ICIC Express Letters
Volume **13**, Number **2**, February **2019**

ICIC International ©2019 ISSN 1881-803X
pp. 113–118

SPATIAL-TEMPORAL APPROACH FOR PREDICTING RAINFALL IN TROPICAL COUNTRY

Indrabayu[1], Surya Aditama[1], A. Ais Prayogi[1], Saleh Pallu[2]
Andani Achmad[3] and Intan Sari Areni[3]

[1]Department of Informatics Engineering
[2]Department of Civil Engineering
[3]Department of Electrical Engineering
Faculty of Engineering, Gowa Campus
Universitas Hasanuddin
Jl. Poros Malino KM. 6, Bontomarannu, Gowa 92119, Sulawesi Selatan, Indonesia
{ indrabayu; andani; intan }@unhas.ac.id; { suryaadi4848; ais.prayogi; salehpallu }@gmail.com

Received August 2018; accepted November 2018

Abstract. *This research aims to develop a system that could predict rain in Makassar region based on spatial data from Ambon and Palembang with sampling resolution per 30 minutes. This spatial region is selected based on correlation of temperature and rainfall between Makassar with Ambon and Palembang. The processed rain data comes from Ambon, Palembang, and Makassar's historical data of meteorological conditions downloaded from wunderground.com with time span between 2009 and 2016. This research predicts rainfall in Makassar based on temperature, humidity, visibility, and dew point in Ambon and Palembang using extreme learning machine. The prediction process is done twice by using two different datasets; 2009-2014 data is used to predict rainfall in 2015 and 2009-2015 data is used to predict rainfall in 2016. The result of the research obtained average sensitivity of 83.99%, precision of 94.68%, and accuracy of 98.37% for 2015 and average sensitivity of 89.22%, precision by 93.62%, and accuracy of 98.62% for 2016 for 30 minutes later.*
Keywords: Rainfall prediction, Extreme learning machine, Spatial analysis, Correlation

1. **Introduction.** Climate change as a result of global warming caused the irregular seasonal changes in Indonesia. This can be seen within the cities affected by monsoon rainfall pattern such as Makassar. Rainfall level is one of the parameters related with the impact of climate change. Hence, in climatology and meteorology research, rainfall parameter becomes an important subject to predict.

Indrabayu has done several researches regarding rain prediction using different methods. In 2011, rainfall prediction in Makassar is done by using wavelet-neural network. The research was using local data of Makassar from the local Meteorological Climatological and Geophysical Agency (BMKG) with the accuracy of 81.09% [1]. In 2013, Indrabayu et al. proposed a new approach for rainfall prediction by combining Support Vector Machine (SVM) and fuzzy logic. The result of this research shows that SVM-fuzzy accuracy is higher than NN-Fuzzy's on predicting rainfall in 2009 and 2010 [2].

The next research by Indrabayu et al. is to compare ARIMA and ASTAR statistic methods with Genetic Algorithm-Neural Network (GA-NN) Artificial Intelligence (AI) method to predict rainfall which use empirical data. The result shows that ASTAR provides the best prediction compared with other methods in terms of Root Mean Square Error (RMSE) value in 2009 and 2010 which is 0.07 and 0.09 [3].

DOI: 10.24507/icicel.13.02.113

More efforts in implemeting advanced statistic methods were conducted in 2013. Two statistic methods, i.e., Auto Regressive Integrated Moving Average (ARIMA) and Adaptive Splines Threshold Autoregressive (ASTAR) were compared in terms of accuracy in predicting daily rainfall with the result of ASTAR has a better prediction than ARIMA. In 2010, the average of ASTAR's Root Mean Square Error (RMSE) is lower than ARIMA's that is 0.1373 and 0.2942 respectively [4].

Sreekanth et al. have done classification such between rain and storm with several rainfall parameters such as temperature, humidity, sea level pressure, wind speed, and wind direction. The dataset used is obtained from www.wunderground.com for 2010 from various locations with 362 samples. This research performed system performance analysis based on accuracy level with the used method that is Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The result of the research shows that accuracy level of ELM is higher than other methods [5].

A spatio-temporal for foracasting rainfall was also conducted in Australia [7]. The research shows thoroughly comprehensive classification of similar region to reduce complexity and only imply for semi-longterm that is monthly prediction. Another approach in spatio-temporal was also conducted in Spain with daily forecasting and mainly using rainfall gauge data [8].

Several researches about rainfall predictions had been done but most of them still focuse on temporal analysis without considering the effect of the parameter from other location of other station (spatial) with other multiple metherological parameters. Observing period of time is also long or semi-long, i.e, yearly, monthly and daily basis. Hence, this research predicts the rainfall in Makassar city based on spatial data or parameter from other cities such as Ambon and Palembang. The focus of this research is to determine the most suited parameters in predicting rainfall from eight meteorological parameters that are available and build rainfall prediction system with sampling resolution of 30 minutes.

The structure of this paper is as follows. In Section 2, the proposed system is described. In Section 3, the prediction results are discussed. And finally, Section 4 concludes the paper and discusses the future work.

2. **Proposed System.** The rainfall prediction system based on spatial data is done by using eight parameters such as rainfall, temperature, dew point, sea level pressure, humidity, visibility, wind speed, and wind direction. Those data are obtained from www.wunderground.com for data per 30 minutes from 2009-2016. Software used to predict rainfall is MATLAB 2016a. The proposed rainfall prediction system is shown in Figure 1.

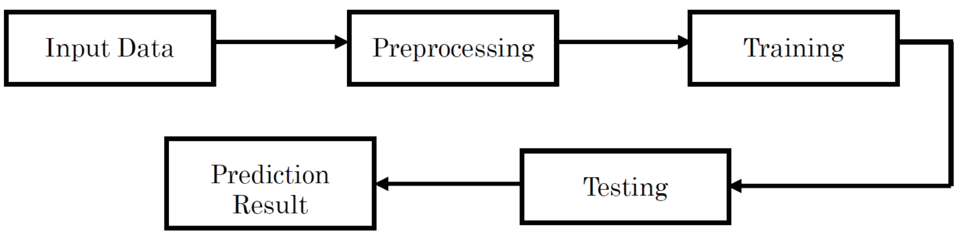

FIGURE 1. Block diagram of system's implementation

On preprocessing phase, the input data passed through correlation, interpolation, and normalization. Correlation is done to measure the strength of the relationship between the input variable and the expected output variables. The stronger the connection is, the more accurate the prediction result is. The relation between correlation coefficient and the correlation level is shown in Table 1.

After the correlation calculation between the parameters is done, then the variable correlation with the highest coefficient is temperature (T) followed by sea level pressure

Table 1. The relationship between coefficient and correlation level [6]

Correlation coefficient	Correlation level
$\rho = 1$	Positive perfect correlation
$0.80 \leq \rho < 1$	Very high correlation
$0.60 \leq \rho < 0.80$	High correlation
$0.40 \leq \rho < 0.60$	Medium correlation
$0.20 \leq \rho < 0.40$	Low correlation
$0.00 < \rho < 0.20$	Very low correlation
$\rho = 0$	Does not have any linear correlation

Table 2. Correlation of rain parameters

Precipitation Makassar	T	SLP	H	DP	V	WS-WD
Palembang	0.690761	0.4925	0.581028	−0.07732	0.057464	0.0591
Lagging	Lag −1	Lag −5	Lag 0	Lag 0	Lag 0	Lag 0
Ambon	0.503652	0.595495	0.401769	0.188952	−0.04035	0.0211
Lagging	Lag −1	Lag −8	Lag 0	Lag −4	Lag 0	Lag 0

(SLP), humidity (H), dew point (DP), visibility (V), wind speed (WS), and wind direction (WD) and then obtained the highest correlation between these variables to change lag −1 to lag −10 as in Table 2.

Interpolation is a way to find a value between known multiple data points. In real world, interpolation can be used to predict a function, where the function is undefined with a formula, but defined with data or table, i.e., table from experiment result; there are several interpolation based on the function, such as linear interpolation, quadratic interpolation, and polynomial interpolation. Normalization done in this research is using minimum-maximum transformation, so every input data is transformed to become within range of 0-1.

On training phase, training data used is 80% from the total data. The training is done to find the optimal weight and bias that is going to be used on testing process. The steps of training with ELM are described as below.

Step 1: Initialize all weight and bias with random numbers.

Step 2: If stopping condition unfulfilled, then do Step 3 to Step 7.

Step 3: Every input unit x_i $(i = 1, 2, \ldots, N)$ receives signal and forwards them to all hidden layer units.

Step 4: Calculate every hidden layer unit z_j $(j = 1, 2, \ldots, m)$ by adding up weighted input signals:

$$z\_net_j = b_j + \sum_{i=1}^{n} x_j w_{ji} \tag{1}$$

b_j is the bias for unit j of layer l and w_{ji} is the weight from unit j to the unit i.

Next, calculate output form hidden layer with activation function

$$g(z\_net_j) = \frac{1}{1 + e^{-z\_net_j}} \tag{2}$$

Step 5: Calculate matrix H with the size of $n \times m$.

$$H = \begin{pmatrix} g(w_{11} \cdot x_1 + b_{10}) & g(w_{12} \cdot x_1 + b_{20}) & \cdots & g(w_{1n} \cdot x_1 + b_{n0}) \\ g(w_{21} \cdot x_2 + b_{10}) & g(w_{22} \cdot x_2 + b_{20}) & \cdots & g(w_{2n} \cdot x_2 + b_{n0}) \\ g(w_{31} \cdot x_3 + b_{10}) & g(w_{32} \cdot x_3 + b_{20}) & \cdots & g(w_{3n} \cdot x_3 + b_{n0}) \end{pmatrix} \tag{3}$$

After obtaining matrix H, then calculate H^+ which is a pseudo-inverse matrix from matrix H that will be used on searching of weight value between hidden layer and output layer. Equation of H^+ can be seen in Equation (4).

$$H^+ = (H^T H)^{-1} H^T \tag{4}$$

Then find weight to output layer β.

$$\beta = H^+ t_i \tag{5}$$

t_i is the target of training process.

Step 6: Calculate the output value using the equation of:

$$\sum_{j=1}^{m} \beta_j g(z\_net_j) = y \tag{6}$$

Step 7: Calculate error value in output unit.

$$E = ||y - t|| \tag{7}$$

The optimal condition on training process is obtained when input unit 15 and hidden node 30 with RMSE (Root Mean Square Error) value of 0.0008608.

System performance's measurement is using confusion matrix where there are four result representation, that is True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TN value is the total of negative data detected correctly, while FP is a negative data detected as a positive data. Meanwhile, TP is a positive data detected correctly. FN is the opposite of TP, so the data will be positive, but detected as a negative one. Confusion matrix model is shown in Table 3.

Table 3. Confusion matrix

Class	Positive classified	Negative classified
Positive	TP	FN
Negative	FP	TN

Based on TP, FP, TN and FN value, we can get the accuracy value, precision and sensitivity (recall) by using Equations (8) to (10) respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

3. **Results and Discussion.** The performance of rainfall prediction is using confusion matrix. For 2015 prediction, train data from 2009-2014 is used. Meanwhile for 2016 prediction, train data from 2009-2015 is used. The confusion matrix of 2015 and 2016 prediction result is shown by Table 4 and Table 5 respectively.

Based on confusion matrix on Table 4 then the values of sensitivity, precision and accuracy are obtained, based on Equations (8)-(10) shown in Table 5 for 2015 & 2016 prediction.

The results are shown by the percentage rate of the prediction accuracy of 98.37% and 98.62% respectively for the 2015 and 2016 predictions showing that the system designed with this ELM algorithm has worked well. In the future, spatial prediction systems can be done by piloting in several areas around Makassar.

Table 4. Confusion matrix of 2015 & 2016 prediction result

Prediction 30 minutes later (2015)					Prediction 30 minutes later (2016)			
Confusion Matrix	Sunny	Light	Medium	Heavy	Sunny	Light	Medium	Heavy
Sunny	18470	55	0	0	18484	49	0	0
Light	190	1256	4	0	230	1274	19	0
Medium	0	53	228	0	0	44	203	3
Heavy	0	0	28	61	0	0	3	36

Table 5. Sensitivity, precision and system accuracy value analysis for 2015 & 2016 prediction

	Results of analysis 30 minutes later (%) for 2015			Results of analysis 30 minutes later (%) for 2016		
	Precision	Sensitivity	Accuracy	Precision	Sensitivity	Accuracy
Sunny	98.98	99.70	98.37	98.77	99.73	98.62
Light	92.08	86.62		93.19	83.65	
Medium	87.69	81.13		90.22	81.20	
Heavy	100.00	68.53		92.30	92.30	
Average	94.68	83.99		93.62	89.22	

4. **Conclusions.** This rainfall prediction system with Extreme Learning Machine (ELM) method is using five from eight available types of metrological data based on their level of correlation that is rainfall, temperature, humidity, dew points, and visibility. Architecture model of neural network on Makassar's weather condition consists of 15 units of input neurons on input layer, 30 and 40 units of hidden neurons on hidden layer, and 1 unit output (sunny, light rain, medium rain, and heavy rain) on output layer. For 2015, the accuracy level is obtained by 98.37%. Meanwhile for 2016 prediction, the accuracy level is obtained by 98.62%.

Sampling resolution used in this study is data per 30 minutes and provides high accuracy. Future research is expected that sampling used in rain forecasting such as Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Indonesia can use a short-term sampling resolution. A faster sampling resolution will provide a more accurate meteorological element correlation and lead to better prediction results.

Acknowledgment. This work is partially supported in part by the Universitas Hasanuddin (UNHAS) via the PDUPT-Ristekdikti Grant, Indonesia.

REFERENCES

[1] Indrabayu, Rainfall prediction in Makassar region using wavelet neural network method, *Jurnal Elektrikal Enjiniring*, vol.9, pp.50-59, 2011.
[2] Indrabayu, A. Achmad, M. S. Pallu and N. Harun, A new approach of expert system for rainfall prediction based on data series, *Int. J. Eng. Res. Appl. (IJERA)*, vol.3, pp.1805-1809, 2013.
[3] Indrabayu, A. Achmad, M. S. Pallu and N. Harun, Statistic approach versus artificial intelligence for rainfall prediction based on data series, *Int. J. Eng. Technol. (IJET)*, vol.5, pp.1962-1969, 2013.
[4] Indrabayu, A. Achmad, M. S. Pallu and N. Harun, Numerical statistic approach for expert system in rainfall prediction based on data series, *International Journal of Computational Engineering Research*, vol.3, pp.81-87, 2013.
[5] M. S. Sreekanth, R. Rajesh and J. Satheeshku, Extreme learning machine for the classification of rainfall and thunderstorm, *J. Appl. Sci.*, vol.15, pp.153-156, 2015.
[6] *Weather Underground*, About Our Data [WWW Document], Weather Undergr. URL /about/data (accessed 11.6.17).

[7] M. Montazerolghaem, W. Vervoort, B. Minasny and A. McBratney, Spatiotemporal monthly rainfall forecasts for south-eastern and eastern Australia using climatic indices, *Journal of Theoretical and Applied Climatology*, vol.124, pp.1045-1063, 2016.

[8] A. F. Militino, M. D. Ugarte, T. Goicoa and M. Genton, Alternative spatio-temporal models for daily rainfall prediction in Navarre, *METMAVII Workshop*, 2014.