figure 5. Cokriging Interpolation Result Map (a) January, (b) February, (c) March, and (d) April 2024

The map in Fig. 5 shows the results of the January-April 2024 rainfall estimation using cokriging. The contour map of the distribution of rainfall in the city of Surabaya is grouped with different color gradations. The highest rainfall interpolation in January is in the Ampelgading area, indicated by the lighter map contour color, meaning that the January rainfall interpolation results in the Genteng area are 760-780 mm/month. The lowest rainfall interpolation results are in the Kalipare area, indicated by the darker map contour color, meaning that the January rainfall interpolation results in the Pakis hamlet area and its surroundings range from 660-680 mm/month. The results of rainfall interpolation in the city of Surabaya in January vary from around 660-780 mm/month.

3.3 Spatial Interpolation Using RNN

The interpolation process is carried out using the RNN method, which uses latitude, longitude, height, and rainfall data from nearby areas as input. To determine the best network architecture, try several neurons in the hidden layer with the number of neurons limited to 5, 10, and 15 neurons in the hidden layer. The number of neurons in the output layer is one, namely in the form of the estimated value of the rainfall at the ith location. Root Mean Square Error (RMSE) is the square root of the mean squared error (MSE). Taking the root does not affect the relative ranks of models, but it yields a metric with the same units as y, which conveniently represents the typical or standard error for normally distributed errors [39]. The RMSE value of the RNN model based on the number of neurons in the hidden layer can be seen in Table 1.

Table 1. RMSE of RNN Model by Hidden Layer Neurons

Month	RMSE				
	5 neurons	10 neurons	15 neurons	20 neurons	25 neurons
January	31.8190	31.4110	**15.9801**	22.6978	17.3431
February	22.4209	20.2998	**19.0117**	20.2998	34.4188
March	23.6189	24.3695	20.0946	21.2365	**19.0572**
April	11.9375	8.4599	**7.0489**	8.1623	10.9900

Based on Table 1, the number of neurons used in the hidden layer in January, February, and April was 15 neurons because it had the smallest RMSE value. Interpolation of residual values in March has the smallest

RMSE value with a number of neurons in the hidden layer of 25 neurons. The RNN architecture for interpolation results for January rainfall can be seen in Fig. 6.

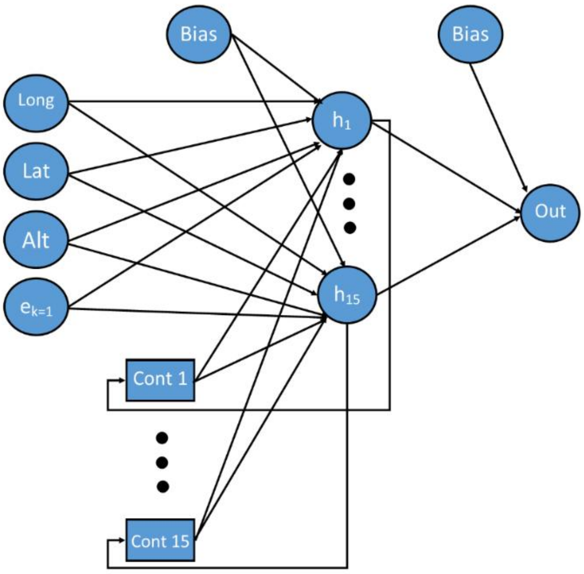


Figure 6. Residual RNN Architecture from Interpolation of Rainfall in January

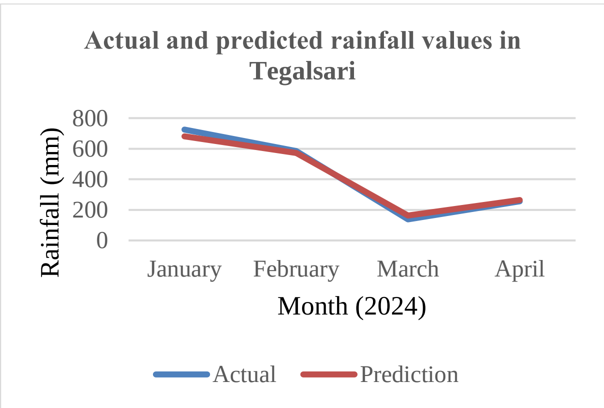
Based on Fig. 6, the number of neurons in the input layer is 4. The number of neurons in the hidden layer for each location follows the results in Table 1. The number of neurons in the context layer is the same as the number of neurons in the hidden layer. The number of neurons in the output layer is 1. The error target used is 0.0001 with a maximum iteration of 1000. In the interpolation of January rainfall data, an activation model is used from the input layer to the hidden layer using the sigmoid activation method. The activation function in the hidden layer to the input layer uses a linear activation method. The number of neurons in the hidden layer used is 15.

Based on the network architecture in Fig. 5, the function formed for interpolating January rainfall can be written as follows:

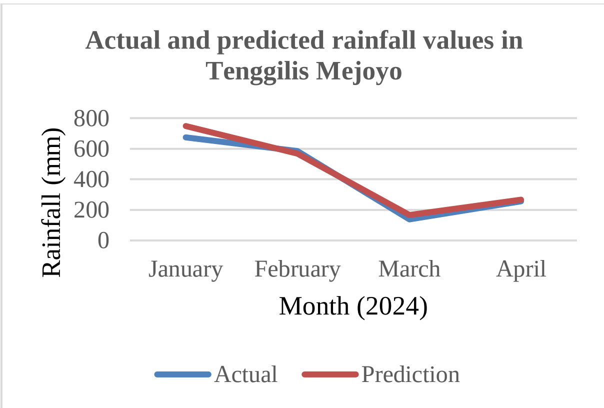
January Rainfall Interpolation Model:

$$y_k = \mathrm{w}_{0k} + \sum_{\mathrm{j}=1}^{15} \left(\frac{1}{1 + e^{-\left(\sum_{i=1}^{4} \sum_{j=1}^{15} x_i v_{ij} + \sum_{h=1}^{15} \sum_{j=1}^{15} y_h u_{hj} + v_{0j}\right)}} \right) \mathrm{w}_{\mathrm{jk}} + e_k$$

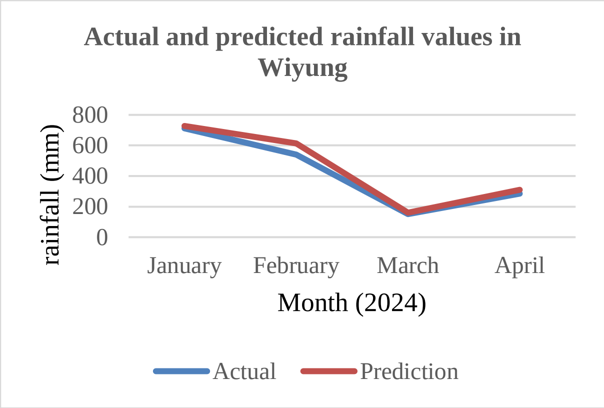
The results of rainfall interpolation using the RNN interpolation method are presented in Fig. 7.



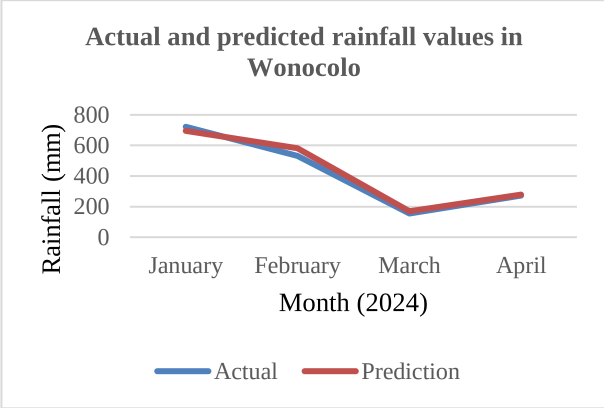
(a)



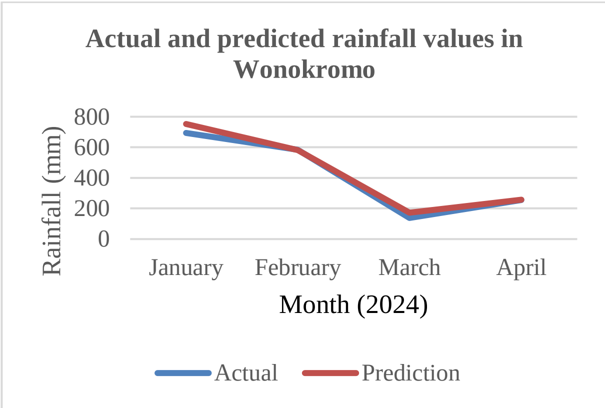
(b)



(c)



(d)



(e)

Figure 7. Plot of RNN Interpolation (a) Actual and Predicted Rainfall Values in Tegalsari, (b) Actual and Predicted Rainfall Values in Tenggilis Mejoyo, (c) Actual and Predicted Rainfall Values in Wiyung, (d) Actual and Predicted Rainfall Values in Wonocolo, (e) Actual and Predicted Rainfall Values in Wonokromo

Based on the plot between the actual value and the interpolated value, the interpolated value-RNN has a higher diversity when compared to the actual value. The Root Mean Square Error (RMSE) was calculated for the RNN interpolation method to evaluate its accuracy relative to Cokriging.

Table 2. RMSE Value from the Interpolation

Location		Rainfall (mm)			
		January	**February**	**March**	**April**
Tegalsari	Actual	725	584	138	255
	Prediction	680.8394	572.3268	163.0038	265.2305
Tenggilis Mejoyo	Actual	674	584	138	255
	Prediction	748.1353	566.7817	166.4277	266.5332
Wiyung	Actual	712	540	151	284
	Prediction	727.5042	614.0448	160.3877	309.8776
Wonocolo	Actual	722	532	155	272
	Prediction	695.343	581.4022	169.6113	278.2572
Wonokromo	Actual	694	584	138	255
	Prediction	752.6313	582.6005	172.5659	256.5382
RMSE		**48.6514**	**40.885**	**24.2064**	**13.7757**

Based on Table 2, the RMSE value from the RNN method shows that the rainfall interpolation results gave the best RMSE value in April, with an RMSE value of 13.7757. Table 2 presents the Root Mean Square Error (RMSE) values derived from rainfall interpolation across five districts in Surabaya: Tegalsari, Tenggilis Mejoyo, Wiyung, Wonocolo, and Wonokromo for January to April. The results indicate a clear trend of decreasing RMSE values from January (48.6514) to April (13.7757), suggesting improved interpolation accuracy during months with lower rainfall variability. This trend highlights the robustness of the applied interpolation method in capturing rainfall patterns during more stable climatic conditions. Among the districts, Wonokromo exhibits the lowest RMSE in April, demonstrating the method's ability to deliver accurate predictions in specific locations with consistent rainfall distribution. Conversely, higher errors observed in January suggest challenges in modelling rainfall patterns during periods of extreme variability, often associated with the peak of the rainy season. The results of this study are in line with previous research that demonstrated the reliability of Cokriging for rainfall interpolation in Indonesia. At the same time, our findings show that the Elman RNN can provide competitive or even superior accuracy, particularly during periods of lower rainfall variability. This supports earlier studies that highlighted the potential of neural networks in hydrological applications, while also emphasizing the role of dataset characteristics in shaping model performance. Unlike studies based on longer or denser datasets, the present work relied on limited spatial and temporal coverage, which underscores both its novelty and its limitations. These insights confirm that RNNs can serve as a useful complement to traditional geostatistical methods, while future research should aim at broader validation using larger datasets.

Despite these promising results, some limitations of the dataset should be acknowledged. Although the dataset in this study consists of only 31 districts (26 for training and 5 for testing), the results still provide meaningful insights for spatial rainfall interpolation in Surabaya. The relatively small dataset size may limit the robustness and generalizability of the RNN model, as neural networks generally benefit from larger amounts of training data. Consequently, potential biases or uncertainties may arise due to limited spatial coverage and variability. Nevertheless, by integrating elevation and spatial coordinates, the model was able to capture essential geographical patterns and deliver competitive accuracy compared to Cokriging. Future research should expand the dataset to include longer temporal coverage and additional monitoring stations, which would improve the reliability and applicability of the model in broader hydrological contexts. Overall, the present findings already demonstrate the applicability of RNN for urban hydrological analysis in Surabaya.

This study demonstrates the potential of Recurrent Neural Networks (RNNs) in spatial rainfall interpolation, showing competitive accuracy compared to the Cokriging method. The findings suggest that RNN can effectively capture spatial and temporal patterns of rainfall, making it a promising tool for hydrological applications in urban areas such as Surabaya. However, several limitations should be acknowledged. The dataset was restricted to 31 districts with a short temporal coverage (January–April 2024), which may reduce the robustness and generalizability of the results. Future research should address these limitations by incorporating longer rainfall records, additional monitoring stations, and other environmental variables to strengthen the model's reliability. Despite these constraints, the present findings provide a meaningful contribution and demonstrate the applicability of RNN for urban hydrological studies, offering a foundation for more comprehensive analyses in the future.

4. CONCLUSION

This study applied Cokriging and Recurrent Neural Network (RNN) methods for spatial rainfall interpolation in Surabaya. The results showed that the RNN model achieved competitive accuracy compared to Cokriging, particularly during months with lower rainfall variability. By integrating elevation and spatial coordinates, the RNN was able to capture essential geographical patterns and improve the representation of rainfall distribution. Although the dataset was limited to 31 districts with a short temporal coverage, the findings still provide valuable insights and confirm the potential of RNN as a complementary tool to geostatistical methods. Future research should incorporate larger datasets, longer temporal records, and additional predictor variables to further enhance the robustness and generalizability of the model.

Author Contributions

Danang Ariyanto: Conceptualization, Data Curation, Methodology, Writing – Original Draft, Supervision. A'yunin Sofro: Funding Acquisition, Project Administration, Writing – Review and Editing. Riskyana Dewi I Puspitasari: Data Curation, Formal Analysis, Visualization. Riska Wahyu Romadhonia: Software, Investigation, Validation. Hernando Ombao: Supervision, Writing – Review and Editing. All authors have read and approved the final manuscript.

Funding Statement

This research was funded by Universitas Negeri Surabaya (UNESA) through its internal research program Penelitian Dasar Non-APBN in 2024. The funding agency had no role in the design of the study, data collection, analysis, interpretation, or writing of the manuscript.

Acknowledgment

The authors would like to express their deepest gratitude to Universitas Negeri Surabaya (UNESA) for the continuous institutional support provided throughout the course of this research. The university's encouragement, academic environment, and access to facilities have been instrumental in enabling the authors to complete the study. The support extended reflects UNESA's strong commitment to advancing research and innovation, and the authors greatly appreciate the opportunity to contribute to this academic endeavor.

Declarations

The authors declare no competing interests.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement, including grammar, spelling, and clarity. The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. All final text was reviewed and approved by the authors.

REFERENCES

[1] Hartanto, S. Humaidi, E. Frida, N. Ananda, and M. Sinambela, "SPATIAL EVALUATION RAINFALL ESTIMATION ON WEATHER RADAR USING MARSHALL–PALMER REFLECTIVITY–RAINFALL RATE IN BANTEN," *Journal of Water and Land Development*, no. No 62, pp. 193–200, 2024. doi: https://doi.org/10.24425/jwld.2024.151567.

[2] E. P. Dewi, R. Juniatmoko, V. Arida, F. Fachruddin, and P. Pribadyo., "HIDROLOGI TEKNIK DAN AGROKLIMATOLOGI," 2023.

[3] E. T. Asmorowati, A. Rahmawati, D. Sarasanty, A. A. Kurniawan, M. A. Rudiyanto, E. Nadya, M. W. Nugroho and Findia., *DRAINASE PERKOTAAN*. Perkumpulan Rumah Cemerlang Indonesia, 2021.

[4] Ö. Baydaroğlu and I. Demir, "TEMPORAL AND SPATIAL SATELLITE DATA AUGMENTATION FOR DEEP LEARNING-BASED RAINFALL NOWCASTING," *Journal of Hydroinformatics*, vol. 26, no. 3, pp. 589–607, 2024. doi: https://doi.org/10.2166/hydro.2024.235.

[5] A. Pamungkas and S. Purwitaningsih, "IS SURABAYA BEING PLANNED AS A LOW-RISK CITY?," *International Review for Spatial Planning and Sustainable Development*, vol. 9, no. 1, pp. 78–92, 2021. doi: https://doi.org/10.14246/irspsda.9.1_78.

[6] D. Lettenmaier, "JOURNAL OF GEOGRAPHY & NATURAL ROLE OF RAINFALL IN HYDROLOGICAL MODELING AND WATER RESOURCE PLANNING," vol. 14, no. 1000319, pp. 1–2, 2024. doi: https://doi.org/10.35841/2167-0587.24.14.319.

[7] E. Pilecka, J. Zięba, and D. Szwarkowski, "ANALYSIS OF THE BEHAVIOUR OF THE HIGH AND STEEP SLOPE OF A ROAD MADE THROUGH WASTE UNDER THE INFLUENCE OF RAINFALL," *Archives of Civil Engineering*, vol. 67, no. No 2, pp. 293–307, 2021. doi: https://doi.org/10.24425/ace.2021.137169.

[8] M. Sholihin and S. E. Puspita Ghaniy Anggraini, *ANALISIS DATA PENELITIAN MENGGUNAKAN SOFTWARE STATA*. Penerbit Andi, 2021.

[9] M. R. A. Prasetya and A. M. Priyatno, "PENANGANAN IMPUTASI MISSING VALUES PADA DATA TIME SERIES DENGAN MENGGUNAKAN METODE DATA MINING," *Jurnal Informasi dan Teknologi*, pp. 52–62, 2023. doi: https://doi.org/10.37034/jidt.v5i2.324

[10] M. I. Syukur, "PENERAPAN METODE ROBUST KRIGING PADA DATA CURAH HUJAN WILAYAH SULAW ESI SELATAN UNTUK MENGESTIMASI ADANYA OUTLIER YANG DISEBABKAN OLEH DATA HILANG," 2023, *Universitas Hasanuddin*.

[11] A. P. Putra, "SPATIAL PATTERN OF SULFUR DIOXIDE DISPERSION AND AFFECTED SETTLEMENT AREAS IN SERANG REGENCY BANTEN PROVINCE POLA SPASIAL DISPERSI GAS SULFUR DIOKSIDA DAN AREA PERMUKIMAN TERDAMPAK DI KABUPATEN SERANG," *Jurnal Sains dan Teknologi Mitigasi Bencana*, vol. 15, no. 1, 2020. doi: https://doi.org/10.29122/jstmb.v15i1.4117

[12] F. Rizkiansyah, "IMPLEMENTASI METODE HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)–ARTIFICIAL NEURAL NETWORK (ANN) DALAM PERAMALAN HARGA MATA UANG KRIPTO," 2022.

[13] N. J. C. Konin, Y. A. N'go, G. E. Soro, and B. T. A. Goula, "IMPACT OF RAINFALL TRENDS ON FLOOD IN AGNÉBY WATERSHED," *Journal of Water and Land Development*, no. No 52, pp. 9–20, 2022. doi: https://doi.org/10.24425/jwld.2021.139938.

[14] N. Annisa, "ANALISIS SEBARAN SALINITAS AIR BERBASIS SISTEM INFORMASI GEOGRAFIS (STUDI KASUS SUNGAI TALLO)," 2023, *Universitas Hasanuddin*.

[15] Y. Saliba and A. Bărbulescu, “A COMPARATIVE EVALUATION OF SPATIAL INTERPOLATION TECHNIQUES FOR MAXIMUM TEMPERATURE SERIES IN THE MONTREAL REGION, CANADA,” *Romanian Reports in Physics*, vol. 76, no. 1, pp. 1–15, 2024. doi: https://doi.org/10.59277/RomRepPhys.2024.76.701.
[16] M. Zhang, D. Yu, Y. Li, and L. Zhao, “DEEP GEOMETRIC NEURAL NETWORK FOR SPATIAL INTERPOLATION,” *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 2022. doi: https://doi.org/10.1145/3557915.3561008.
[17] A. Antal, P. M. P. Guerreiro, and S. Cheval, “COMPARISON OF SPATIAL INTERPOLATION METHODS FOR ESTIMATING THE PRECIPITATION DISTRIBUTION IN PORTUGAL,” *Theoretical and Applied Climatology*, vol. 145, no. 3, pp. 1193–1206, 2021. doi: https://doi.org/10.1007/s00704-021-03675-0
[18] P. Pradabmook and T. Laosuwan, “ESTIMATION OF PM10 USING SPATIAL INTERPOLATION TECHNIQUES,” *International Journal on Technical and Physical Problems of Engineering*, vol. 13, no. 4, pp. 33–39, 2021.
[19] G. Urva *et al.*, *PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia, 2023.
[20] S. Rifky, L. P. I. Kharisma, A. R. Afendi and I. Zulfa, *Artificial INTELLIGENCE: TEORI DAN PENERAPAN AI DI BERBAGAI BIDANG*. PT. Sonpedia Publishing Indonesia, 2024.
[21] M. A. Faishol, E. Endroyono, and A. N. Irfansyah, “PREDICT URBAN AIR POLLUTION IN SURABAYA USING RECURRENT NEURAL NETWORK–LONG SHORT TERM MEMORY,” *Jurnal Ilmiah Teknologi Informasi*, vol. 18, no. 2, pp. 102–114, 2020. doi: https://doi.org/10.12962/j24068535.v18i2.a988
[22] Z. Chen, Z. Li, X. Tang, L. Chen, and N. Chen, “STPNet: A RECURRENT NEURAL NETWORK FOR SPATIOTEMPORAL PROCESSES PREDICTIVE LEARNING,” *Journal of Supercomputing*, vol. 81, no. 1, 2025. doi: https://doi.org/10.1007/s11227-024-06823-1.
[23] H. Febrianto, A. Fariza, and J. Hasim, “URBAN FLOOD RISK MAPPING USING ANALYTIC HIERARCHY PROCESS AND NATURAL BREAK CLASSIFICATION (CASE STUDY: SURABAYA, EAST JAVA, INDONESIA),” 2016, pp. 148–154. doi: https://doi.org/10.1109/KCIC.2016.7883639.
[24] Y. R. Savitri, R. Kakimoto, R. A. Begum, N. Anwar, W. Wardoyo, and E. Suryani, “THE APPLICATION OF AHP TO DETERMINE THE PRIORITY DRAINAGE SYSTEM ON FLOOD MITIGATION IN SURABAYA – INDONESIA,” *Journal of Disaster Research*, vol. 17, no. 3, pp. 431–443, 2022. doi: https://doi.org/10.20965/jdr.2022.p0431.
[25] C. M. Duff, “USING SPACETIME GEOSTATISTICAL ANALYSIS TO IMPROVE INTERPOLATION,” *Journal of Hydrology*, 2025. doi: https://doi.org/10.1016/j.jhydrol.2024.131987.
[26] T. Page, “INTERPOLATION OF RAINFALL OBSERVATIONS DURING EXTREME EVENTS: COMPARING NNI, OK, AND CK,” *Hydrological Processes*, vol. 36, no. 11, p. e14758, 2022. doi: https://doi.org/10.1002/hyp.14758.
[27] Q. Hu, Z. Li, L. Wang, Y. Huang, Y. Wang, and L. Li, “RAINFALL SPATIAL ESTIMATIONS: A REVIEW FROM SPATIAL INTERPOLATION TO MULTI-SOURCE DATA MERGING,” *Water*, vol. 11, no. 3, 2019. doi: https://doi.org/10.3390/w11030579.
[28] M. A. Faishol, E. Endroyono, and A. N. Irfansyah, “PREDICT URBAN AIR POLLUTION IN SURABAYA USING RECURRENT NEURAL NETWORK–LONG SHORT TERM MEMORY,” *Jurnal Ilmiah Teknologi Informasi*, vol. 18, no. 2, pp. 102–114, 2020. doi: https://doi.org/10.12962/j24068535.v18i2.a988
[29] I. F. Damanik, “PENERAPAN METODE JARINGAN SYARAF TIRUAN UNTUK MEMPREDIKSI NILAI TUKAR PETANI SUBSEKTOR TANAMAN PANGAN MENGGUNAKAN ALGORITMA CONJUGATE GRADIENT POLA-REBIERI,” *SkripsiKu-2022*, vol. 1, no. 2, 2022.
[30] M. Zakaria, L. Pagiling, and W. O. S. N. Alam, “SISTEM PENYIRAMAN OTOMATIS TANAMAN SEMUSIM BERBASIS JARINGAN SARAF TIRUAN MULTILAYER PERCEPTRON,” *Jurnal Fokus Elektroda: Energi Listrik, Telekomunikasi, Komputer, Elektronika Dan Kendali*, vol. 7, no. 1, p. 35, 2022. doi: https://doi.org/10.33772/jfe.v7i1.24050
[31] W Tri Zahrotun, “PERAMALAN HARGA GABAH KERING PANEN DI INDONESIA MENGGUNAKAN METODE HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE EXOGENOUS (ARIMAX)–FEED FORWARD NEURAL NETWORK (FFNN),” 2022, *Universitas Muhammadiyah Semarang*.
[32] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A CRITICAL REVIEW OF RECURRENT NEURAL NETWORKS FOR SEQUENCE LEARNING,” pp. 1–38, 2015.
[33] C.-C. Peng and G. D. Magoulas, “SEQUENCE PROCESSING WITH RECURRENT NEURAL NETWORKS,” *Encyclopedia of Artificial Intelligence*, pp. 1411–1417, 2011. doi: https://doi.org/10.4018/978-1-59904-849-9.ch207.
[34] G. R. Kanagachidambaresan, A. Ruwali, D. Banerjee, and K. B. Prakash, “RECURRENT NEURAL NETWORK,” *Programming with TensorFlow: Solution for Edge Computing Applications*, pp. 53–61, 2021.
[35] J. L. Elman, “FINDING STRUCTURE IN TIME,” *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990. doi: https://doi.org/10.1207/s15516709cog1402_1.
[36] A. Bond, P. M. Léger, and S. Sénécal, *THE EFFECT OF DEVICE-AFFORDANCE ALIGNMENT WITH THE USER GOAL ON USER EXPERIENCE*, vol. 12423 LNCS. Springer International Publishing, 2020. doi: https://doi.org/10.1007/978-3-030-60114-0_3.
[37] R Core Team, “R: A LANGUAGE AND ENVIRONMENT FOR STATISTICAL COMPUTING,” 2023, *Vienna, Austria*.
[38] RStudio Team, “RSTUDIO: INTEGRATED DEVELOPMENT ENVIRONMENT FOR R,” 2023, *Boston, MA*.
[39] T. O. Hodson, “ROOT MEAN SQUARE ERROR (RMSE) OR MEAN ABSOLUTE ERROR (MAE): WHEN TO USE THEM OR NOT,” *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–10, 2022. doi: https://doi.org/10.5194/gmd-2022-64
[40] E. J. Pebesma, “MULTIVARIABLE GEOSTATISTICS IN S: THE GSTAT PACKAGE,” Computers & Geosciences, vol. 30, no. 7, pp. 683–691, Dec. 2004. doi: https://doi.org/10.1016/j.cageo.2004.03.012.
[41] B. Gräler, E. Pebesma, and G. Heuvelink, “SPATIO-TEMPORAL INTERPOLATION USING GSTAT,” The R Journal, vol. 8, no. 1, pp. 204–218, Aug. 2016. doi: https://doi.org/10.32614/RJ-2016-014.