

A TIME-DELAY CASCADING NEURAL NETWORK ARCHITECTURE FOR MODELING TIME-DEPENDENT PREDICTOR IN ONSET PREDICTION

Agus Buono, Imas Sukaesih Sitanggang, Mushthofa and Aziz Kustiyo

Department of Computer Science, Bogor Agricultural University, Bogor, Indonesia

Received 2013-12-07; Revised 2013-12-16; Accepted 2014-01-29

ABSTRACT

The occurrence of rain before the real start of a rainy season often mislead farmers into thinking that rainy season has started and suggesting them to start planting immediately. In reality, rainy season has not started yet, causing the already-planted rice seed to experience dehydration. Therefore, a model that can predict the onset of rainy season is required, so that draught disaster can be avoided. This study presents Time Delay-Cascading Neural Network (TD-CNN) which deals with situations where the response variable is determined by a number of time-dependent inter-related predictors. The proposed model is used to predict the onset in Pacitan District Indonesia based on Southern Oscillation Index (SOI). The Leave One Out (LOO) cross-validation with series data 1982-2012 are used in order to compare the accuracy of the proposed model with the Back-Propagation Neural Network (BPNN) and Cascading Neural Network (CNN). The experiment shows that the accuracy of the proposed model is 0.74, slightly above than the two other models, BPNN and CNN which are 0.71 and 0.72, respectively.

Keywords: Southern Oscillation Index, Rainy Season, Onset, Back-Propagation, Cascading Neural Network

1. INTRODUCTION

Empirical data has suggested that agricultural productivity in Indonesia is affected by climate variability. The occurrence of extreme climate conditions has been increasing since 1950, both in frequency and intensity, as reported in the International Disaster Database (Boer and Perdinan, 2008), giving rise to large loss, in particular, to the national rice productivity. A total of around 80% of disasters in rice agriculture are caused by flood and drought conditions, respectively at around 10 and 70% (Boer and Subbiah, 2005). The remaining disasters are caused by other factors, including pests. The research result by Pasaribu *et al.* (2010) shows that national rice production has lost about 1.2 million tones, due to drought in 2007. According to the data in Pacitan on the time frame from 1982 to 2009 indicates that around 90% of drought-related disasters occurred in the drought

season (May, June, July and August) while the rest occurred in the early rainy season (November and December). This is an interesting fact, because theoretically, water should be in abundance during this period. Further investigations upon this phenomena showed that drought occurred when there is a delay in the onset of rainy season which is preceded by a false rain. This phenomenon occurred when there is a heavy rain around the expected onset of rainy season and farmers intuitively concluded that rainy season has started, but it fact it has not and rain has not occurred over the next 2 dasarians (tens-of-days). This led to the condition where the already-planted rice fields to experience water-deficiency, leading to crop failure. This has occurred on 1997, 2006 and 2007 leading to a drought of over 3000 hectares of rice fields in Pacitan district Indonesia. **Figure 1** shows how this has occurred; taking as an example the 1997 case with a total area of drought equals 3595 hectares.

Corresponding Author: Agus Buono, Department of Computer Science, Bogor Agricultural University, Bogor, Indonesia

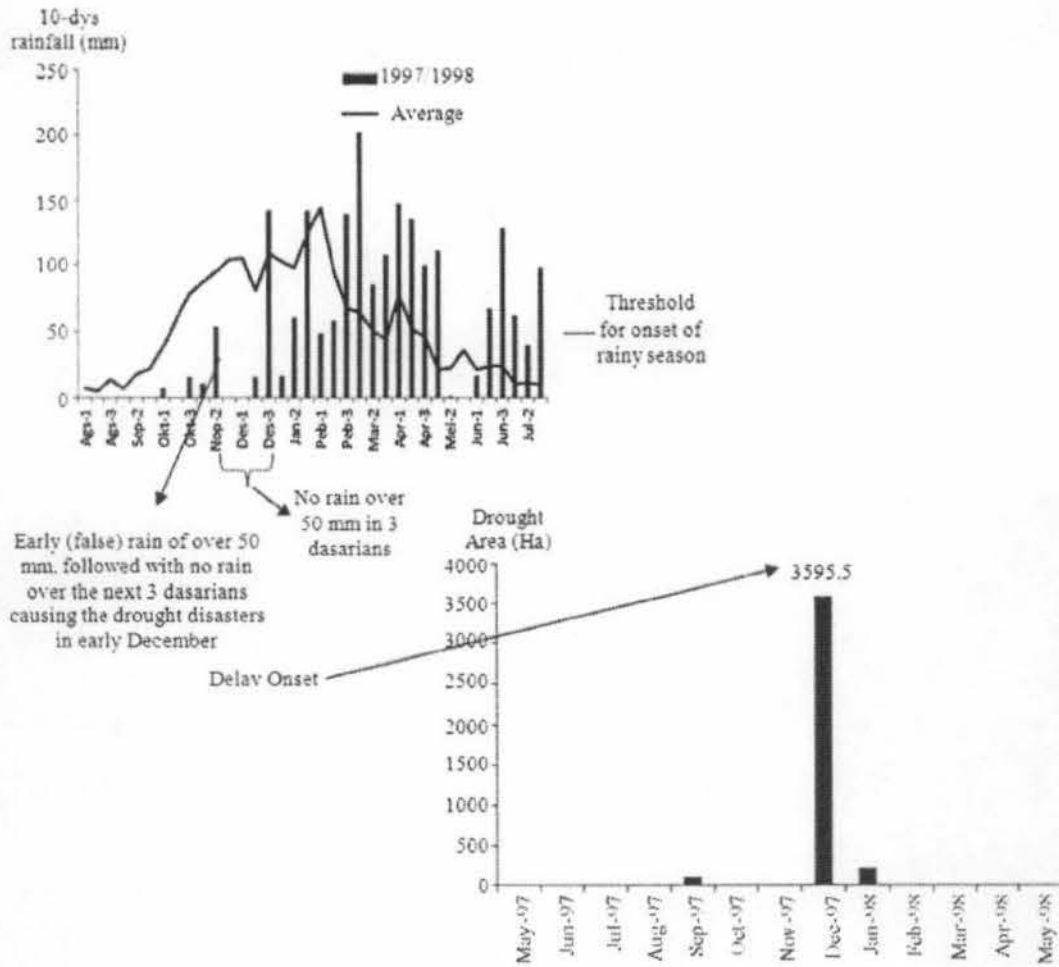


Fig. 1. Drought disaster due to false rain in Pacitan district Indonesia (year 1997)

Based on the above facts, it has been deemed necessary for the government and relevant agencies to provide accurate information regarding the onset of rainy season to mark the start of the planting of rice fields. With such information, it is hoped that farmers will start the planting only when it is certain that the onset of rainy season has been reached, therefore avoiding the disaster of crop failure due to drought.

To help the government in determining the onset of rainy season, a system is required that can predict the onset accurately. Robertson *et al.* (2007) showed that the onset of rainy season in Indonesia is closely related to global climate phoneme, one of which can be measured used the Southern Oscillation Index

(SOI). Research conducted by Buono *et al.* (2012) showed that artificial neural networks models using SOI as predictors are able to predict the onset of rainy season in the district of Indramayu with correlation results of up to 0.8. Other research conducted by Larasati (2012) showed that models using Support Vector Regression (SVR) are able to predict the onset of rainy season in the same district with correlation results of 0.7. Meanwhile, Buono and Mushthofa (2012) used fuzzy inference systems with SOI as predictors and obtained a correlation result of 0.68.

The shortcoming of the previous models was that they do not accurately reflect the natural characteristics of the underlying system. Global

climate indexes as well as the onset of rainy systems are phenomena which contain the aspect of time series and lag time. The developed models do not accommodate such characteristics. In this study, an artificial neural network model is developed from the cascaded multi-layer perceptron to model the time series aspect of SOI as predictor in predicting the onset. This model is intended to elaborate the characteristic of time lag from the existing data, in the hope that it will provide a better accuracy compared to previous models. To compare the performance of the proposed system, this study will also present the result of the experiment using the models trained using the Back-Propagation Neural Network (BPNN) and the Cascading Neural Network (CNN).

In what follows, the paper will be structured according to the following: Part 2 described the research method used, the data as well as the processing performed. The next part, Part 3 will be focused on the discussion regarding the research results. Finally, Part 4 will provide concluding remarks for this research.

2. DATA AND RESEARCH METHOD

2.1. Data

In this study, the predictor variable used to predict the onset is the Southern Oscillation Index (SOI). SOI data are obtained from <http://www.bom.gov.au/climate/current/soihtml.shtml>. SOI is an index which reflects the condition of the Pacific Ocean in comparison with the ocean around Indonesia. The choice of using this index is based on the fact that seasons in Indonesia are affected by the condition of the Pacific Ocean.

The monthly data of SOI are determined based on the pressure differences between Tahiti and Darwin Fig. 2, formulated as follows (AGBM, 2013):

$$SOI = 10 * \frac{(P_{diff} - P_{diffav})}{SD_{diff}}$$

Where:

- Pdiff = (average Tahiti MSLP for the month)-(average Darwin MSLP for the month),
- Pdiffav = long term average of Pdiff for the month in question and
- SD (Pdiff) = long term standard deviation of Pdiff for the month in question.

MSLP = Mean Sea Level Pressure

The response variable to be predicted is the onset of rainy season. According to the Indonesian Bureau of Meteorology, Climatology and Geophysics (BMKG), the onset of rainy season is defined by the threshold on the total precipitation of at least 50 mm over the course of one dasarian (ten days) and followed by the next dasarian. Therefore, the variable to represent the onset of rainy season can be measured in terms of dasarians. Figure 3 illustrated the process of determining the onset of rainy season. It can be seen that on the second dasarian of December (35th dasarian), the rain volume is above 50 mm and continued on for the next few dasarians. Therefore, on that year, the onset of rainy season occurred on the 35th dasarian. On the second dasarian of November, the rain volume is above 50 mm, but not for the next few dasarian and hence we can conclude that the onset did not occur at the second dasarian of November.

From Fig. 4, it can be seen that the anomaly of rainy season onset from the year of 1982 to 2009 varies from -5 (ahead 5 dasarians) and +5 (delayed 5 dasarians). Based on this data, the average of onset falls in the 32th dasarian (second dasarian of November). This means that the onset ranges at the earliest from the 27th dasarian (third dasarian of September) and at the latest on the 37th dasarian (first dasarian of January).

According to empirical data, the onset can be delayed or advanced from the normal (average) values. The advancement and delay of the onset depends on the variation of the condition of the Pacific Ocean, which in turn is reflected by the values of the SOI.

2.2. Research Method

Figure 5 shows the research methods block diagram. The method consists of six steps: Problem formulation, data collection, data analysis, neural network model building (TD-CNN, BPNN and CNN), testing the model and finally result discussion and documentation. The BPNN model used in this layer is the multi-layer perceptron with one hidden layer and training using the standard back-propagation training algorithm as described in (Fauset, 1994). Meanwhile, the CNN and the training algorithm is according to the description presented in (Engelbrecht, 2007).

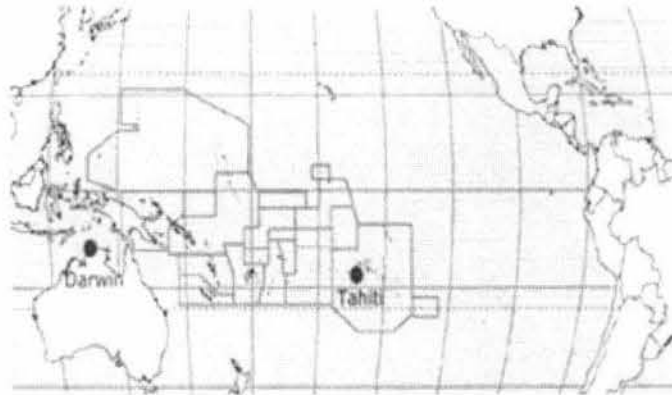


Fig. 2. Position of Darwin and Tahiti in SOI Calculation

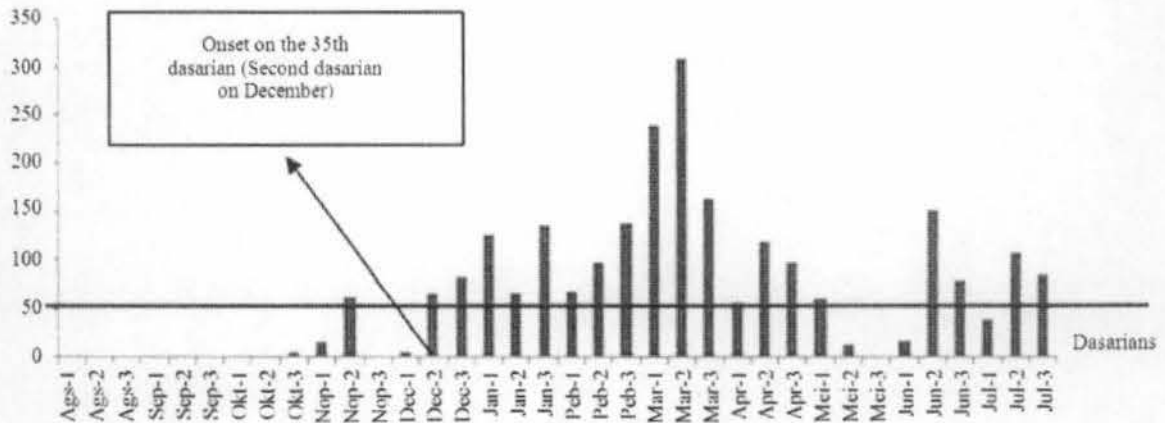


Fig. 3. Determining the Onset

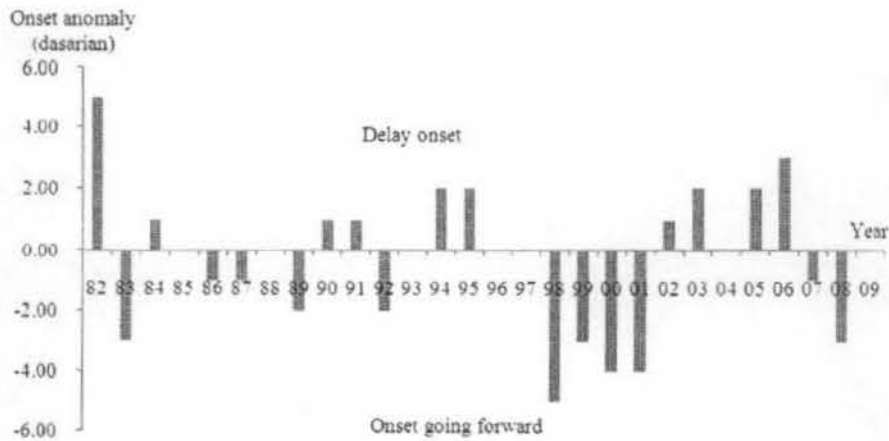


Fig. 4. Variability of onset data

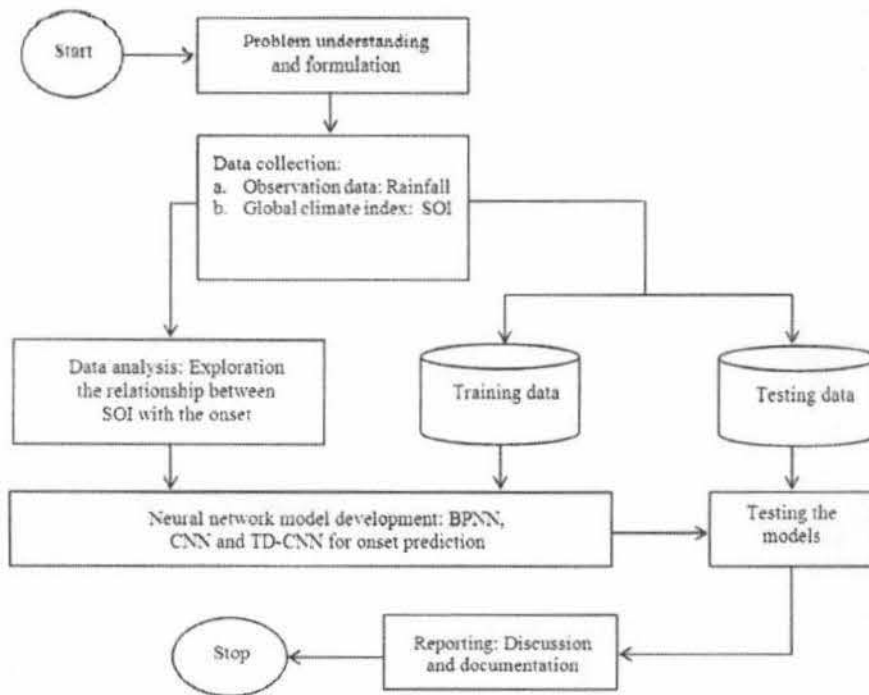


Fig. 5. Research Methods Block Diagram

3. RESULTS AND ANALYSIS

In this section, we present the relationship among the considered variables, model architecture and compare their accuracies.

3.1. Relationship of the Considered Variables

Based on the available data, it can be seen that SOI gives a significant influence on the onset of rainy season, as shown in Fig. 6. The first one is the apparent significant correlations between the onsets with SOI in the months of May to October. This means that the onset can be modeled using the SOI variables of May to October. However, based on this data, the onset usually falls in October, which makes the appropriate time to predict the onset around August or beginning of September. Therefore, the SOI data in the months of September and October cannot be used for prediction, since the SOI data of September can only be obtained in around the start of October and similarly for the October, the SOI is available in November. To wait for these data to be available would mean that it is too late to do the prediction.

The second result is that there is an apparent time series relation between the values of SOI. This fact allows us to build the model for predicting the onset using the

SOI data of the months of May to August, while still taking into account the SOI of September and October. The strategy employed here is to perform a cascading between SOI prediction models based on the SOI of the previous months. Therefore, the first model to represent the time series relationships between the SOI data will have an aspect of time-delay. The neural network architecture which allows such relationships is the cascading neural network. Therefore, the model called the Time-Delay Cascading Neural Network (TD-CNN) will be built and its performances will be compared to the standard models: BPNN and CNN.

3.2. TD-CNN Architecture and Algorithm

Based on the reasoning explained in Section 3.1, the structure of the TD-CNN to be built is presented in Fig. 7. To ease the formulation of the training algorithm, we define the following terminology for the network:

- u_{1j} = The weights between the SOI of May ($i = 1$) to SOI of August ($i = 4$) and bias 1 ($i = 0$) to neuron $H1_j$, with $j = 1, 2, 3, \dots, p$
- u_{2j} = The weights between the SOI of May ($i = 1$) to SOI of September ($i = 5$) and bias 3 ($i = 0$) to neuron $H2_j$, with $j = 1, 2, 3, \dots, p$

- u_{3ij} = The weights between the SOI of May ($i = 1$) to SOI of October ($i = 6$) and bias 5 ($i = 0$) to neuron H_{3j} , with $j = 1, 2, 3, \dots, r$
- w_{1i} = The weights between SOI of May to SOI up to neuron layer H1 and bias 2 to neuron SOI_{Sep} , with $I = 0, 1, 2, 3, \dots, 4+p$
- w_{2i} = the weights between SOI of May to SOI up to September, neuron layer H2 and bias 4 to neuron SOI_{Okt} , with $I = 0, 1, 2, 3, \dots, 5+p$
- w_{3ij} = The weights between SOI of May to SOI of October, neuron layer H3 and bias 6 to neuron of AMH (onset, with $I = 0, 1, 2, 3, \dots, 6+p$ and $j = 1, 2, 3, \dots, 14$ (the number of weather stations)

Training for this network used the back-propagation (Fauset, 1994) with adding some link in order to make it suitable to the model being constructed.

3.3. Accuracy

To perform testing for the prediction method used, we need to divide the data we have into training and testing data. One of the ways this can be performed is by

using the LOO (Leave One Out) Cross Validation. The advantage of using this method is that we can simulate the condition as if all data are used for testing. This step is important, since there are not many data available for this research and to obtain a new data, we have to wait for at least a year. In a LOO, all instances will have the chance of being a training data and testing data.

Figure 8 provides the comparison of accuracies of the three models visually. The interesting thing to see from the above figure is that in BNN, the predicted values are not linearly correlated with the observed values. This is in contrast with the CNN and TD-CNN results. In CNN, the predicted values have linear correlation with the observed value. However, the error values are still too large. In TD-CNN, we can see a linear correlation between the predicted and the observed values. However, we can see that there is a systematic deviation. This shows that if we had performed a calibration of the predicted values of TD-CNN, then its prediction results would be better than the other two methods.

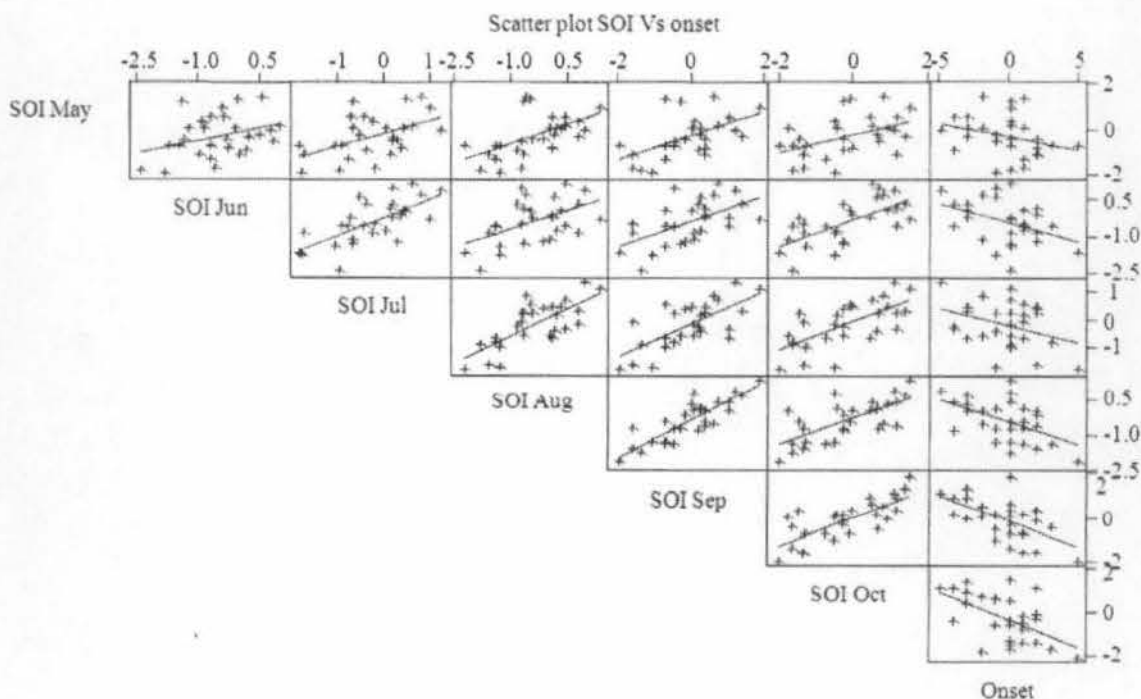


Fig. 6. Scatter plot of SOI Vs onset in Pacitan district

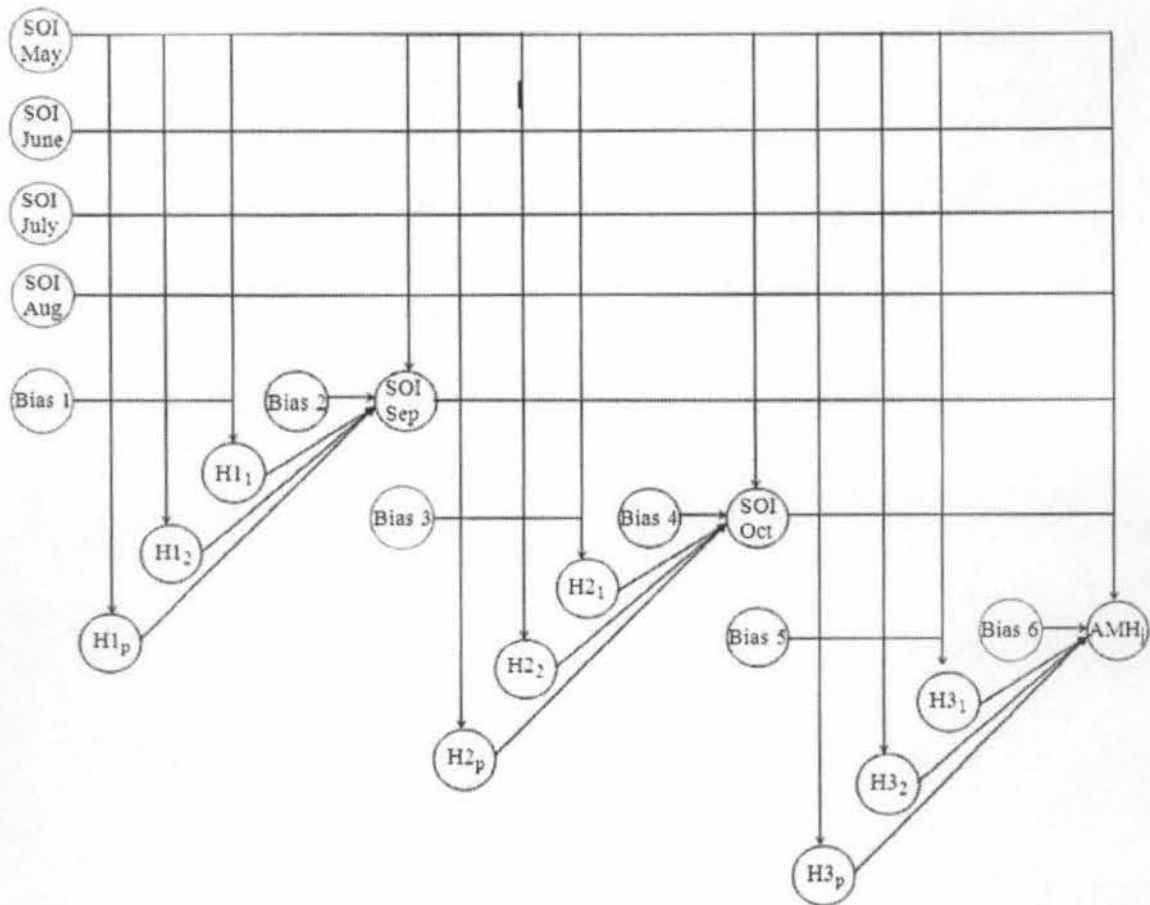


Fig. 7. TD-CNN architecture

This is due to the fact that the predicted values of TD-CNN have a linear correlation to the observed values and the error values of the TD-CNN are smaller than the ones in CNN. The correlation value for TD-CNN is also the largest, reaching 0.74. In addition, the correlation value for BNN and for CNN is 0.71 and 0.72, respectively. The previous works that applied the fuzzy method, support vector regression and the BNN using the training method Levenberg-Marquardt result the correlation value of 0.68, 0.7 and 0.8, respectively (Buono *et al.*, 2012). This study shows that the TD-CNN method gives the comparable correlation value i.e., 0.74.

Figure 9 illustrates the cause of the 1997 drought disaster in Pacitan. On that year, the rainfall on the 30th dasarian (end of November) reached more than 50 mm, which caused the farmer to assume that the

rainy season has started and began planting their rice fields. Observational data showed that over the next 6 dasarians, only one dasarian has a rainfall above 50 mm. Therefore, the rice fields experience water deficiency up to 3595 Ha. From this figure, we can conclude that on the first dasarian (or 37th, beginning of January) rainfall reaches over 50 mm and continued on for consistently for the almost all the next dasarian on the rainy season. According to the definition by BMKG, we can conclude that the (real) onset of rainy season occurred on the beginning of January. From the result of this experiment, the predicted onset of rainy season for this particular year is on the 36th dasarian. Therefore, if only this information had been available to the farmers, the drought disaster of year 1997 could have been avoided.

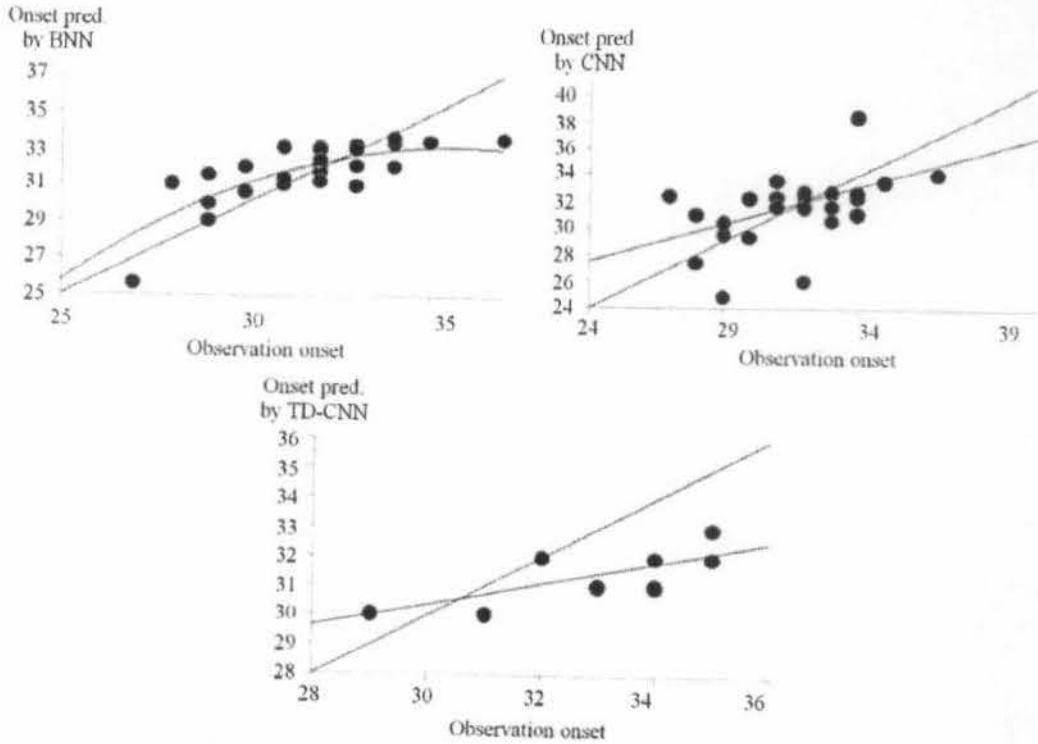


Fig. 8. Scatter plot of predicted onset Vs observation onset by BNN, CNN and TD-CNN

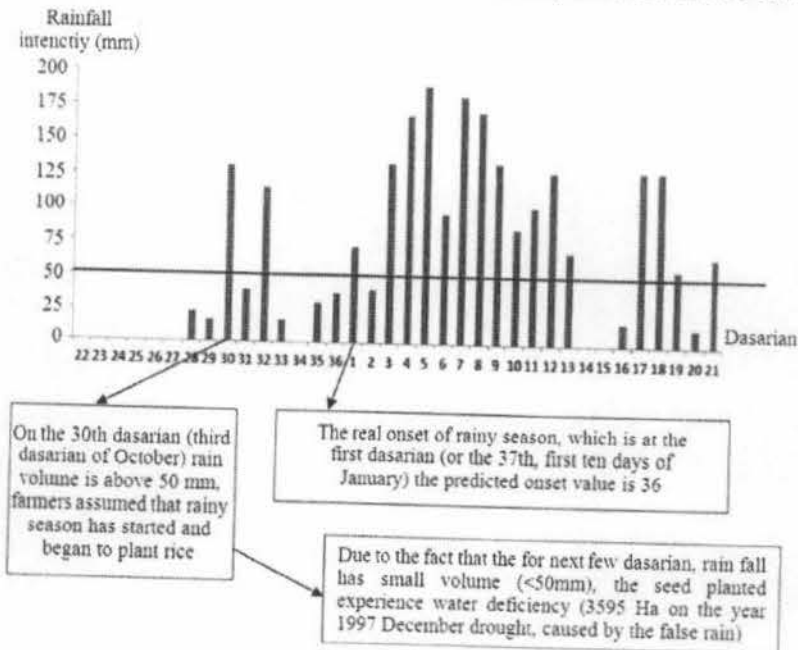


Fig. 9. A logical explanation of the drought disaster in 1997

4. CONCLUSION

Based on the experiment, it can be concluded that the TD-CNN model has a higher accuracy, which is 0.74, compared to the other two existing standard models, the BPNN and CNN, with accuracies 0.71 and 0.72, respectively. This is due to the fact that in the TD-CNN model, the response variable, which is the onset of rainy season, is also formulated using SOI in September and October, which is closer to the onset. While on the other two models, the response variables can only be formulated using SOI up to August.

This research has two limitations. Firstly, the global parameters that are used as predictors have not represented all global conditions influencing climate in Indonesia, particularly in Java Island. Secondly, the parameters for modeling have not given optimum results yet. Therefore, the future works include increasing the number of predictor variables and to perform a more effective model parameters selection. The predictor variables that can be used is the index that represents the condition of the Hindian Ocean which also affects the climate in Indonesia (especially Java island). To obtain more effective model parameters, we will use optimization techniques based on empirical data in estimating the model's parameters.

5. ACKNOWLEDGEMENT

The researchers thanks to the Directorate of Higher Education, Indonesia for financial support and the Centre for Climate Risk and Opportunity Management in Southeast Asia and Pacific (CCROM-SEAP), Indonesia for providing the rainfall data.

6. REFERENCES

- AGBM, 2013. Climate glossary. Australian Government, Bureau of Meteorology.
- Boer, R. and A.R. Subbiah, 2005. Agriculture drought in Indonesia. In: *Monitoring and Predicting Agricultural Drought: A Global Study*, Boken, V.K., A.P. Cracknell and R.L. Heathcote (Eds.), Oxford University Press, New York, ISBN-10: 019516234X, pp: 330-344.
- Boer, R. and Perdinan, 2008. Adaptation to climate variability and climate change: Its socio-economic aspect. Proceeding of Workshop on 'Climate Change: Impacts, Adaptation and Policy in South East Asia, Economy and Environmental Program for Southeast Asia.
- Buono, A. and Mushthofa, 2012. An implementation of fuzzy inference system for onset prediction based on southern oscillation index for increasing the resilience of rice production against climate variability. Proceedings of the International Conference on Advanced Computer Science and Information Systems, Dec. 1-2, IEEE Xplore Press, Depok, pp: 281-286.
- Buono, A., A. Kurniawan and F. Ahmad, 2012. Prediction the beginning of the wet season based on southern oscillation index using neural network. Proceedings of the Seminar Nasional Aplikasi Teknologi Informasi, Jun. 15-16, Yogyakarta, pp: B-27-B-32.
- Engelbrechth, A.P., 2007. Computational Intelligence: An Introduction. 2nd Edn., John Wiley and Sons, Ltd., England, ISBN-10: 0470035617, pp: 628.
- Fauset, L.V., 1994. Fundamentals of Neural Networks: Architectures, Algorithms and Applications. Prentice Hall, Englewood Cliffs, ISBN-10: 0133341860, pp: 461.
- Larasati, R., 2012. Prediction of rainy season onsets with southern oscillation index by using support vector regression. Indonesian Scholars International Convention, London.
- Pasaribu, S.M., I.A. Setiajie, N.K. Agustin, E.M. Lokollo and H. Tarigan *et al.*, 2010. Development of rice farming insurance to handle 75% of risk in rice production due to the flood, drought and disease. Res. Report, Center Social Analy. Economic Agricultural Policy.
- Robertson, A.W., A.V.M. Ines and J.W. Hansen, 2007. Downscaling of seasonal precipitation for crop simulation. *J. Applied Meteorol. Climatol.*, 46: 677-693. DOI: 10.1175/JAM2495.1