Rainfall Data Modeling with Artificial Neural Networks Approach

S Astutik1*, H Pramoedyo1, N S Rahmi1, D Irsandy1, and R H P Y Damayanti1

1Statistics Department, FMIPA, Universitas Brawijaya, Indonesia

*Email : suci_sp@ub.ac.id

Abstract. Rainfall is one of the important information that widely used in various fields. Rainfall data involving location information is referred to as spatial rainfall data. Some of the model approaches to spatial rainfall data are the Vector Autoregressive (VAR) model, state-space, Markov chain stochastic model, and Geographically Weighted Regression (GWR). However, these models have not been able to produce predictions of the occurrence of no rain (zero value) or extreme values. Currently, theoretical modelling is mostly approached by artificial neural network (ANN) techniques. The purpose of this study is to model spatial rainfall data in East Java, Indonesia in 2020 with the ANN approach which is supported by several variables such as location and elevation information. The ANN used backpropagation and Rprop by combining the learning rate and layer which is then obtained the RMSE value. The results show that the best model has the smallest RMSE of 1.22 when the learning rate is 0.15 on 11 layers using Rprop algorithm.

Keywords : Rainfall, ANN, backpropagation, Rprop

1. Introduction

Rainfall is the important attribute to identification of climate change, select a cropping pattern, hydrological cycles, and others. Accurate information about rainfall is very important for water planning and availability, as a basis for disaster mitigation efforts, and supporting daily human activities. However, rainfall is the elements of the hydrological cycle that is difficult to reach and model because of the complexity of the atmosphere that produces rainfall with various variations both in space and time. Therefore, proper modeling of rainfall data is needed in predicting rainfall so that it can be implemented in predicting climate change, determining planting time, hydrological modeling of rain flows, modeling early detection of floods in flood disaster mitigation, and others.

Rainfall modeling has been developed by researchers using various techniques, including: rainfall model with Bayesian state space approach on spatio-temporal data [1], rainfall data modeling with ZIG [2,3], rainfall model using the Vector Autoregressive (VAR) approach [4], rainfall modeling with Geographically Weighted Regression (GWR) [5], prediction of annual rainfall patterns by discussing the HMM Maximum Likelihood approach in Nigeria [6]. However, rainfall modeling has a nonlinear nature so that it has not been able to predict the occurrence of no rain (zero data) or extreme data.

Artificial Neural Network (ANN) is an artificial intelligence method that has currently been developed for a theoretical modeling approach. The ANN algorithm is an attractive inductive approach in predicting rainfall with its high nonlinearity, flexibility, and data-based learning in building models without information about the flow process. In addition, modeling with ANN has the desired attributes and the ability to learn from examples without requiring explicit physical data. Several studies related to ANN in modeling rainfall relationships have been carried out, including: Interpolation of rainfall data with ANN [7]; ANN for modeling rainfall runoff from watersheds in Bali [8]. Applications of ANN for
12-month forecasting of rainfall and evaporation indices in Nigeria [9]. This study is limited to annual rainfall. This study aims to model spatial rainfall data with the ANN approach.

This study aims to model and evaluate spatial rainfall data using the ANN approach. The results are expected to provide benefits in the development of science and technology, especially statistics with spatial rainfall data and provide benefits in other fields such as hydrology, health, and economics.

2. Methodology
This study uses the ANN method to predict a daily rainfall in East Java. ANN is a predictive model that can predict rainfall with a better random pattern of rain events. Neural networks are a new knowledge paradigm [10] that imitates the human brain in the process of solving and storing memories. ANN applications have been widely applied in predictions in the fields of climatology and hydrology. The model developed in this study is expected to be better at achieving extreme values for better rainfall prediction results.

2.1. Artificial neural networks
ANN is an information processing system that has a similar trait to biological neural networks [8]. It is a representation of the human brain consisting of neurons. Neurons in the human brain have the same trait in ANN, consisting of groups called layers. Neurons in every layer are connected to other adjacent layers. The strength of the relationship between adjacent neurons is represented in the strength of the relationship or weight.

In general, there are three layers of ANN called input layer, hidden layer, and output layer. The input layer as receiver an input that describes a problem from the external environment. The hidden layer as receiver the input from the input layer, and then carry the output to the output layer and send it to the user.

One layer of Perceptron introduced by Rosenblatt (1957) [11]. It is a simplest form ANN. The input is given directly to the output through a weight relation. About 1960, Multi Layer Perceptron (MLP) was developed and included in the feed-forward network category. In a feed-forward network, weighted connections provide activation only in the forward direction from the input layer to the output layer (Figure 1).

![Figure 1. Feed forward network.](image)

The output from the ANN for an input pattern depends on the weight of the relationship between neurons. For pattern recognition applications in ANN, an initial training is needed. It can be used in receiving input from outside. An ANN can solve complex problems if it uses the right weight values between neurons in different layers. The right weight value is obtained through the training process, which is a process to change the weights between neurons. Because of it, a network can solve a problem in the study case.

ANN has three components, namely the network architecture, determining the weight of the association, and the activation function [12]. The network architecture determines the success of the target to be achieved because not all problems can be solved using the same architecture. The artificial neural network architecture consists of a single layer network, a multiple layer network, and a
competitive layer network [13]. The activation function is a function for each neuron that is designed to limit the output of the neuron. In general, the activation function has a value between 0 to 1 or -1 to 1. The activation function in hidden layer neurons is needed to apply nonlinear forms to the neural network and plays an important role in determining neuron output at various values.

The determination of the matrix weights through training is an urgent step in ANN model. There are two types of training mechanisms, namely supervised and unsupervised training. Supervised training requires external supervision to guide the training process. This algorithm apply some of input-output data that are applied as examples. Next, calculate the difference among output estimation from network and the actual output. It is applied to change the weight of the network so that the output is the same with the target and the probles are solved.

ANN is a predictive model that can predict rainfall with a better random pattern of rain events. Neural network is a new knowledge paradigm [10] which imitates the human brain in the process of solving and storing memory. ANN applications have been widely applied in predictions in the field of climatology and hydrology. The model developed in this study is expected to be better in achieving extreme values for better rainfall prediction results.

\[ y_k = \frac{1}{1 + e^{-w_{kj}h_j}} \]  
(1)

\[ h_j = \frac{1}{1 + e^{-2w_{ij}x_i}} \]  
(2)

\[ \sum w_{ij}X_i = w_0 + w_{ij}X_1 + \cdots + w_{Tj}X_T \]  
(3)

w<sub>ij</sub> is weight between matrix X (predictor variable) and matrix H (hidden layer), v<sub>ij</sub> is weight between matrix H and matrix Y (response variable), and Y<sub>T</sub> is rainfall at T-location.

2.2. Backpropagation algorithm

Supervised training has the mechanism called backpropagation algorithm [14] is commonly used in engineering applications. The idea of backpropagation is the iterative application of the chain rule to calculate the effect of each weight in the network with respect to the error function [15]. The network is given a pair of patterns consisting of the input pattern and the desired pattern. When a pattern is given to the network, the weights are changed to minimize the difference between the output pattern and the desired pattern. This exercise is done repeatedly so that all patterns issued by the network can meet the desired pattern. The choice of learning rate has an important influence on the time to reach convergence. If learning rate is small than Learning process need a long time,. However, the weight value is far from the minimum weight if the learning rate is large, [15].

2.3. Resilient propagation algorithm (Rprop)

The problem of backpropagation algorithm like the weight is far from the minimum weight can be solve with the new algorithm, namely algorithm of Resilient Propagation. Rprop is a new efficient learning scheme by adapting directly by weighting indicated on local gradient information. The Rprop algorithm uses positive and negative signs that show the direction of weight adjustment [14]. The adaptation rule works as follows: any partial derivatives of weights and biases which in two successive iterations differ in sign indicating that the improvement of the last value is too large it has skipped the local minimum, the value of improvement ij is reduced by a factor of . If in two iterations the order of the sign of the derivative remains, the improvement value is increased by a factor of + to accelerate the convergence of the error surface area. The update value remains the same, if the child is zero, the weight is reduced by the update value if the derivative is positive (increasing error), but the update value is added if the derivative is negative, mathematically can follows equation (4) and equation (5).
\[ \Delta w_{ij}(t) = \begin{cases} -\Delta_i j(t), & \text{if } \frac{\partial E}{\partial w_{ij}} > 0 \\ +\Delta_i j(t), & \text{if } \frac{\partial E}{\partial w_{ij}} < 0 \\ 0, & \text{else} \end{cases} \] (4)

\[ w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)} \] (5)

2.4. Data preprocessing
Data preprocessing is the initial stage in the research (15). In this study, the min-max scaler method can be written as equation (6).

\[ X_{\ast} = \frac{\text{minRange} + (x - \text{minValue})(\text{maxRange} - \text{minRange})}{\text{maxValue} - \text{minValue}} \] (6)

2.5. Data resources
Data in this study obtained from website of BMKG. The data that applied in this study is rainfall data in 2020 at 11 rain stations in East Java which is 8 stations as training data and 3 stations as testing data. The eleven weather stations is as a spatial unit.

East Java has astronomically located between 111° east longitude and 7°,12'-8°,48' south latitude. East Java has an area of 47,963 km² with the northern boundary is the Java Sea, the east is bordered by the Bali Strait, the south is the Indian Ocean, and the west is bordered by the province of Central Java. East Java Province in general has a tropical climate with two changes in the seasons, namely the dry season and the rainy season. Until December, all areas in East Java have entered the rainy season. Almost every day it rains in all areas with light to heavy intensity which is observed through 11 Meteorology, Climatology and Geophysics stations. Rain is influenced by the topographic conditions of each region.

Rainfall is defined as rainwater that has a certain height that collects in a rain gauge, does not percolate, does not flow, and does not absorb. The height of this water is usually expressed in millimetre. Rainfall in 1 mm means that in an area of one square meter a flat place can accommodate one mm of rainwater.

Rain intensity is the amount of rainfall per unit time. The longer the rain lasts, the higher the intensity [16]. Enough rain has benefits for life including in agriculture and daily needs. However, if the intensity of rainfall in an area is too high, it causes a flood disaster and if the availability of water is very less and the rainfall is very low, it will cause a water crisis. The height of rainfall that occurs between regions will be different. An area will have these different characteristics. Therefore, there are topographical factors of the area indicated by latitude, longitude, and altitude.

3. Results and discussion
The analytical method used is descriptive analysis and ANN using backpropagation and Rprop algorithm. The analysis is used to determine the general description and characteristics of the data. The results are summarized in Table 1. Most of the rainfall in East Java is 8.396 mm.

| Table 1. Descriptive statistics. |
|-------------------------------|
| Mean (mm) | Minimum (mm) | Maximum (mm) |
| 8.3955    | 5.3596       | 17.55262948  |

Based on Table 1, it shows that the lowest daily rainfall in 2020 is at the Banyuwangi meteorological station, which is equal to 5.3957 mm, while the highest rainfall occurs at the Nganjuk geophysical station of 17.5526 mm. Geographically, the Banyuwangi meteorological station is in the eastern part of East Java.
Java Province and the Nganjuk geophysical station is in the western part of East Java Province. The rainfall in the eastern part of the island is lower than the western part. This is as a result of the influence of the dipole mode, namely the phenomenon of the interaction of the seawater atmosphere in the Indian Ocean [17]. Visually, rainfall in East Java is shown in Figure 2 that the highest rainfall in western of East Java include Kabupaten Nganjuk, Kabupaten Ponorogo, Kabupaten Madiun, Ponorogo, and Kediri.

This study implements ANN modeling with two algorithms, namely backpropagation and Rprop to obtain a suitable and more efficient model for predicting rainfall in East Java. The learning rates used are 0.01, 0.05, 0.1, and 0.15 with 10 to 15 hidden layers. The best ANN model is chosen with the smallest RMSE value. The results of the RMSE for each algorithm are presented in Table 2 and Table 3.

The backpropagation algorithm produces the smallest RMSE value at a learning rate of 0.1 in 10 hidden layers, which is 2.45. While the Rprop algorithm produces the smallest RMSE of 1.22 with a learning rate of 0.15 on 11 hidden layers. Overall the RMSE value in the Rprop algorithm is smaller than that of backpropagation. In the Rprop algorithm, it can be seen that the RMSE spreads between 1.22 to 6.96, while in backpropagation the RMSE spreads between 2.45 to 9.88. That is, the best ANN model in this case is when the learning rate is 0.15 with 11 hidden layers with the Rprop algorithm. Rprop is a new efficient learning scheme by adapting directly by weighting based on local gradient information. The best ANN model is the basis for predicting rainfall involving topographic information. The results of rainfall predictions on both algorithm in East Java are shown in Figure 3 and Figure 4.

### Table 2. RMSE value of ANN model with backpropagation algorithm.

| Learning Rate | 10  | 11  | 12  | 13  | 14  | 15  |
|---------------|-----|-----|-----|-----|-----|-----|
| 0.01          | 5.45| 5.21| 6.32| 4.61| 4.13| 9.88|
| 0.05          | 4.29| 7.41| 6.03| 4.36| 5.93| 4.38|
| 0.1           | 2.45| 6.25| 6.51| 6.14| 4.72| 7.97|
| 0.15          | 4.25| 7.44| 6.90| 7.15| 5.66| 5.68|
Table 3. RMSE value of ANN model with Rprop algorithm.

| Learning Rate | Hidden Layer |
|---------------|--------------|
|               | 10 | 11 | 12 | 13 | 14 | 15 |
| 0.01          | 2.99 | 3.17 | 4.14 | 2.59 | 4.22 | 2.10 |
| 0.05          | 2.08 | 1.52 | 2.21 | 6.96 | 4.60 | 2.18 |
| 0.1           | 1.95 | 2.42 | 1.85 | 2.06 | 2.52 | 5.28 |
| 0.15          | 1.84 | 1.22 | 2.81 | 2.07 | 2.05 | 3.05 |

Figure 3. Prediction of east java rainfall using backpropagation algorithm

Figure 4. Prediction of east java rainfall using Rprop algorithm

Rainfall prediction map in East Java with backpropagation algorithm is divided into 4 parts, namely very low, low, medium, and high rainfall. This means that there are no areas in East Java with very high rainfall. Almost part of East Java has relatively low rainfall. In addition, showed on Figure 3, it can be stated that the distribution of rainfall in East Java is different from the actual conditions (Figure 2).

The prediction map for rainfall in East Java in 2020 with the Rppop algorithm looks similar to the actual conditions. Areas with relatively high rainfall (High-very high) occur in the western part of East Java. However, there is a relatively low difference in rainfall between the predicted results and the actual conditions. On the map, the prediction results show that areas with low rainfall are higher than areas with very low rainfall. While the actual data shows the opposite, namely areas with very low rainfall more than areas with low rainfall.
Based on the two results of prediction of rainfall in both algorithms, it can be seen that the Rprop algorithm is better than the backpropagation algorithm. That's because the rainfall prediction map in East Java on the Rprop algorithm is more similar to that of backpropagation.

4. Conclusion
This study implements the ANN model to predict rainfall in East Java at eleven weather stations. Not only rainfall, but also modeling is done by involving topographical conditions at the eleven weather stations, namely spatial information in the form of astronomical locations and elevations. ANN modeling uses backpropagation and Rprop algorithms to obtain a suitable model with a high level of accuracy. The results of the analysis show that rainfall in East Java by involving spatial information can be modeled through ANN with backpropagation and Rprop algorithms. However, based on RMSE the Rprop algorithm shows better results than the backpropagation algorithm. The ANN model with the Rprop algorithm predicts rainfall in various locations in East Java.

5. Acknowledgment
We express our gratitude to the Rector of Universitas Brawijaya and Dean of Faculty of Mathematics and Natural Sciences, Universitas Brawijaya for providing the financial support in the DPP/SPP grant. We are also thankful to the reviewers for the valuable comments and suggestions on the earlier draft paper.

References
[1] Astutik S, Iriawan N, Nair G, Suhartono and Sutikno 2013 Bayesian state space modeling for spatio-temporal rainfall disaggregation. *Int. J. Appl. Math. Stat.* 37 26-37
[2] Astutik S, Pramoedyo H and Solimun 2013 ZIG Model on Intermittent and Positively Skewed Rainfall Data Proeceeding First International Conf. Sci. 1 184–190
[3] Dzupire N C, Ngare P and Odongo L 2018 A Poisson-Gamma Model for Zero Inflated Rainfall Data *J. Probab. Stat.* 2018 1-12
[4] Sukarna, Abdy M and Asnida B 2015 Vector Autoregressive Model for Forecast Rainfall in Makassar *JMathCoS, J. Math. Comput. Stat* 1 1-8
[5] Kumari M, Singh C K, Bakimchandra O and Basistha A 2017 Geographically Weighted Regression Based Quantification of Rainfall-Topography Relationship and Rainfall Gradient in Central Himalayas *Int. J. Climatol.* 37 1299-1309
[6] Lawal A, Abubakar U, Danladi H and Gana A 2016 Prediction of Annual Rainfall Pattern Using Hidden Markov Model (HMM) in Jos, Plateau State, *Nigeria J. Appl. Sci. Environ. Manag.* 20 617-622
[7] Karmakar S, Kowar M K and Guhathakurta P 2009 Spatial Interpolation of Rainfall Variables using Artificial Neural Network *Int. Conf. Adv. Comput. Commun. Control.* 09 547-552
[8] Dharma I S, Putera I A and Ardana P D H 2011 Artificial Neural Networks untuk Pemodelan Curah Hujan-Limpasan pada Daerah Aliran Sungai (DAS) di Pulau Bali *Bumi Lestari J. Environ.* 11 9-22
[9] Ogunrinde A T, Oguntunde P G, Fasinmirin J T and Akinwumi A S 2020 Application of Artificial Neural Network for Forecasting Standardized Precipitation and Evapotranspiration Index : A Case Study of Nigeria *Engineering Reports* 2 1-18
[10] Syarifuddin M, Koesmarno Y and Pramudia A 2009 Pengembangan Model Jaringan Syaraf Tiruan untuk Prediksi Curah Hujan Bulanan dan Pemanfaatannya Bagi Perencanaan Pertanian di Kabupaten Subang dan Karawang *Forum Pascasarj.* 32 169-178
[11] Hung N Q, Babel M S, Weesakul S and Tripathi N K 2009 An Artificial Neural Network Model for Rainfall Forecasting in Bangkok, Thailand *Hydrol. Earth Syst. Sci.* 13 1413-1425
[12] Pratama R A and Anifah L 2016 Peramalan Beban Listrik Jangka Panjang Provinsi D.I. Yogyakarta Menggunakan Neural Network Backpropagation *J. Tek. Elektro.* 5 37-47
[13] Wuryandari M D and Afrianto I 2012 Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation dan Learning Vector Quantization pada Pengenal Wajah *J. Komput. Dan Inform.* 1) 45-51
[14] Riedmiller M and Braun H 1993 Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm *IEEE Int. Conf. Neural Networks* 1 586-591

[15] Nasution D A, Khotimah H H and Chamidah N 2019 Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN *J. Comput. Eng. Syst. Sci.* 4 78-82

[16] Hendri A 2015 Analisis Metode Intensitas Hujan pada Stasiun Hujan Pasar Kampar Kabupaten Kampar *Annu. Civ. Eng. Semin.* 1 297-304

[17] Mulyana E, 2002 Pengaruh Dipole Mode Terhadap Curah Hujan Di Indonesia *J. Sains Teknol. Modif. Cuaca.* 3 39-43