

APPLICATION OF ANFIS TECHNIQUE FOR PREDICTION OF RAINFALL IN THE SOUTH TANGERANG CITY

Wayan SUPARTA^{1*}, RESDIANSYAH² and Safitri JAYA³

DOI: 10.21163/GT_2020.152.01

ABSTRACT:

Flood phenomenon is an annual scourge that often occurs at several critical points in the territory of Indonesia, one of which is in the South Tangerang City, Banten. This paper aims to predict rainfall as an early detection model by exploring the application of artificial intelligence techniques such as the Adaptive Neuro-Fuzzy Inference System (ANFIS). The proposed technique was built with various input structures and membership functions, then trained and tested to evaluate its capabilities. Daily meteorological data such as relative humidity, temperature, and partial pressure of water vapour are chosen as intermediate parameters to predict rainfall. While rainfall measurement data is used as a comparison or validation. By analysing the three scenarios that have been carried out, it was found that rainfall can be estimated well. The prediction results with ANFIS Fuzzy-C-Means (FCM) algorithm are promising to detect floods that can later be integrated with more complex systems, especially types of floods at various points in the City of South Tangerang and surrounding areas.

Key-words: ANFIS, Rainfall, Urban Floods, South Tangerang City.

1. INTRODUCTION

Indonesia is clearly a country with a disaster-prone zone, starting from floods, landslides, earthquakes to tsunamis. With a remarkable tropical climate where the country tends to be hot, the temperature is not the main parameter of climate in this country. In other words, rainfall plays an important role in climate change, especially during the rainy season, where rainfall occurs quite high in each year. The rainy season usually triggers frequent of a natural disaster such as landslides and floods. Floods can be caused by various factors such as the condition of the catchment area, rain duration and intensity, land cover, topographic conditions, and drainage network capacity (Ghazavi et al., 2016; Fujiki and Renard, 2018). While the basic factor of a disaster is the result of climate change. Among the types of floods that are very upsetting besides flash floods are urban floods, where people's lands have now shifted to various types of new housing.

Based on flood forecasts in Indonesia, December, January and February throughout the year are potentially high flood months. Similar to neighboring cities like Jakarta, the South Tangerang is also no exception to this flood phenomenon. The South Tangerang City is located in the eastern part of Banten Province, at the coordinates of 106°38"- 106°47"E and 06°13'30" - 06°22'30"S. Most of this area is lowland and has a relatively flat topography with a slope of 0 - 3% (average) while the altitude of the region is between 0-25 m above sea level. Looked from the climate condition, the intensity of the sun is around 50% and the highest rainfall (December - January) is reaching 264.4 mm. Until 2019, there are 13 locations are regularly subscribed to flooding because of inadequate drainage (Kompas, 2018). The Ciputat Block is the most flood point and other areas adjacent to the Universitas Pembangunan Jaya (UPJ) are Jombang and Cantiga Blocks.

With the above climatic and topographic conditions, the development of an early warning system (EWS) that can provide fast, accurate, targeted, easily accepted and understood, reliable and sustainable information is indispensable. The previous system was works by measuring the water

^{1,3}Universitas Pembangunan Jaya, Department of Informatics, South Tangerang City, Banten 15413, Indonesia, Corresponding author: wayan.suparta@upj.ac.id, safitri.jaya@upj.ac.id

² Universitas Pembangunan Jaya, Department of Civil Engineering, South Tangerang City, Banten 15413, Indonesia, resdiansyah.mansyur@upj.ac.id.

level in the river and then classified according to the danger level. The disadvantage of this system is that new flooding can be predicted if the river's water level is at an almost overflowing height. This causes the warning given too short and the community does not have time to prepare self-preservation. In addition, the existing monitoring system in this region cannot be recorded automatically so that long-term historical data cannot be utilized for the benefit of future investigations. In this study, a softcomputing method will be proposed with consideration of low costs, easy to analyze and still have good accuracy. One form of hybrid softcomputing technique to be developed is an Adaptive Neuro Fuzzy Inference System (ANFIS) model. ANFIS is a combination of fuzzy logic and artificial neural networks. Exploration of the ANFIS model in this study is expected to predict flooding.

ANFIS model has been proposed to develop a forecasting seasonal rainfall in Victoria of southern Australia to anticipate floods (Mekanik et al., 2015) and for spatial flood predictions such as Hong et al. (2017) and Bui et al. (2018). However, the model applied is also very dependent on regional characteristics. The purpose of this study is to predict urban flood trough rainfall parameter using ANFIS. These urban floods become a driver in the development of a softcomputing in order to improve the accuracy of the flood early detection system. The focus of research is to obtain the capability of the models built to predict future flooding based on the data that has been trained. The data collected will also be analyzed to carry out patterns and flood trends. For preliminary investigation, rainfall in the South Tangerang City will be used as a testing model (training and validation). The model developed will help provide alternative prediction in anticipating floods that can be used as references by decision makers in flood mitigation and educate the public or users with the aim of providing early weather warnings, especially related to flooding.

2. METHODS

2.1. Dataset and location

A number of measurement data is collected to support the study. The primary data are obtained from Indonesia Agency for Meteorology, Climatology and Geophysics or BMKG (*Badan Meteorologi, Klimatologi dan Geofisika*).

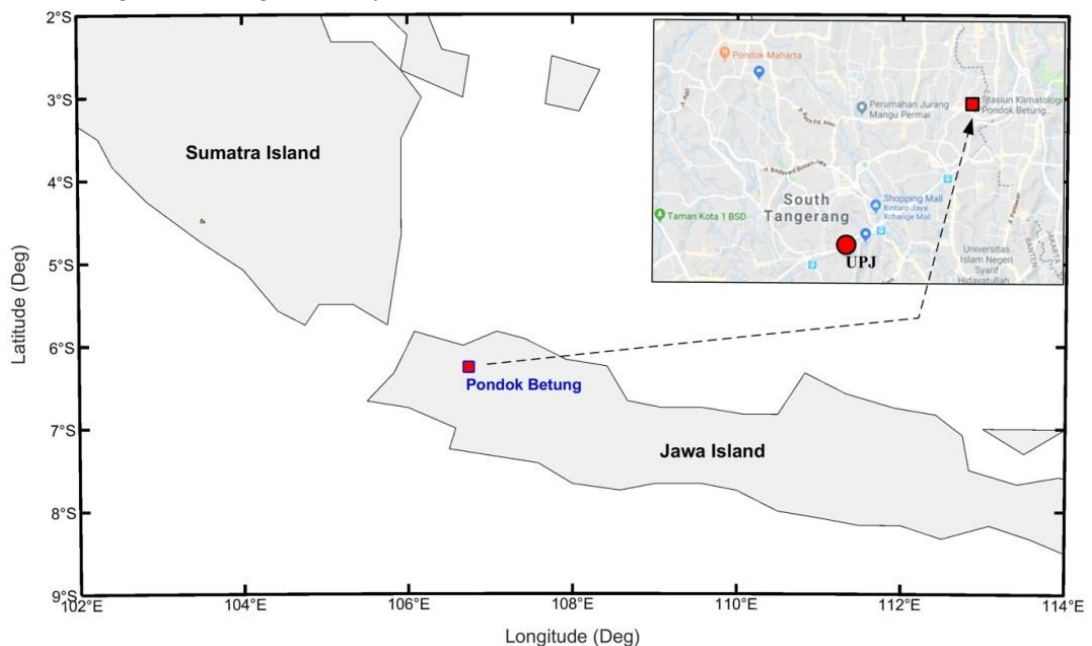


Fig. 1. The location of study and floods point by Google Maps in South Tangerang City.

The data collected includes rainfall, relative humidity, and surface temperature. Because the BMKG in one district only managed one weather station, this study will use data at the South Tangerang City which located in Pondok Betung station (geographically: 6°15'40.22"S, 106°44'56.26"E). The meteorological data provided by BMKG is available on a daily basis. This existing data is utilized entirely to predict floods. In this work, ten years of data will be employed to make an accurate prediction of rainfall. The location of meteorology station used in this study is depicted in **Fig.1**.

Since no surface pressure data provided free by BMKG, the partial pressure of water vapour (P_w) is employed. It can be obtained from the relative humidity (RH) and temperature as recommended by the World Meteorological Organization (WMO) Technical Note No. 8 as in Suparta et al. (2008) and defines as follows:

$$P_w = \frac{RH}{100} \exp\left(-37.2465 + 0.213166T_K - 2.56908 \times 10^{-4} T_K^2\right) \quad (1)$$

where P_w is in mbar and T_K is mean temperature in Kelvin.

2.2. ANFIS model development for rainfall predictions

The approach used for rainfall prediction is the utilization of ANFIS techniques rather than Artificial Neural Network (ANN) (Tierra, 2016). One of the advantages of ANFIS, it has a similar concept to the concept of human thinking where they can learn something from the data that has trained. Furthermore, it can self-learning effectively and memory abilities of neural networks and brings a more stable training process (Huang et al., 2018). This is the idea of Tagaki-Sugeno-Kang (Jang, 1993; Jang & Sun, 1995) which apply the first order of the ANFIS model. The approach can use types of supervised learning and unsupervised learning algorithms. In this work, a neural multilayer advanced feed network with supervised learning processes will be selected. From the various types of supervised learning algorithms, the backpropagation-learning algorithm (BPL) is chosen as a flood prediction system. The choice of BPL is because it has been widely applied. About 90% of backpropagation has been successfully applied in many fields, such as finance, handwriting pattern recognition, voice pattern recognition, system control, medical image processing, and light prediction made by NASA.

Fig. 2 shows the basic ANFIS architecture with two inputs and one output. In general, ANFIS is constructed with five layers, i.e. the input layer as antecedent parameters, three hidden layers with two constant parameters and one consequent parameter, and one output layer. In more detail about the ANFIS concept, authors can refer to Suparta and Alhasa (Suparta & Alhasa, 2013; 2016a; 2016b, 2013; Suparta & Putro, 2017). In brief, the model can be explanation as follows.

Layer 1: Every node in this layer transforms input into degree between 0 and 1 (fuzzification) which so-called premise parameter. It is an activation function with membership function such as gauss, triangular, trapezoidal, and generalized Bell. The node output is given by,

$$O_{1,i} = \mu_{A_i}(x), \quad i = 1,2 \text{ and } O_{1,i} = \mu_{B_{i-2}}(y), \quad i = 3,4 \quad (1)$$

where x or y is the input to the node i ; A_i or B_{i-2} is a linguistic label of fuzzy set associated with this node, $O_{1,i}$ is the membership functions (MFs) grade of fuzzy set. Assuming the Gaussian membership function is used.

$$\mu_{A_i}(x) = \exp\left[-\left(\frac{x-c_i}{2a_i}\right)^{2b}\right] \quad \mu_{B_{i-2}}(x) = \exp\left[-\left(\frac{x-c_i}{2a_i}\right)^{2b}\right] \quad (2)$$

where $\{a_i, b_i, c_i\}$ are the parameter set of the membership function in premise parameter that can change the shape of the membership function. As the value of these parameters' changes, the gauss shaped functions will vary accordingly

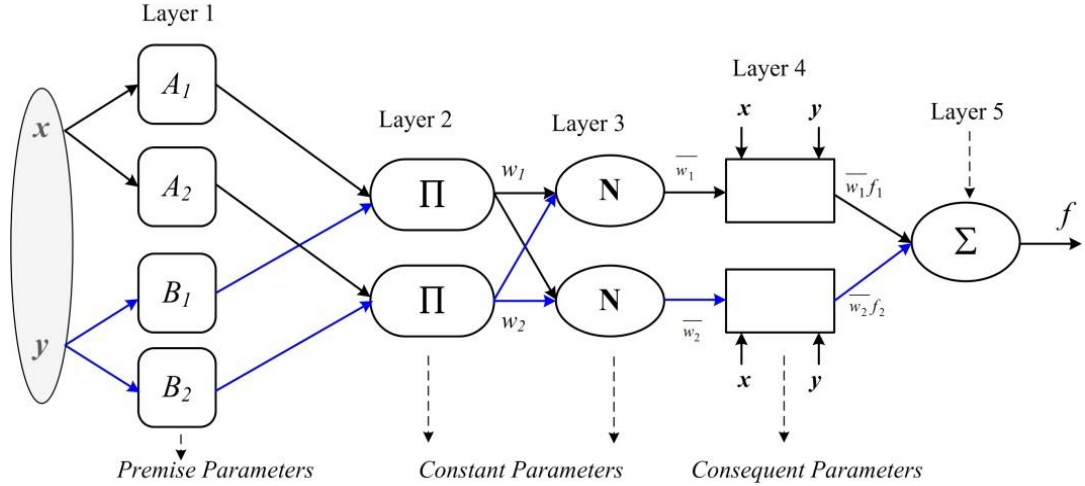


Fig. 2. Basic ANFIS architecture with two inputs and one output.

Layer 2: Every node in this layer will estimate of each incoming signal to each neuron using the product operator (Π) or T-norm operator. The output is the product of multiplying the incoming signals to entire node and demonstrates the firing strength of a rule.

$$O_{2,i} = w_i = \mu_{A_i}(x) * \mu_{B_i}(y), i = 1,2 \quad (3)$$

Layer 3: Every node in this layer labelled N will normalize all input signals to calculate the ratio of each i^{th} rule's firing strength to the sum of all rules firing strengths as

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad (4)$$

where, w_i is the firing strengths of the second layer, and \bar{w}_i is taken as the normalized firing strength.

Layer 4: Each node in the fourth layer calculates the contribution of the i^{th} rules toward the total output and the function is defined as

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (5)$$

where $(p_i x + q_i y + r_i)$ are the parameter set on first order of Sugeno fuzzy inference system model.

The parameters in this layer are referred to as consequent parameters.

Layer 5: The single node in this layer labelled Σ will summarize all the weighted output values from overall of all incoming signals.

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (6)$$

2.3. ANFIS Fuzzy C–Means clustering algorithm

Specifically, this work employed an ANFIS based on Fuzzy C–Means clustering algorithm (ANFIS FCM) for rainfall prediction. ANFIS FCM is one of the artificial intelligence algorithms that is proven successful in predicting the relationship between input and output parameters (Fattahi, 2016). Bezdek et al. (1984) has been introduced the FCM algorithm as a data clustering in which each data point belongs to a cluster to a degree specified by a membership grade. FCM partitioned the collection of n vector $Y_b, i = 1, 2, \dots, n$ into C fuzzy groups and finds a cluster center in each group, it then minimizes a cost function of inequality measure. The stages of FCM algorithm are (1) the cluster

centers, c_i , $i = 1, 2, \dots, C$ is randomly chosen from n points $\{Y_1, Y_2, Y_3, \dots, Y_n\}$. The membership of matrix \mathbf{U} is then computed following the equation:

$$U_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij}/d_{kj})^{\frac{2}{m-1}}} \quad (7)$$

where, $d_{ij} = \|c_i - c_j\|$ is the Euclidean distance between i th cluster center and j th data point, and m is the fuzziness index. Then, the cost function according to the equation (3) is computed. The process is stopped if it is below a certain threshold.

$$J(U, c_1, \dots, c_3) = \sum_{k=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d_{ij}^2 \quad (8)$$

Finally, a new C fuzzy cluster centers c_i , $i = 1, 2, \dots, C$ is computed using the following equation:

$$C_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \quad (9)$$

2.4. Data processing

In this study, meteorological parameters that were selected will be trained and tested to find which combinations of parameters have the best correlation with the highest accuracy to be selected as the input of model. The size of the data being trained is ten years with 70% for training and 30% for testing. The output of the system will be tested to match with the target given to the system. If not, then the learning algorithm will be improved by providing new network architecture or learning. The results of rainfall prediction through the ANFIS method are also evaluated with the RMSE (Root Mean Squared Error) and Mean Squared Error (MSE). Mathematically, RMSE and MSE are expressed as (Wang & Lu, 2018).

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2} \quad \text{and} \quad MSE = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \quad (10)$$

where \hat{y}_k is the prediction value, y_k is the actual value, and n is the data number.

Note that both RMSE and MSE express the average model prediction error in the unit of the observed variable. The MSE parameter is used when data contains unexpected values that must be considered, i.e. the value is too high or too low. In this case, rainfall data on certain days can reach extreme values. The value of both parameters can range from 0 to ∞ and ignores the direction of the error. Prediction results will obtained better if both values are close to zero. On the other hand, if RMSE is small enough, ANFIS is considered to have succeeded in carrying out the training process.

The detailed process of developing a rainfall prediction model using ANFIS, it can be applied as a system for early detection of flooding. For this work, three scenarios of combination input by using surface temperature, relative humidity, and partial pressure of water to predict rainfall are employed as shown in Fig. 3. The first scenario of learning process, it has been used one input, i.e. relative humidity (RH). The second scenarios is with two inputs (temperature average (T_{avg}) and RH). While the third scenario is using three inputs (minimum temperature (T_{min}), RH, and partial pressure of water vapour (P_w)). The selection of these parameters is based on trial and error to find which parameter is best as an intermediary. Overall, each output is predicted using ANFIS through the learning process of training, testing, and validation.

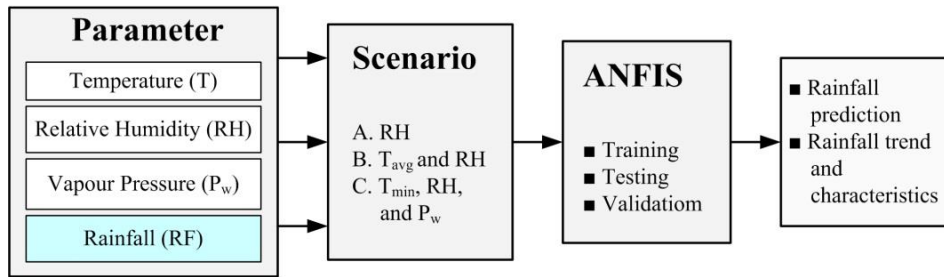


Fig. 3. The scenario of research to predict rainfall.

3. RESULT AND DISCUSSION

Based on the scenario in **Fig. 3**, the daily data for 10 years (2009 – 2018) is divided into 70% (2556 data) for training and 30% (1096 data) for validation. Furthermore, 70% of training data is also divided into 70% for training (1789 data) and 30% for testing (767 data). The training set is used to fit the model to see how the model learns from this data. While the validation set is used to evaluate a given model. On the other hand, the test set is used to provide an unbiased evaluation of a final model fit on the training dataset. It is only used once a model is completely trained (using the train and validation sets). The estimation results of rainfall that has been carried out with three scenarios (A, B, and C) are presented in **Fig. 4**. As shown in the figure, the rainfall prediction (blue) follows the trend of rainfall data (gray) in (a) training, (b) testing, and (c) validation. However, the daily rainfall from measurement is around 34% higher than the rainfall with the estimation technique using ANFIS.

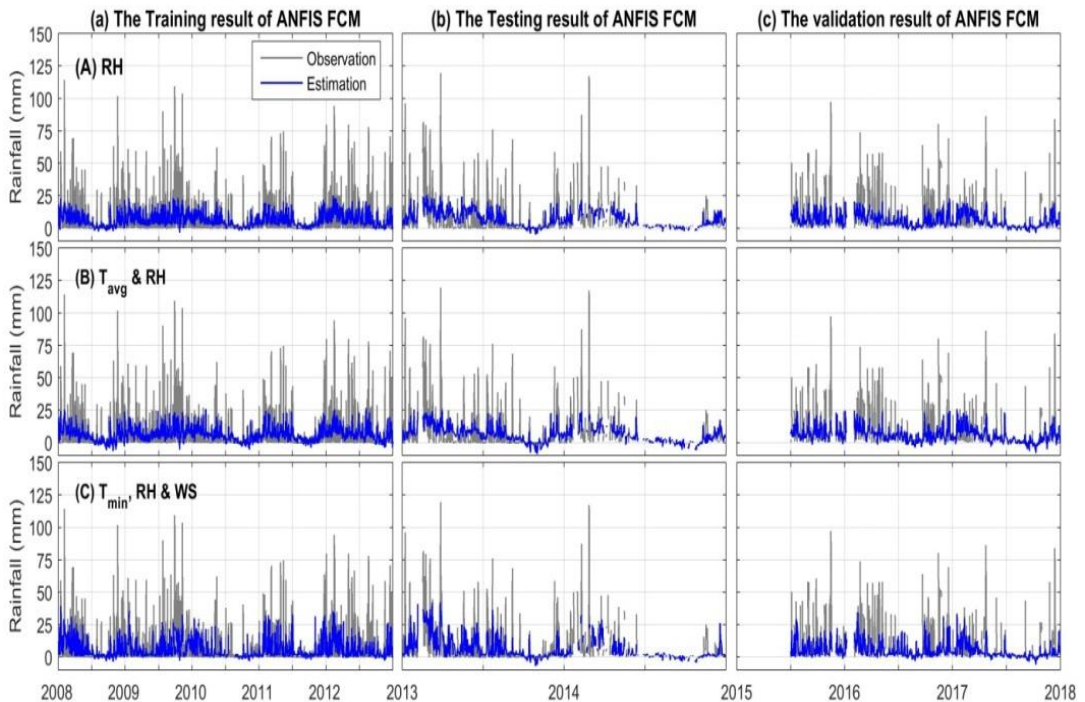


Fig. 4. Daily variation of rainfall estimated from ANFIS FCM model compared with daily rainfall data at BMKG Pondok Betung station during the (a) training, (b) testing and (c) validation process for the input of (A) RH, (B) T_{avg} and RH, and (C) T_{min} , RH, and P_w .

The relationship rainfall together with MSE values between observation and estimation for the three scenarios during the training, testing, and validation processes is depicted in **Fig. 5**. The result shows that overall correlation coefficients (r) to have a low relationship at the 99% confidence level with almost similar MSE values except for *scenario C* where during the training process shows a moderate relationship with a small MSE value. The figure also shows that good learning outcomes in the training process will also produce good estimates. In the case of using daily rainfall, it appears that the estimated daily rainfall value for the three scenarios is almost the same. With due regard to cost efficiency, measurement equipment, and time, with almost the same results, relative humidity can be considered as the best input model or intermediate parameter to estimate rainfall values.

Because the daily data in **Fig. 4** can only see rainfall trends between measurements and estimations and the correlation is also low, the monthly data is processed to show clearly the variations and differences in values between observations and estimates. **Fig. 6** shows the monthly average rainfall results in the training, testing and validation processes for each scenario. The figure also shows the correlation coefficient (r), MSE and the difference in rainfall between observations and estimated average (δ). It is clear that the rainfall data from the measurement is still higher around 28 mm than the estimated average. However, rainfall trends clearly follow one another. Note that the testing result show a strong correlation where the training result is with moderate relationship. Note that the testing results show a strong correlation while the training results show a moderate relationship.

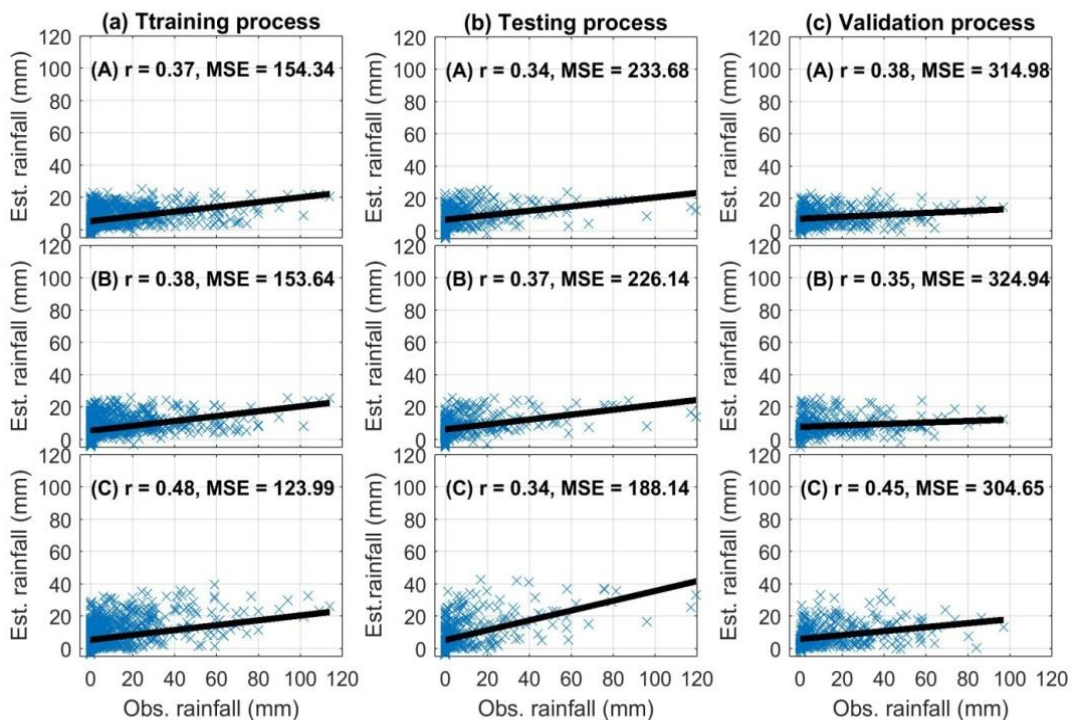


Fig. 5. The relationship of rainfall between observation and estimation for the three scenarios during the training, testing, and validation processes.

Table 1 shows the RMSE value obtained during the learning process for training, testing, and validation. It is seen that the RMSE value for daily and monthly data is similar for training and testing. In the validation process, the RMSE value for monthly is 38% lower than the daily RMSE. This indicates that the monthly RMSE value provides better accuracy of the predicted results.

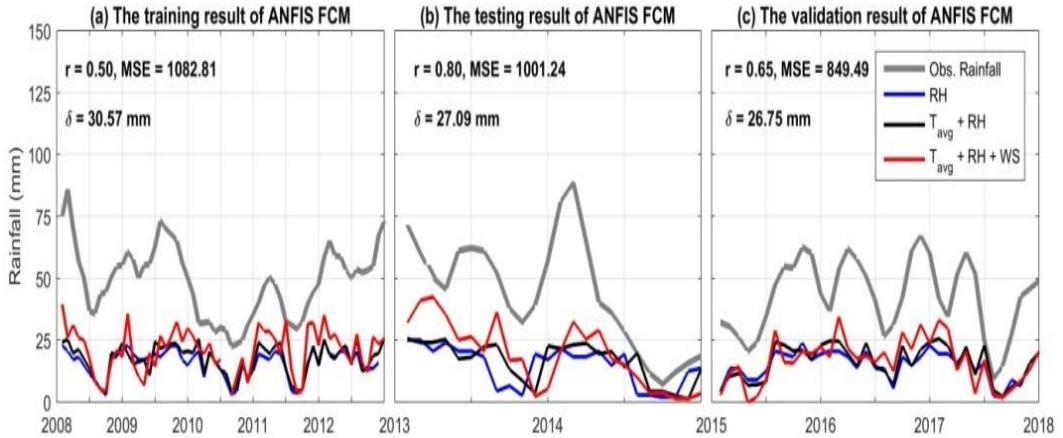


Fig. 6. Monthly average of variation of rainfall estimated from ANFIS FCM model compared with daily rainfall data at BMKG Pondok Betung station during the (a) training, (b) testing, and (c) validation process for the input of (A) RH, (B) T_{avg} and RH, and (C) T_{min} , RH, and P_w .

Table 1.

RMSE results from daily and monthly rainfall based on three scenarios (A, B, and C) during training, testing, and validation for a period of 10 years (2009 - 2018)

| Learning Process | The evaluation of the model in each scenario | | | | | |
|------------------|--|-------|-------|---------|-------|-------|
| | Daily | | | Monthly | | |
| | A | B | C | A | B | C |
| Training | 12.20 | 12.17 | 10.43 | 12.20 | 12.17 | 10.42 |
| Testing | 11.92 | 11.73 | 10.53 | 11.92 | 11.73 | 10.53 |
| Validation | 17.75 | 18.03 | 17.45 | 11.03 | 11.21 | 10.71 |

Focusing on daily analysis, in each case, the model learns well the sample data (training dataset) where the testing result (test dataset) is capable to capture all the training data in their estimation. This is indicated by almost similar RMSE in each scenario, except for the validation result where the bias or error is 32% higher as compared to training and testing results. Prediction of rainfall using scenarios A and B are different in error by 9% and 8% as compared to scenario C respectively. Overall, the combination input with T_{min} , RH, and P_w or scenario C is the best result for the prediction of rainfall in South Tangerang City.

5. CONCLUSIONS

Using the ANFIS Fuzzy-C-Means (FCM) technique, rainfall data on a daily basis in South Tangerang City, Indonesia has been successfully estimated. From all the scenarios, the RSME of training and testing and validation processes of prediction on a daily basis is around 11% and 18% respectively. The model is capable to estimate rainfall with accuracy more than 80%.

However, the model that have been built with membership structures and functions are still not able to capture or predict extreme rainfall. This condition is possible because the intermediate parameters chosen are only limited to the amount of data available namely temperature, humidity and partial pressure of water. In future, other meteorological parameters such as station surface pressure

and other parameters need to be inputted and combined with other existing parameters such as temperature and humidity and the results of the estimation can also be compared with hydrological data such as water level. At present, meteorological data in the form of surface pressure are not provided free of charge by BMKG. Further analysis will also look at what rainfall categories potentially triggering urban flash floods.

ACKNOWLEDGMENTS

This study is supported under the Internal Research Fund (001/PER-P2M/UPJ/05.19) by the Universitas Pembangunan Jaya. The authors highly appreciate the Meteorology, Climatology and Geophysics Agency (MCGA) of Indonesia or BMKG for providing surface meteorological data.

REFERENCES

- Bezdek, J. C., Ehrlich, R. & Full, W. (1984) FCM: Fuzzy C-Means Clustering Algorithm. *Computer & Geosciences*, 10(2-3), 191-203.
- Bui, D. T. et al. (2018) New Hybrids of ANFIS with several optimization algorithms for flood susceptibility modeling. *Water*, 10, 1210, doi:10.3390/w10091210.
- Fattahi, H. (2016) Adaptive neuro fuzzy inference system based on fuzzy c-means clustering algorithm, a technique for estimation of tbm penetration rate. *Int. J. Optim. Civil Eng.*, 6, 159-171.
- Fujiki, K., & Renard, F. (2018) A geographic analysis of post-disaster social impacts on a municipal scale – a case study of a potential major flood in the Paris region (France). *Geographia Technica*, 13 (2), 31-51.
- Ghazavi, R., Rabori, A. M. & Reveshty, M. A. (2016) Modelling and assessment of urban flood hazards based on rainfall intensity-duration-frequency curves reformation. *Nat. Hazards Earth Syst. Sci. Discuss.*, doi: 10.5194/nhess-2016-304.
- Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.-X., Chen, W., Kougiyas, I. & Kazakis, N. (2017) Flood susceptibility assessment in hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci. Total Environ.* 621, 1124–1141.
- Huang, M., Zhang, T., Ruan, J. & Chen. X. (2017) A new efficient hybrid intelligent model for biodegradation process of DMP with fuzzy wavelet neural networks. *Sci. Rep.*, 7, 41239, doi: 10.1038/srep41239.
- Jang, J. S. R. (1993) ANFIS: adaptive network-based fuzzy inference systems. *IEEE Transactions on Systems Man and Cybernetics*, 23, 665–685.
- Jang, J. S. & Sun, C. T. (1995) Neuro-fuzzy modeling and control. *Proceedings of the IEEE* 83(3), 378-406.
- Kompas (2018) Flood-prone points in South Tangerang.
”<https://megapolitan.kompas.com/read/2018/10/26/19180951/ini-titik-titik-rawan-banjir-di-tangerang-selatan>”
- Mekanik, F., Imteaz, M. A. & Talei, A. (2015) Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. *Clim. Dyn.* 1-15.
- Suparta, W., Abdul Rashid, Z. A., Mohd Ali, M. A., Yatim, B., & Fraser, G. J. (2008) Observations of Antarctic precipitable water vapor and its response to the solar activity based on GPS sensing. *Journal of Atmospheric and Solar- Terrestrial Physics*, 70, 1419-1447.
- Suparta, W. & Alhasa, K. M. (2013) A comparison of ANFIS and MLP models for the prediction of precipitable water vapor. *Proceeding of the 2013 IEEE International Conference on Space Science and Communication (IconSpace)*, 243-248.
- Suparta, W. & Alhasa, K. M. (2016a) *Modeling of Delays Using ANFIS*. Germany: Springer International.
- Suparta, W. & Alhasa, K. M. (2016b) Precipitable water vapor using an adaptive neuro-fuzzy inference system in the absence of the GPS network. *Journal of Applied Meteorology and Climatology* 55(10), 2283-2300, doi: 10.1175/JAMC-D-15-0161.1.
- Suparta, W. & Putro, W. S. (2017) Comparison of tropical thunderstorm estimation between multiple linear regression, Dvorak, and ANFIS. *Bulletin of Electrical Engineering and Informatics* 6(2), 149-158, doi: 10.11591/eei.v6i2.648.
- Tierra, F., (2016) Nonlinear and discontinuities modeling of time series using artificial neural network with radial basis function. *Geographia Technica*, 11 (2), 102 -112.
- Wang, W. & Lu, Y. (2018) Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP Conference Series: Materials Science and Engineering*, 324 (2018) 012049, doi:10.1088/1757-899X/324/1/012049.