A Monsoon Onset and Offset Prediction Model Using Backpropagation and Moron Method: A Case in Drought Region

Syeiva Nurul Desylvia
Department of Computer Science
Bogor Agricultural University
Bogor, Indonesia
syeiva.nd@gmail.com

Taufik Djatna
Postgraduate Program of Agroindustrial Technology and Graduate Program of Computer Science
Bogor Agricultural University
Bogor, Indonesia
taufikdjatna@ipb.ac.id

Agus Buono
Department of Computer Science
Bogor Agricultural University
Bogor, Indonesia
pudesha@gmail.com

Abstract—**First day (onset) and last day (offset) of monsoon are ature phenomena which are important elements at cultivation tages in agriculture. These 2 sets of time value influent harvest erformance but it is difficult to predict onset and offset at lrought region. One of technique that can be used to solve nentioned problem is prediction technique which is one of data nining task. In this research, Feed Forward Backpropagation BPNN) were combined with Moron method to predict onset and ffset at drought region. Data used were daily rainfall data from 983 to 2013. This experiment used 2 kind of BPNN models and hey used 5 different values for learning rate (alpha) from range .01 to 0.2. Root Mean Square Error (RMSE) is used to evaluate esulted prediction models along with correlation value and tandard deviation of error for better understanding. For BPNN nset model, lowest RMSE value at alpha 0.15 is 32,0546 and owest RMSE value for BPNN offset is 26,6977 at alpha 0.05. Developed model has been able to use for prediction, but the result vas still not close enough to actual data. In order to achieve a etter model with lower RMSE, it is neccesary to improve model rchitecture and to specify some methods to obtain certain umber of input layer based on Southern Oscillation Index (SOI) lata.**

Keywords—Rainfall; Backpropagation; Prediction; Moron nethod.

I. Introduction

Indonesia is a tropical climate country with only 2 seasons vhich are monsoon and summer. In drought region, it is difficult o predict first day of monsoon (onset) and last day of monsoon offset) in Indonesia. Data mining task can be used to predict nset and offset in Indonesia. There are exist some of data nining tasks (e.g. classification, outlier analysis, clustering nalysis, and mining frequent pattern) and the most suitable task or prediction is classification task. Classification technique is a rocess to discover a model or a function which generalize or istinguish data classes or concepts. Model is constructed based n analysis of training data (labeled sample data). Resulted nodel can be used to predict unlabeled sample data (testing data) 1].

Patterns about onset and offset at each year are important in cultivating sector. The farmers need to know when the start and the end of monsoon are. It is because they need to make a plan for their crops such as what kind of crops that suitable are, when they are going to start cultivating the crops, or at which day the crops will be ready.

There are some researches concerning rainfall prediction. Awan and Maqbool [2] constructed a rainfall prediction model in Pakistan using Artificial Neural Network (ANN) namely Backpropagation and Learning Vector Quantization. The result showed advantage of neural network than it was of statistical model. Modeling for rainfall prediction also had been implemented in India [3]. In this research, a model based on ANN was compared with a model based on regression. As a result, ANN model was better than regression model. Abhishek et al. [4] compared 3 kinds of ANN for rainfall prediction modeling. Modeled ANNs were Feed Forward Backpropagation, Layer Recurrent, and Cascaded Feed Forward Backpropagation. As the outcome of their research, Feed Forward Backpropagation algorithm was better based on resulted MSE. Wan et al. [5] modeled Backpropagation (BP) for rainfall prediction at JiangXi Province, China. Preprocessing technique applied the combination of k-means clustering method and discriminant analysis. This study used 6 extreme rainfall index (PQ90, PX5D, PINT, PFL90, PNL90, PXCDD) as predictor. Based on resulted BP performance, constructed model had been capable to predict rainfall well.

Another research had been done by Nikam and Meshram [6]. In this research, Bayesian Prediction was used to construct rainfall prediction model and the data was climate data from Indian Meteorological Departement (IMD) Pune. The data attributes were Temp, Station Level Pressure, Mean Sea Level Pressure, Relative Humidity, Vapor Pressure, Wind Speed, and Rainfall. Converting numeric data into categorical data based on given range (discretization) was one of preprocessing technique that was implemented. The best resulted accuracy was 96.15% at Delhi City dataset. Further research is rainfall prediction using Fuzzy Inference System [7]. The data were METAR 20 years data at Cairo airport station (HECA) [1972 – 1992] and METAR

 978-979-1421-23-2/15/$31.00

5 years data at Mersa Matruh station (HEMM). Parameters which were used as input were relative humidity, total cloud cover, wind direction, temperature, and surface pressure. For evaluation model, this research applied rainfall cases data and 6 hours-before data. The outcome based on Brier Score (BS) and Fraction Skill Score (FSS) showed that the constructed model had been sufficient for rainfall prediction. This research still needed to be developed, for example, to add input parameter and to combine FIS method with ANN.

Models constructed at some of mentioned researches had been capable to predict rainfall well. However, at those researches, on which day monsoon starts and on which day monsoon ends have not been discovered yet. Those researches are only predicting rainfall and measuring model performance based on resulted prediction. Therefore, in this research, models to predict monsoon onset and to predict monsoon offset was developed.

The objective of this research is to construct Backpropagation Neural Network (BPNN) model along with Moron Method [8] as feature extractor to predict monsoon onset and offset. In this case, Waingapu historical data was deployed. Waingapu is one of city at Nusa Tenggara Timur, Indonesia. BPNN was used as classifier to predict on which day monsoon onset and offset in Waingapu are. Root Mean Square Error (RMSE) was applied to evaluate the model prediction performance. The data was temporal rainfall data at Waingapu and Southern Oscillation Index (SOI) data, which was used as predictor.

In order to achieve the outcome, the first step was feature extraction using Moron Method for getting monsoon onset and offset, the second step was the implementation of correlation method to determine the number of BPNN model input layer, and the last step was BPNN modeling for monsoon onset and offset prediction with SOI data as predictor. BPNN in combination with Moron Method can be used to predict monsoon onset and offset in Waingapu with sufficiently low error rate (RMSE). Also, the combination of BPNN and Moron method can indicate at which day the first day of monsoon and at which day the last day of monsoon are. The first day and the last day of monsoon are difficult to predict but this developed BPNN model can be utilized to predict them.

This paper is organized as follow: Section 2 describe the method that is used in this research, section 3 describe the result and discussion of this research, and section 4 describe the conclusion.

II. METHOD

The method consist of 4 main steps, that is, data collection, model construction, and evaluation. Fig. 1 shows flowchart of mentioned method.

Data collection. The data in this research are:

1) Daily rainfall data at Waingapu City obtained from BMKG (Badan Meteorologi, Klimatologi dan Geofisika), from 1983 to 2013. There are 71 rain monitoring stations and 5 data attribute namely Tanggal, Tahun, Bulan, TGL, and Hujan. Attribute used only Hujan which is temporal numeric data.

2) Southern Oscillation Index (SOI) data is obtained from National Oceanic Atmospheric Administration (NOAA) and can be downloaded from http://www.esrl.noaa.gov/.

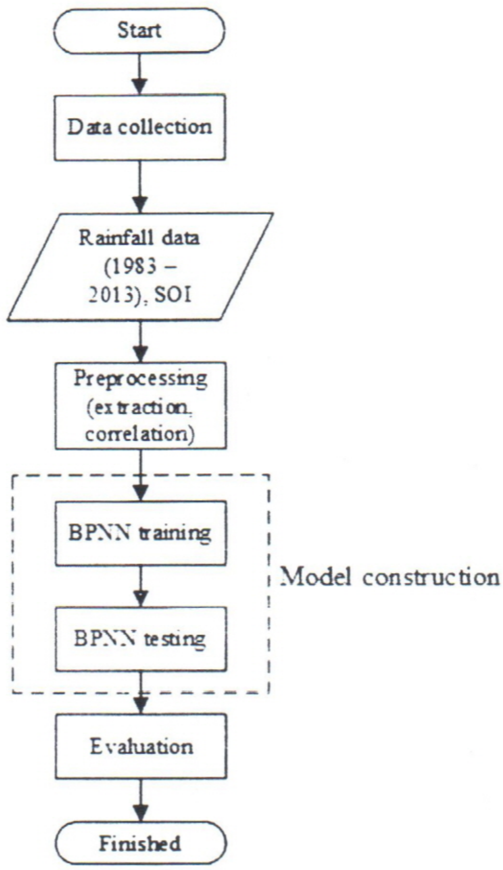

Fig. 1. Example of a figure caption. *(figure caption)*

Preprocessing. Preprocessing step is used to prepare data before be utilized as input for the model. This process consists of feature extraction to determine onset (the first day of monsoon) and offset (the last day of monsoon) using Moron method and the selection of months as BPNN input model using correlation. First process is to specify onset using Moron method that consists of several steps. In this technique, onset determined by accumulating rainfall in 5 days. Onset will be start if accumulated rainfall in 5 days is more than 40 mm. Next step, accumulation is conducted again for consecutive 10 days since onset is determined. If during 30 days since onset is choosen has accumulated value at least 5 mm, onset determined before is selected as onset (not canceled). These same steps are conducted again for determining offset but rainfall data is reversed. Next process is to correlate onset and offset with SOI data (only August to July). The first day of monsoon in Waingapu is September so June is choosen as model input because June is 3 months before monsoon. The last day of monsoon in Waingapu is March so, with the same method, December is choosen. Then, resulted correlation is conducted to determine how many months are used as input at BPNN model (how many months before the month that has been selected).

Model construction. This research applies Backpropagation Neural Network (BPNN) also called multilayer perceptron. A multilayer perceptron with 1 hidden layer (unit Z) is indicated by Fig. 2. At the network, output unit (unit Y) and hidden unit have bias. Bias at output unit Y_k is denoted by w_{0k} and bias at hidden unit Z_j is denoted by v_{0j}. Those biases act as weight at connection from unit which it's output always 1. BPNN training consists of 3 steps that are feed forward from input training data, error backpropagation computation, and adjusting weight. After training process, only feed forward step is applied to the network. More than 1 hidden layer might be usefull for some applications but 1 hidden layer is enough for most of applications. BPNN activation function should have some important characteristics such as continuous function, derivable, and monotonly not decreased. For computation simplicity, derived activation function should be convenient to compute. The most commonly used activation function are binary sigmoid and bipolar sigmoid. [9]. Random weight initialization is used in the beginning of BPNN training. In this research, 5 different learning rate (alpha) values is tested to obtain alpha value with minimum resulted error. The data that is used to predict is the same data that is used to train (training data).

Evaluation. Root Mean Square Error (RMSE) is used to measure model performance. RMSE is square root of MSE [10]. The MSE of an estimator $\hat{\theta}$ with respect to an unknown parameter θ is defined as:

$$\mathrm{MSE}=E\left[\left(\hat{\theta}-\theta\right)\right]=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i-Y_i\right)^2 \tag{1}$$

$\hat{Y}_i$ is vector from n predictor and Y_i is vector from actual value. RMSE defined as:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i-Y_i)^2} \tag{2}$$

III. Result And Discussion

Preprocessing. This step yields onset and offset which are obtained using Moron method. Onset and offset is indicated at Fig. 3 and Fig. 4. Based on those Figures, onset and offset value is fluctuating and some values are far different from other values. Those different values might be extreme weather.

Next step is to compute correlation between onset and offset with SOI data from August to July. Table 1 denotes resulted correlation.

Correlation value between SOI data at June with onset is low so SOI data which is used as input model is July and August data. Then, correlation value between SOI data at December with offset is sufficient so SOI data which is used as input model is December, October and November.

Model construction. There are 2 BPNN models which are developed namely onset prediction model and offset prediction model. Those models differ at the number of input layer. Table 2 shows the architecture of mentioned models.

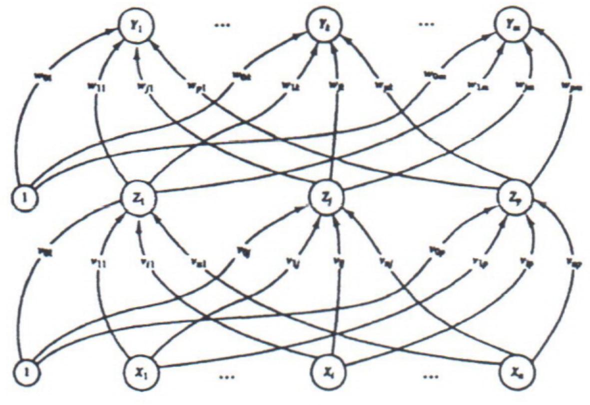

Fig 2 Backpropagation neural network [9]

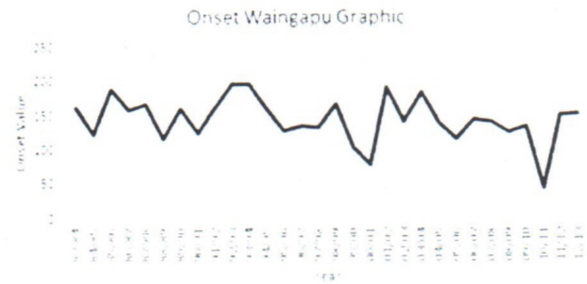

Fig. 3 Onset at 1983 to 2013

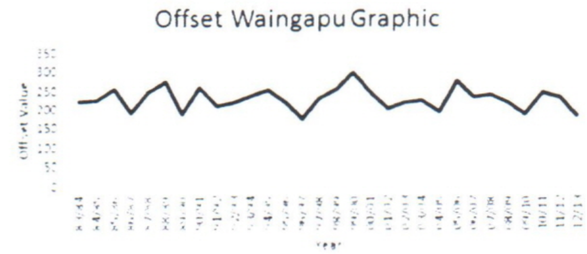

Fig 4 Offset at 1983 to 2013

For those models, activation functions are purelin and tansig and training function is trainlm. There are 5 different values for learning rate (alpha) parameter in this research. They are 0.01, 0.05, 0.1, 0.15, and 0.2. Fig. 5 indicates prediction result of BPNN onset model and Fig. 6 indicates prediction result of BPNN offset model with alpha 0.01. Based on those Figures, both models are capable to predict onset and offset subject to actual data. On the other hand, model architecture still needs to be advanced further in order to achieve prediction result closer to actual data. To measure interrelationship between actual data and prediction result, Table 3 represents correlation between each model with actual data and Fig. 9 indicates plot of the same information. It can be concluded from correlation result that the diverse of alpha values range tested in this research needs to be improved. This development is based on correlation values which are similar to each other. Therefore, in the future work, it is necessary to increase alpha values range to obtain better interrelationship between actual data and prediction result.

Additionally, in this research, 0.15 is the best alpha value because it resulted the highest correlation value in both models.

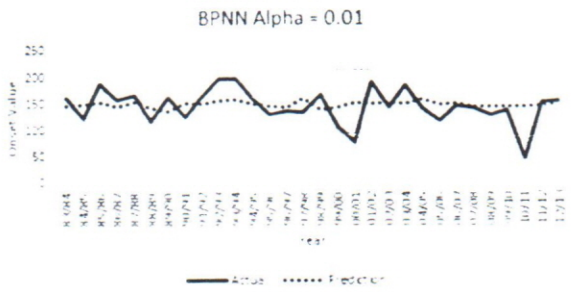

Fig. 5. Resulted onset prediction with alpha 0.01

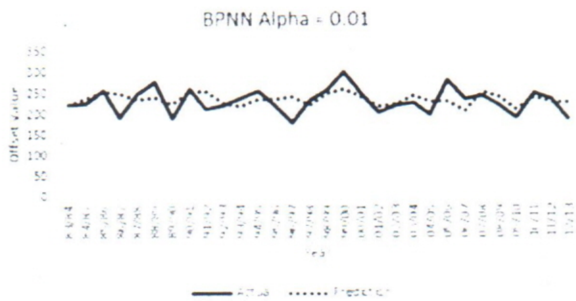

Fig. 6. Resulted offset prediction with alpha 0.01

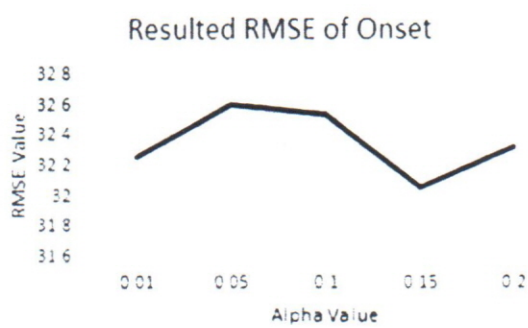

Fig. 7. Onset RMSE

Evaluation. RMSE, from (2), is used to evaluate each model. Fig. 7 and Fig. 8 denote RMSE of BPNN onset model and offset model. Based on resulted RMSE, the best onset model is at alpha 0.15 with RMSE 32.0546 and the best offset model is at alpha 0.05 with RMSE 26.6977. Furthermore, standard deviation of error is also utilized to estimate model performance. Error obtained in this study is subtraction between actual data with each model prediction result. Table 4 presents standard deviation of error for both models and Fig. 10 also shows the same information. Coming from standard deviation of error result, the amount of dispersion for each model is similar and the data points tend to spread out over a wider range of values. This similar value for each model is identical to correlation value at Table 3 and at Fig. 9, so it can be analyzed that standard deviation of error result is related to alpha values range. Then, it is possible to reduce error value by increasing diversity of alpha value range.

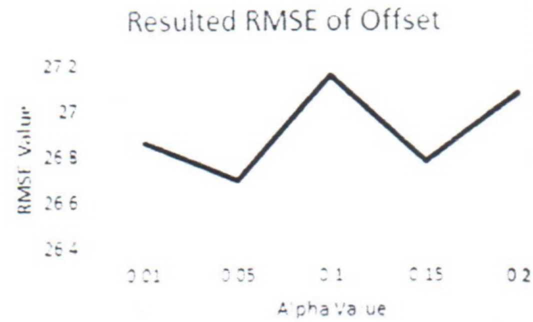

Fig. 8. Offset RMSE

TABLE I. CORRELATION BETWEEN ONSET AND OFFSET WITH SOI DATA

Month	*Onset*	*Offset*
Aug	-0.24	-0.10
Sep	-0.44	0.06
Okt	-0.41	0.22
Nov	-0.46	0.08
Dec	-0.53	0.26
Jan	-0.46	0.40
Feb	-0.43	0.43
Mar	-0.40	0.44
Apr	-0.38	0.40
May	-0.43	0.44
Jun	-0.51	0.45
Jul	-0.04	0.17

TABLE II. BPNN MODELS ARCHITECTURE

Model	*Input Layer*	*Hidden Layer*	*Output Layer*
Onset	2	15	1
Offset	3	15	1

Moreover, Fig. 11 and Fig. 12 indicates Taylor Diagram of onset and offset models as another representation for better understanding. For each Figure, alpha value 0.01, 0.05, 0.1, 0.15, and 0.5 are represented respectively by blue dot, green dot yellow dot, brown dot, and gray dot. At each Figure, the dots that represent each model are close to each other and their correlation values are around 0.2 for onset and 0.4 for offset.

TABLE III CORRELATION BETWEEN EACH MODELS RESULT AND ACTUAL DATA

Alpha	*BPNN Onset Model*	*BPNN Offset Model*
0.01	0.2526	0.3959
0.05	0.2243	0.4005
0.10	0.2580	0.3902
0.15	0.2739	0.4132
0.20	0.2575	0.4073

Fig 9 Correlation of Actual Data with Prediction Result

TABLE IV STANDARD DEVIATION OF ERROR FOR EACH MODEL

Alpha	*BPNN Onset Model*	*BPNN Offset Model*
0.01	32.6550	27.0095
0.05	32.7849	27.1149
0.10	32.7903	26.9980
0.15	32.3604	26.9061
0.20	32.6434	26.7805

Fig 10 Standard Deviation of Error for Both Models

Both models are successful to follow the actual pattern so far, but they still needs to develop further to obtain a better result with higher correlation, lower RMSE, and lower standard deviation of error. Subject to correlation result, RMSE value, and standard deviation of error analysis, it is fundamental to refine model architecture and alpha value variations. It is also a possibility of extreme weather in one or more year in Waingapu which affected onset and offset prediction result.

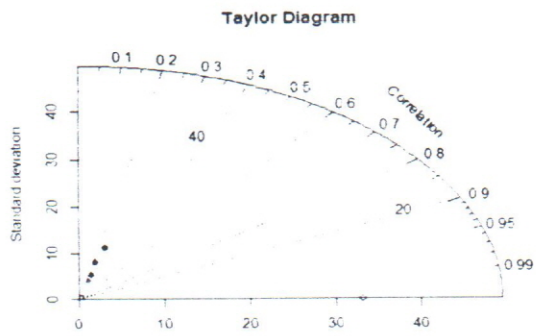

Fig 11 Taylor Diagram of BPNN Onset Model

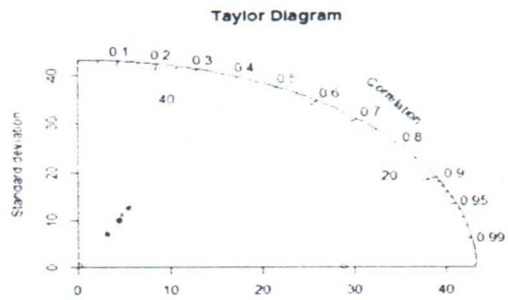

Fig 12 Taylor Diagram of BPNN Offset Model

IV. CONCLUSION

The constructed models have been able to predict the first day (onset) and the last day (offset) of monsoon. Higher diverse in parameter model and improvement in model architecture are potentially advance model performance. The resulted RMSE values from each model indicate that the model has been not adequate yet to predict rainfall onset and offset based on historical data in drought region. So, there is a possibility of good RMSE result if the models is tested based on historical data in high rainfall rate.

This research still needs developing further to obtain better models. In future work, several parameter variation, particularly alpha value, will be tested along with enhancement of model architecture. Besides, some specific methods to choose SOI data months as input model will be considered for model improvement.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Morgan Kauffman, 2012

[2] J. A. Awan and O. Maqbool, "Application of artificial neural networks for monsoon rainfall prediction," in *6th International Conference on Emerging Technologies (ICET)*, 2010

[3] R.P. Shukla, K.C. Tripathi, A.C. Pandey, and I.M.L. Das, "Prediction of indian summer monsoon rainfall using niño indices: A neural network approach," *Atmospheric Research*, 2011.

[4] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," in *IEEE Control and System Graduate Research Colloqium (ICSGRC 2012)*, 2012.

[5] D. Wan, Y. Wang, N. Gu, and Y. Yu, "A novel approach to extreme rainfall prediction based on data mining," in *2nd International Conference on Computer Science and Network Technology*, 2012.

[6] V.B. Nikam and B.B. Meshram, "Modeling rainfall prediction using data mining method," in *Fifth International Conference on Computational Intelligence, Modelling, and Simulation*, 2013.

[7] S.A. Asklany, K. Elhelow, I.K. Youssef, and M. Abd El-wahab, "Rainfall events prediction using rule-based fuzzy inference system," *Atmospheric Research*, 2011.

[8] V. Moron, A.W. Robertson, and R. Boer, "Spatial coherence and seasonal predictability of monsoon onset over Indonesia," *Journal of Climate*, 2009.

[9] L. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice Hall, 1994.

[10] J. Devore, *Probability and Statictics for Engineering and The Sciences*. Cengage Learning, 2010.