

APPLICATION OF MULTIVARIATE ANFIS FOR DAILY RAINFALL PREDICTION: INFLUENCES OF TRAINING DATA SIZE

Edvin Aldrian¹, and Yudha Setiawan Djamil²

1. Agency for the Assessment and Application of Technology (BPP Teknologi), Jakarta 10340, Indonesia
2. The INSIDE Project Research Associates, Agency for the Assessment and Application of Technology (BPP Teknologi), Jakarta 10340, Indonesia

E-mail: edvin@webmail.bppt.go.id

Abstract

This study investigates the use of multi variable Adaptive Neuro Fuzzy Inference System (ANFIS) in predicting daily rainfall using several surface weather parameters as predictors. The data used in this study comes from automatic weather station data collected in Timika airport from January until July 2005 with 15-minute time interval. We found out that relative humidity is the best predictor with a stable performance regardless of training data size and low RMSE amount especially in comparison to those from other predictors. Other predictors shows no consistent performances with different training data size. Performances of ANFIS reach a slightly above 0.6 in correlation values for daily rainfall data without any filtering for up to 100 data in a time series. The performance of ANFIS is sensitive to the magnitude and scale differences among predictors, thus suggesting introducing a transforming and scaling factor or functions. Application of multivariate ANFIS is relatively new in Indonesia. However, results presented here indicate some promises and possible roadmaps for improvements.

Keywords: ANFIS, humidity, rainfall

1. Introduction

Rainfall is a stochastic process, whose upcoming event depends on some precursors from other parameters such as the sea surface temperature for monthly to seasonal time scales, the surface pressure for weekly to longer than daily time scale and other atmospheric parameters for daily to hourly time scale. The latter atmospheric parameters could be temperature, relative humidity and winds. Variability of weather and climatic factors, especially those atmospheric parameters will be the major forcing for daily precipitation event. If we could recognize such a variability pattern and use it for future trajectory, daily rainfall prediction is very much feasible. One method to understand the underlying pattern is by using a pattern trainer using a neural network system until reaches the minimum trained errors. Presently, such facility is possible by using a combination of a neural network and Fuzzy logic system.

Beside those parameters, rainfall in Indonesia is determined mainly by local processes such as the orography that control the diurnal pattern and land atmosphere interaction due to the land cover type. The combination of land sea interaction and orographic effect will introduce convective type of precipitation,

which is torrential type (pour heavily within a very short period) and hence creates difficulty in rainfall prediction. One of such difficulty in predicting daily rainfall is also experienced in PT Freeport, Papua with its complex topography.

The concession area of the PT Freeport Indonesia is located in Timika, south Papua Island that has complex weather characters, while the area spans from the coastal swamp into high glacier mountain (Front of north wall glacier). This area experiences high amount of rainfall all year round (more than 300 rainy days) and has the monsoonal pattern [1]. It is also mentioned that this area receives high rainfall amount all year round and with no significant differences between dry and rainy seasons and is categorized as the non-seasonal area by the Indonesian Bureau of Meteorology (BMG). That BMG report also mentions about local orographic effect due to local topography. Thus, there is a need to conduct local climate assessment that could provide comprehensive picture of the local climate character quantitatively in term of spatial as well as temporally. The location of Timika, which is closed to the Pacific Ocean, is also affected by the global weather pattern such as the El Nino Southern Oscillation (ENSO), although the mountainous barrier in the middle of the Papua Island isolates Timika as the shadow area of large phenomena in the Pacific.

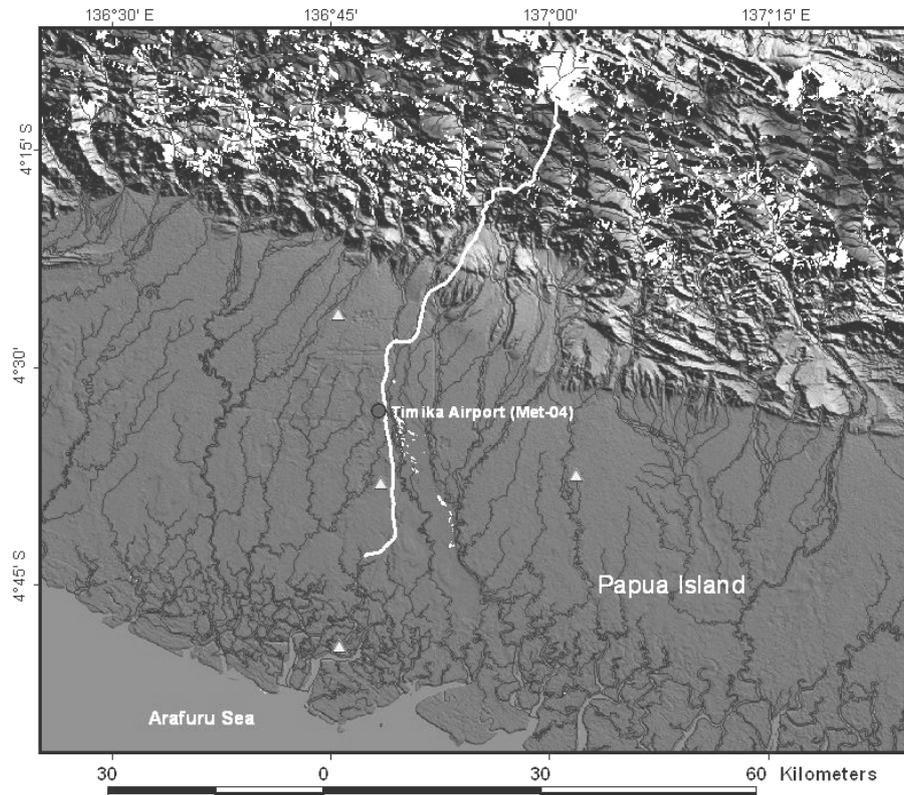


Figure 1. Location of the observation site or the Timika Airport in the PT Freeport concession area along with other automatic weather stations (triangle marks)

Variety of weather condition will have implication to the activity of the exploration, transportation, communication etc. by the PT Freeport Indonesia in Timika. For that reason, it is compulsory to apply a local weather prediction with local weather data that has been acquired since 1991 in 10 locations (Figure 1) distributed in the concession area of the PT Freeport Indonesia and use them for optimal engineering and management works. For that reason, climate assessment will help to identify appropriate predictor parameters for local weather prediction, while there is possibility of different prediction system for the dry and wet season.

In this study, we apply multi parameter Adaptive Neuro Fuzzy Inference System (ANFIS) for daily rainfall prediction in a location of PT Timika using series data of relative humidity, temperature, pressure and rainfall. Moreover, we use the hindcast method to predict and validate daily rainfall using historical record. Basic assumption made here is that rainfall will not generate its own variability and fluctuation but rather the other ambio parameters. Limitation of this study is only a single location of surface parameters (1 dimension) is used, thus the result will disregard spatial relationships with the surrounding. The application of multi variate ANFIS is relatively new in Indonesia. Others use rainfall record to predict rainfall [2-4] or runoff record to predict runoff [5], we will present some results and

discuss possible discrepancies and potential use of ANFIS for daily rainfall prediction.

2. Data and Method

Data. Main data used in this study comes from an automatic weather station, which consists of many in-situ atmospheric surface parameters such as the precipitation amount, relative humidity, temperature, surface pressure, wind direction and magnitude and solar radiation. Only the first four parameters among those are used in this study and they are displayed in Figure 2. The data were recorded by a logger at 15-minute temporal resolution at the AWS Met-04 in the Timika airport (4.52°S, 136.88°E, elevation 37 m) in the concession area of PT Freeport, Timika. We collected data over 184 days from 1 January until 5 July 2005.

From Figure 2, we could identify some local climatic condition according to the precipitation, RH and temperature and pressure data. The maximum precipitation rate occurs between 16 and 22 WITA and there are some days with precipitation early in the morning. Afternoon until night rainfall represents heavy convective type of rainfall (torrential rain). Torrential rainfall type is a rainfall that pouring heavily in a very short period. This type of rainfall is clearly seen in the

data as most of rainfalls are instant (15 minute) rainfall up to 2 hour duration. There is almost no rainfall in the following days of heavy rainfall. The relative humidity and temperature data has a strong correlation with $R=0.95$ with inverse relationship. While the pressure data has two maxima and minima daily, which represent strong mountain and sea breezes. Thus we expect strong influences of in situ weather parameters of other observation sites.

From Figure 2, we derived empirical relationship between rainfall and other parameters and found out that most rainfall occurs at 20.00 WITA (local time) and the variability of temperature, pressure and relative humidity in the morning seems to have significant relationship with the rainfall event at the night after as shown in Figure 3. From the rainfall data, it is clear that almost whole rainfall falls between 16.00 and 22.00 WITA with the maximum intensities normally occur at 20.00 WITA. For daily rainfall prediction, we tried to use the predictor as far ahead from that time but at similar day. The reason is that changes in environmental in-situ parameters (temperature, relative humidity and pressure) foreshadow the upcoming rainfall event. Then we chose some predictor in the morning based on some empirical studies conducted (not shown). Based on

several empirical studies like that, we design our method of research using the following data:

- rainfall (mm) at 20.00 WITA
- temperature ($^{\circ}\text{C}$) at 07.00, 10.00 and its tendency between 07.00 and 10.00 WITA
- surface pressure (mBar) at 07.00, 10.00 and its tendency between 07.00 and 10.00 WITA
- relative humidity (%) 07.00, 10.00 and its tendency between 07.00 and 10.00 WITA

ANFIS procedures. Considering to the inter relationship among several weather parameters as mention above as well as the input preparation for ANFIS as used before (Refs), the following input scheme for the training data has been setup for the temperature, pressure, relative humidity and precipitation rate. The data checking follows the below schemes for each parameters as the major predictor except rainfall itself. Note that historical rainfall rate is determined by the latest history of rainfall plus other predictors in two previous series.

Training data:

$$\text{temp}(t) = [x(t-1) \ x(t)], \text{bar}(t) = [x(t-1) \ x(t)], \text{rh}(t) = [x(t-1) \ x(t)], \text{ch}(t) = [x(t-1)]$$

Data Checking:

$$\text{ch}(t) = [x(t)]$$

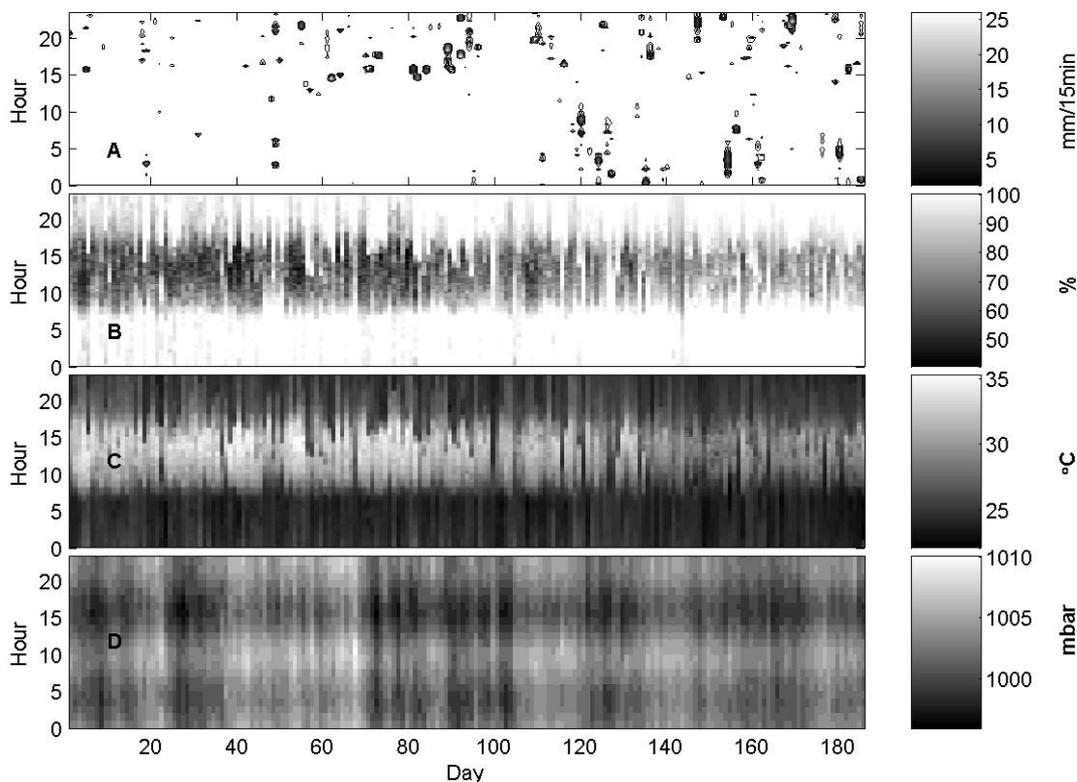


Figure 2. Main data used in this study from the automatic weather station no 4 belongs to PT Freeport, which is located in Timika airport in 15 minute temporal resolution. This figure illustrates the rainfall data (A), relative humidity (B), surface temperature (C) and surface pressure (D). Number of days represents days between 1 January and 5 July 2005 (184 days)

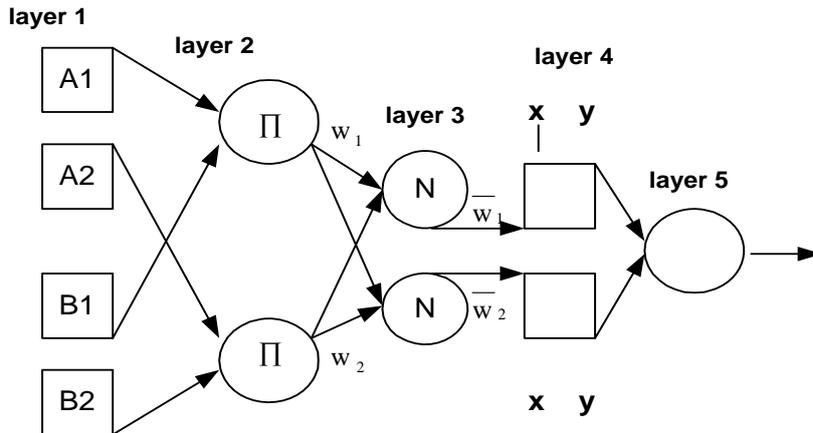


Figure 3. Schematic architecture of the Sugeno Fuzzy model of one order and two inputs [6]

Because of the limitation of ANFIS that could take up only maximum 6 input series, we divided our investigation into different categories of temperature, pressure and relative humidity at different local time series (at 07.00 or 10.00 WITA) and their tendencies between different local time (differences between 07.00 and 10.00 WITA). Then we prepared input series up to maximum 5 input for training data and additional latest rainfall data for ANFIS prediction. We then investigated among all possible combinations of different series for training and prediction and record the performance by their correlations and root mean square errors (RMSE).

ANFIS model setup. According to the ANFIS tools and guidelines, there are two main things should be done in ANFIS model setup as follow:

I. Sugeno ANFIS network setup

ANFIS network as shown in Figure 2 will be setup following the input scheme as described above. This process is conducted with a command `genfis1` with `numMF` is 2 and `mftype` is 'gbellmf'.

II. ANFIS Training processes

The training process proceeded with a command `anfis` with back propagation method and `epoch` number 10. With this training we have ANFIS ready for the prediction. During this training, ANFIS will learn pattern among different time series and within each series itself.

Hindcast rainfall prediction

To perform prediction against daily rainfall data $ch(t) = [x(t+1)]$ the input time series (consist of two or three parameters including rainfall) during the training processes were advanced one time step ahead, i.e. the present day. Then, with the command `evalvis` we enter the ANFIS formula for prediction based on the learning processes of the input data. The process was repeated for with the next observed (not predicted) daily

rainfall data for the next day prediction. As the process move further in time, we will obtain a new time series of daily prediction. The whole process is known as the hindcast method or the prediction using a real observed data in the past. Since the whole process was conducted using real observed data, thus we have limitation on the length of the training data. In this study, we will investigate the influence of the training data length for the performance of ANFIS prediction. We used data length of 40 up to 100 for 184 day observed rainfall. Hence we performed hindcast prediction from data of 41 up to the end (184). Next we continued the same process for training data length 40 up to 100 with step interval 2 and repeated the daily rainfall prediction for other specific length and predictors.

Validation

The validation process was done with two parameters, i.e. the correlation and the root mean square error as defined below:

Correlation:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}}$$

RMS Error:

$$y_j = \sqrt{\frac{\sum_{i=1}^M u_{ij}^2}{M}}, \text{ with } u_{ij} = (A_{ij} - B_{ij})^2$$

For generating those two parameters, real observed (A_{ij}) and predicted (B_{ij}) time series were involved using training data length from 40 to 100 at 2 step interval. In overall we will have performance measuring parameters of 31 for each predictor scheme used.

3. Results

The real time series result of ANFIS prediction is given in Figure 4, with a training data length of 70, predictor of humidity at 07.00 WITA. From that figure, there are two times overshooting occur during the process (22, 23 April and 26 May 2005). On most cases, ANFIS follows the pattern with a time lag of 1 days behind. However, there are still some rainfall values were taken by the system. After a heavy rainfall prediction, the ANFIS prediction will still give residue of rainfall amount that diminish slowly as time passes. Thus, only at the beginning of early prediction, the system catches up the null precipitation values. This phenomenon indicates that ANFIS is not very powerful for a stochastic event such as daily rainfall since the system still carries memories of previous events.

On the next figure, we illustrate several good results of ANFIS in predicting daily rainfall at 20.00 local time using several predictors. These are some of the best results so far for the multi variate ANFIS application. In this study we also investigate the influence of training data length against ANFIS performances. We found out that ANFIS has stable prediction despite of training data length if the system using relative humidity data at 10.00 in the morning (bottom right), follows by the results using relative humidity data at 07.00 (bottom left). Although there is a very significant correlation between relative humidity and temperature variability, when we are using the temperature data directly, the system performance drops significantly. One possible reason is that magnitudes and ranges between those two parameters are different by four times for relative humidity, i.e. 40 – 100 for relative humidity and 23 – 35 for temperature. A further test should be conducted to prove this hypothesis, whether the system is sensitive to the magnitude and range differences. Such a test, however, will be done later on (Figure 7). Another solution is by introducing a transforming function, since in every test performed in this study, we include rainfall data for training as well. The rainfall data has magnitude and range far beyond the relative humidity or even the temperature. If the above hypothesis is correct, then the

rainfall data has more biases in the system to the result. In our early assumption, we try to use other predictor as main trainer for upcoming precipitation pattern (foreshadow predictor). However, it seems that the role of rainfall is still far greater than other predictors, by considering only the magnitude and range differences. For improvement of multivariate ANFIS, magnitude – range should be the major consideration in setting up the possible transforming function.

Also from Figure 5, the performance of ANFIS with other predictors do not show consistent outcome with different training data lengths. We notice maximum correlation values at around 0.6 for all predictors. In general for large training data sizes, the performance of ANFIS will be lower except for the predictor from a tendency between pressure at 07.00 and 10.00. Again, among the rest of predictors, a combination between relative humidity and pressure give a moderate result. Thus, relative humidity is the best predictor so far for predicting daily rainfall at 20.00 WITA.

In comparison to performance test bench using correlation values, the performance analysis using the RMSE does not show persistent relationship between those two measures. In general high correlation values relate to low RMSE, but this is not the case for the whole story. In the same figure, we notice that RMSEs when using relative humidity as the predictor are relatively much lower than those when using other predictors. Hence, this is again evidence that relative humidity is the best predictor for daily rainfall. We also learn that the RMSE performances indicate significant influence of magnitude and ranges. As the surface pressure has large magnitude, so does the RMSE in comparison to temperature and relative humidity. The indication is clearer if we compare the RMSE result when using only pressure and when we use a combination with temperature and relative humidity.

ANFIS performances on several other weak predictors are given in Figure 6. In addition to previous figure, there are some experiments to predict precipitation at 17.00 WITA using similar strong predictors as it was for

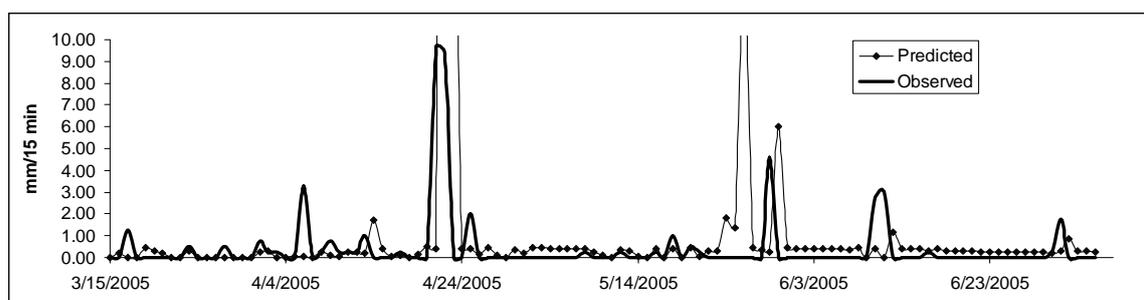


Figure 4. An example of the predicted and observed series of accumulated rainfall at 20.00 WITA after application of multi variate ANFIS prediction with the predictor relative humidity at 7.00 WITA

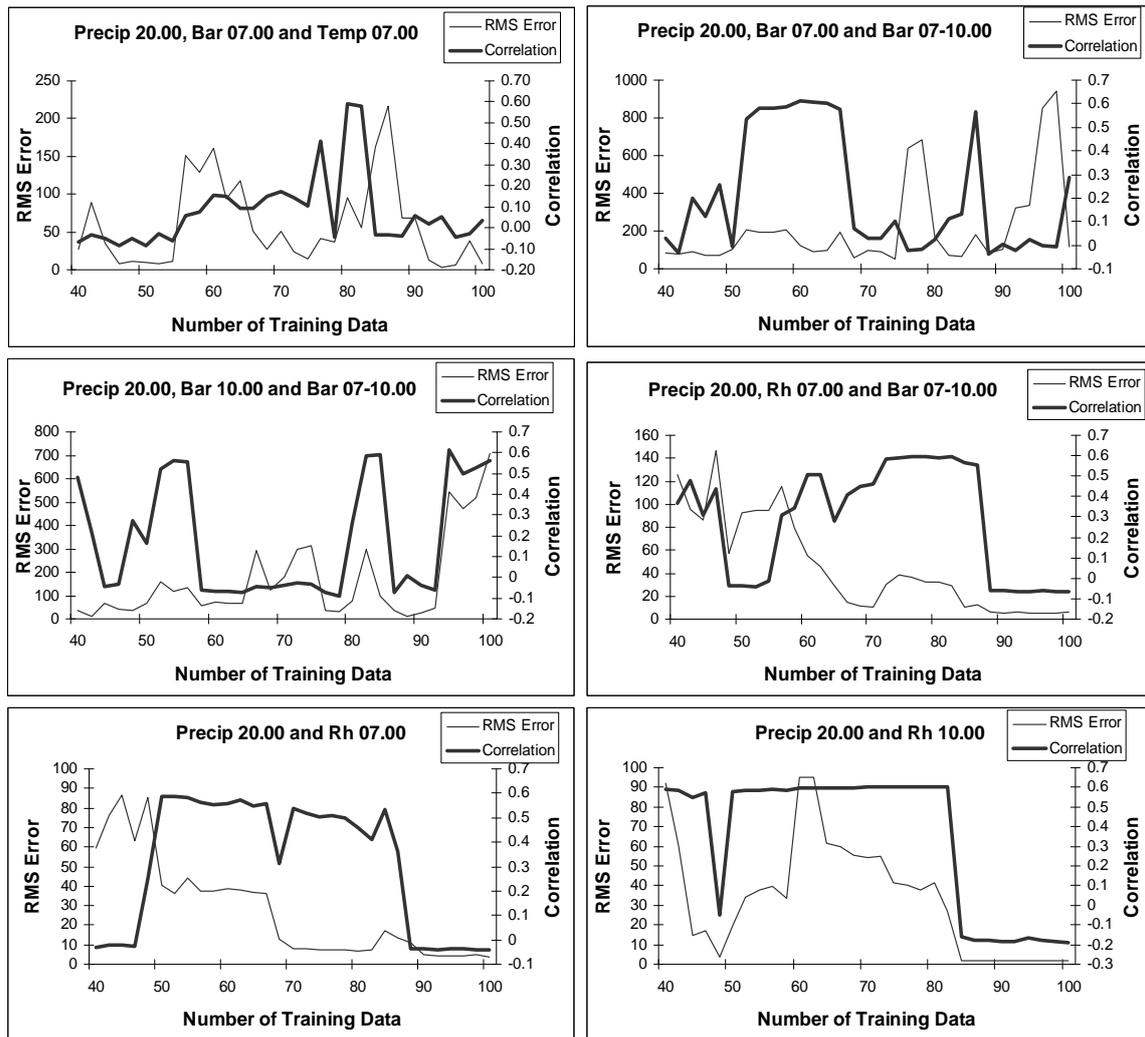


Figure 5. Performances of multi variate ANFIS in predicting daily rainfall values at 20.00 WITA. Performances are measured by their respective correlation values (right axes, solid thick lines) and the RMS error (left axes, solid thin lines) with different predictors. These performances are among the strong predictors in this study

precipitation at 20.00. One of the best performances is shown by a predictor temperature at 10.00. However, when we combine temperature and relative humidity, performances drop. For these unsatisfying performances, we do not relate the correlation results with RMSE.

Results of several experiments performed in this study raise the possibility to combine between predictors with an appropriate transforming function and scalability. In original Sugeno map (Figure 3), transforming function is not described well. We should define clearly how we introduce this transforming function into ANFIS or shall the scaled magnitude be used as the transforming function. Other difficulty comes from the nature of rainfall itself, which is intermittent and stochastic processes, while other parameters have variabilities bounded by some ranges. We shall note that the rainfall

predicted here is limited to rainfall at 20.00 – 20.15 only.

We found previously that there is a significant relationship between temperature and relative humidity data (up to 95% correlation) over all time of observation. We also notice that the best predictor yet is the relative humidity at 10.00 WITA. The correlation between the relative humidity and temperature at that time is slightly lower at R value of -0.922 (notice the negative correlation means inverse relationship). Now we try to prove our hypothesis that scale and magnitude play major role in performance of ANFIS. We transform the temperature values at the level of scale and magnitude similar to that of the relative humidity, i.e. the temperature dataset at 10.00 WITA shall have its maximum and minimum values equal to those of the

relative humidity. The transformation was conducted using the following relationship

$$T'_{10.00} = \frac{(T_{10.00} - T_{10.00}^{\min})(RH_{10.00}^{\max} - RH_{10.00}^{\min})}{(T_{10.00}^{\max} - T_{10.00}^{\min})} + RH_{10.00}^{\min}$$

The result of ANFIS rainfall prediction after data transformation is given in Figure 7. That figure shall be an optimal improvement of Figure 5 bottom left of the best ANFIS performance so far. From that figure, it is clear that the ANFIS performance has significantly improved after data transformation close to that of the

RH 10.00 predictor. Thus the weak ANFIS performance using temperature as predictor is due to inappropriate scale and magnitude and the performance could be gradually increased just by transforming the data. This result brings a consequence of transform different predictors at an appropriate level so that ANFIS could reach its optimal predictive skill. Whether the scale and magnitude of the relative humidity data is already at the optimal level, is still an unresolved problem. However, considering the scale and magnitude of the rainfall data, there is still a space for improvement.

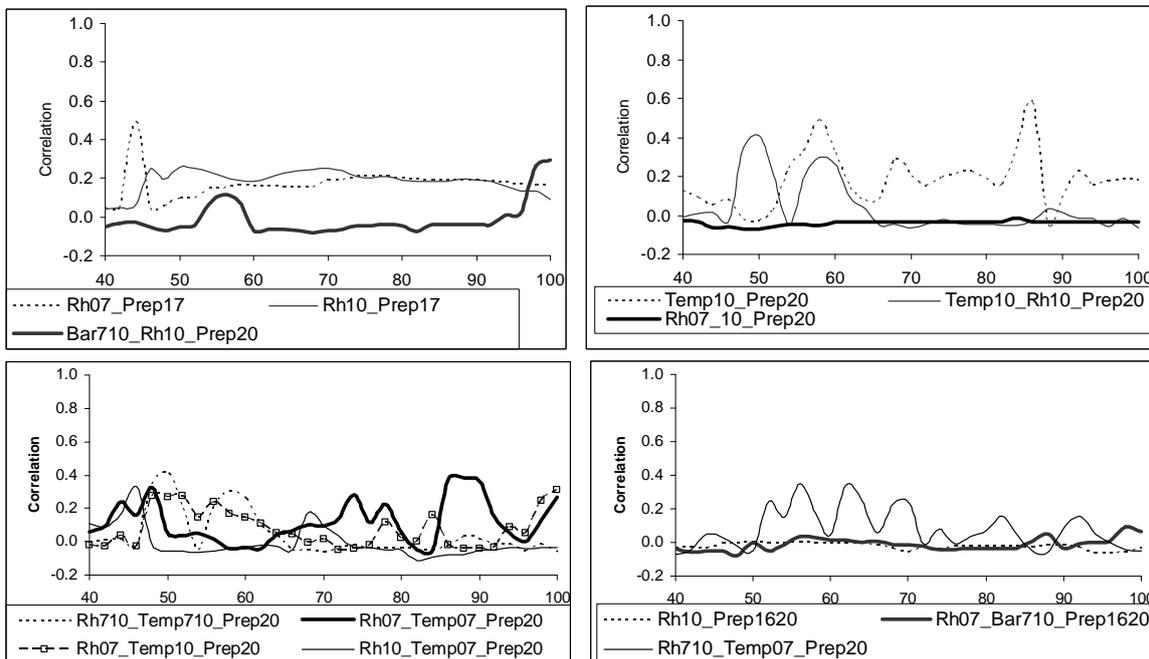


Figure 6. Performances of some weak predictors with their respective weak correlation values around zero

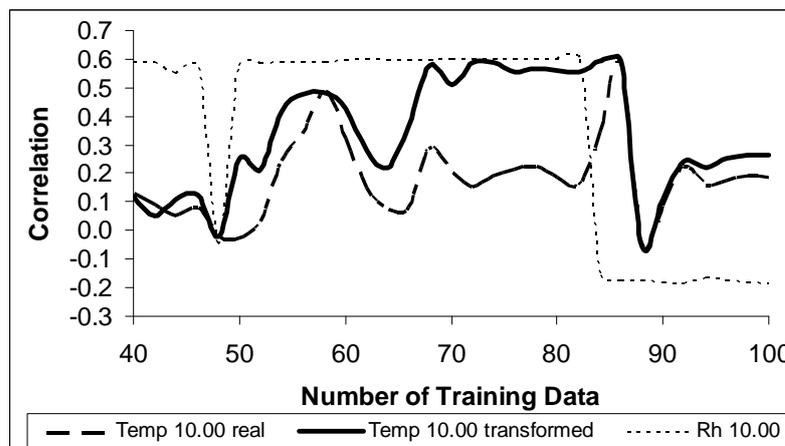


Figure 7. A comparison of performances of multi variate ANFIS in predicting daily rainfall values at 20.00 WITA using RH at 10.00 (thin dashed line) as in Figure 5 bottom right, using temperature at 10.00 (thick dashed line) and using transformed temperature at 10.00 (solid thick line). Transformed temperature data is the temperature data at 10.00 after elevated and scaled at magnitude and range similar to RH at 10.00

4. Conclusions

We have performed a study of multivariate ANFIS application in predicting daily rainfall values. The study focuses on the influence of training data length. The prediction was done on rare data without any filtering and smoothing. We used several parameters recorded by the automatic weather station in Timika airport. We found out that relative humidity at 10.00 and 07.00 in the morning is the best predictor for rainfall occurrence at 20.00 WITA.

The study suggest possible extension of the work by using a transforming function on different predictor that we want to emphasize and scaling the magnitude of some parameters to be equal in range. We found out that ANFIS is sensitive to different magnitudes and scale sizes and not an appropriate tool for a stochastic process like rainfall. There is a need to conduct spatial data analysis in rainfall prediction over the study area.

Acknowledgements

The authors would like to thanks Steffy Rompas for valuable data and information of the Freeport

concession area and its typical problems. The authors are also indebted many thanks to Edi Santoso for fruitful discussion and critics in preparing the study.

References

- [1] BMG, Analisis Iklim di Area Kontrak Karya PT. Freeport Indonesia, Pusat Analisa dan Pengolahan Data, Badan Meteorologi dan Geofisika, Jakarta, 2001.
- [2] Indragustari, Lokakarya Nasional Forum Prakiraan, Evaluasi dan Validasi, Jakarta, 2005.
- [3] T.H. Liong, Pelatihan Aplikasi Matlab Untuk Meteorologi, Bandung, 2005.
- [4] Z.L. Dupe, Noersomadi, Lokakarya Nasional Forum Prakiraan, Evaluasi dan Validasi, Jakarta, 2005.
- [5] F.H. Widodo, J. Arifian, M. Kudsy, S. Tikno, S. Nuryanto, J. Sains dan Teknologi Modifikasi Cuaca (2004) 5, 10 – 21
- [6] J.S.R. Jang, ANFIS: Adaptive-Network-Based Fuzzy Inference System, IEEE Transactions on Systems, Man and Cybernetics (1993) 23, 665-685.