Monthly Rainfall Prediction Using Statistical Downscaling with Combination of Grid Boxes and Adaptive Neuro Fuzzy and Inference System in Lombok

A Safril1* and A Ulfiana2

1Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang Selatan, Banten, Indonesia
2Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah IV, Makassar, Indonesia 90231

*E-mail: agus.safril@stmkg.ac.id

Abstract. Lombok is an island in Indonesia that has more people live in this area. Flood is a natural disaster that cause severe impact on society. To manage the risk of natural disaster, seasonal prediction of rainfall for early warning system are needed. However, prediction is difficult to predict due to El Niño and La Niña. Seasonal prediction of rainfall using Global Circulation Model (GCM) is useful to capture rainfall variability, but has a lack information because it has coarse of grid resolution (more than 200 kilometers). This model cannot provide in the local area so statistical downscaling model is used to get detail information. To optimize the result of prediction from global model, a method to capture the extreme condition is needed. Taking a best predictor that will be used to predict the rainfall is done to get the high performance of model used Singular Value Decomposition (SVD). ANFIS (Adaptive Neuro Fuzzy and Inference System) is a model prediction to capture the rainfall variability. Predictor is chosen based on the physical condition of atmosphere and ocean that have impact on Lombok Region. Selected area of predictor (atmospheric variable) are grid boxes in the Maritime Continent (80°E - 150°E and 12.5°N – 12.5°S) and Nusa Tenggara (105°E - 120°E, 10°S - 0). Eight variable predictors are used for prediction, they are air temperature at 2 m (T2M), geopotential height at 500 mb (Z500), zonal and meridional wind at 850 mb (U850 and V850) and at 200 mb (U200 and V200), sea level pressure (SLP), and air temperature at 850 millibar (T850). Both predictor variables (grid box combination) are used to predict for lead time one month. The result of research shows that rainfall prediction can capture the rainfall variability (correlation 0.70) with RMSE (Root Mean Square Error) is 69.

1. Introduction
Lombok is a small island with a mountain range in it. Moreover, Lombok also has rivers whose upstream is in mountainous area and downstream is in lowlands area. These conditions cause various natural disasters such as droughts, floods, or landslides. In West Lombok, floods occurred several times during the rainy season and experienced drought during the dry season [1].

Besides having the potential hazards of flooding and drought, Lombok is also known as earthquake prone area. This has an effect on potency of landslide. Earthquake and high rainfall area, high potential of the landslide, areas with ground cracks due to the earthquake with steep slope can result in floods and landslides during the rainy season. One of the causes of soil movement or landslides is high rainfall as a trigger [2]. Monthly rainfall predictions are needed to obtain information of rainfall condition in the future so that they can be used for mitigation (actions to reduce the impact of disaster
both life and property). In addition of disaster mitigation needed, information of rainfall prediction also useful for decision making in agriculture caused by climate variability due to ENSO (El Niño Southern Oscillation) [3]. It is because Lombok is also known as the center of paddy, some secondary crops, and vegetables farming [4]. Rain prediction information is used to determine the cropping pattern when planting will begin and the suitable crops.

Monthly rainfall predictions using Global Circulation Model (GCM) can capture rainfall patterns in the next few months, but information obtained in low resolution (more than 200 kilometers). Therefore, a statistical downscaling method is needed to get rainfall prediction in local scale [5]. Predictors that used in this study are various atmospheric variables (air temperature, air pressure, air humidity, etc.), while predictand (which is predicted) is rainfall.

In previous studies related to statistical downscaling using grid box (single) and global predictors for predictions in Indonesia region with grid boxes in the study area (Sumatra) [6]. Predictors characteristics of local grid box area around the study area are needed to obtain local patterns. In that context, research is carried out by combining grid boxes with global and local regions in the grid of study area.

2. Data and Method

2.1. Data
Predictor variables data that will be used for prediction are hindcast data (predictions of the past period) output of the Global Circulation Model (GCM) from National Centers for Environmental Prediction (NCEP) in monthly period with the output of 15 ensemble members. This result is averaged so that the average ensemble is obtained and used for prediction one month ahead (one-month lead time). Data reanalysis can be downloaded at NOAA (https://www.esrl.noaa.gov/) to perform a physical analysis of the selection of predictor boxes. Data resolution for analytical purposes is 2.5° (250 kilometers).

Data of predictor variables used for this study are eight variables consisting of zonal and meridional wind at 850 millibars (U850 and V850) and at 200 millibars (U200 and V200), surface air temperature (T2M), air temperature at 850 millibars (T850), air pressure at sea level (SLP), and geopotential height at 500 millibars (Z500) as in Table 1 offered at 2.5° (250 kilometers) resolution GCM. The 25-year data period is hindcast data (predictions that were run in the past between 1981-2005. Predictor data is used for prediction time period for the next month

Table 1. GCM atmospheric independent variables (wind, air pressure, air temperature, geopotential height) and symbols used (U850, V850, U200, V200, T200, T850, and Z500)

| No | Independent Variables               | Symbols |
|----|-------------------------------------|---------|
| 1  | Zonal wind at 850 millibars         | U850    |
| 2  | Meridional wind at 850 millibars    | V850    |
| 3  | Zonal wind at 200 millibars         | U200    |
| 4  | Meridional wind at 200 millibars    | V200    |
| 5  | Surface air temperature (at 2 meters)| T2M     |
| 6  | Air temperature at 850 millibars    | T850    |
| 7  | Air pressure at sea surface         | SLP     |
| 8  | Geopotential height at 500 millibars| Z500    |

The rainfall data as predictor variable is monthly rainfall data sourced from Indonesia Agency for Meteorology, Climatology and Geophysics (BMKG) from 1981-2005 namely Dasan Tereng (latitude: -8.584052, longitude: 116.18441 E, Elevation: 156 m), Narmada District, West Lombok Regency, West Nusa Tenggara (Figure 1).
The selected predictor areas (grid boxes) are used for predictors (Figure 2), which include:
1. Regional area (Maritime Continent) in the 80°E-150°E, 12.5°N-12.5°S
2. Java Sea (around Lombok) in the 105°E-120°E, 0-10°S.

To obtain the best predictor variable, correlation analysis was performed on all 8 predictor variables (U850, U200, U850, V850, T850, T2M, SLP, and Z500) from the two grid boxes. The result of a predictor window combination of two predictor variables from these two predictor boxes is used as the input for prediction using the statistical downscaling method.

2.2. Method
2.2.1. Grid boxes selection based on atmospheric physical characteristics. Selection of the grid boxes for the predictor area is based on physical characteristics that affect rainfall in the Maritime Continent (Figure 3). The Maritime Continent region has high solar radiation so that it becomes a source of water vapor for Indonesia region [7]. In addition, the islands in Indonesia are influenced by several global
factors such as monsoon and El Nino Southern Oscillation (ENSO) so that the variability of rainfall becomes high [8]. The monsoon pattern is also dominant in the territory of Indonesia. For the Nusa Tenggara region, rainfall pattern is Monsoon A with a peak rainfall [9]. In December-January-February, the wind pattern is westerly (the west coast region of Sumatra, the Western Pacific, and the South China Sea) which brings rainfall in the Nusa Tenggara Islands region as in Figure 3 (big box) (80°E-150°E, 12.5°N-12.5°S), while the small box for predictors is around the Nusa Tenggara region (local characters) (105°E-120°E, 0°-10°S).

![Figure 3. Wind pattern average in December-January-February. Big box (grid box of Maritime Continent area) and small box (Nusa Tenggara region)](image)

2.2.2. SVD correlation by reducing multicollinearity and dominant mode. Singular Value Decomposition (SVD) or Principal Component Analysis (PCA) is a method to see dominant factors that influence the characteristics of rainfall in the study area and eliminate multicollinearity. The SVD method provides a physical interpretation of the results of the relationship between predictor variables and rainfall statistically [10]. The dominant pattern used is mainly in capturing monsoon variability. The Nusa Tenggara region is an area with an annual rainfall pattern that is strongly influenced by the monsoon. This method identifies pairs of fields (matrices) to patterns of space and time. Each pair of fields describes a covariance fraction. In the dominant mode, the strength of the relationship between predictor variables with rainfall is seen with the correlation method. Relationship between variables is strong when approaching 1 and weak when approaching 0 (zero).

The independent variable $X$ is expressed in a matrix with two predictor variables from the GCM model in the grid box 1 (Maritime Continent) and grid box 2 (Nusa Tenggara). The correlation between each predictor variable and Lombok observation rainfall (matrix $Y$) is computed to find the highest correlation. The results of the variables that have the highest correlation are used as predictor variables for each grid box. The best variables in the first and second grid boxes are then used as independent variables.

Next the covariance matrix ($X^T Y$ can be calculated if it has the same number of rows as $Y$ as the time series).

$$C = X^T Y$$ (1)
After the covariance matrix $C$ is formed, then the Singular Value Decomposition (SVD) is calculated and matrices $U$, $V$ and diagonal matrix $L$ is obtained, so that:

$$C = ULV^T$$ (2)

where the singular vector $X$ is the columns in $U$ and the singular vector of $Y$ is the columns of $V$. Then from this matrices mode (coefficient) expansion is obtained in the time series between time series (predictor variables) and precipitation series (predictand) and correlated. Correlation values ranged from 0 to 1. Correlation approaching 1 shows the strong relationship between predictand and predictor variables, when approaching 0 indicates a weak relationship between the two variables.

2.2.3. Regression model to capture extreme rainfall pattern.

Indonesia region is influenced by various climate phenomena so a model is needed to capture the variability. The Maritime Continent region is influenced by several global factors so that the variability of rainfall becomes high. This is caused by the Indonesian Maritime Continent which is predominantly influenced by seasonal variability due to the Asian monsoon [11]. In addition, it is also influenced by the phenomenon of inter-annual variability such as El Niño-La Niña and IOD (Indian Ocean Dipole) alone or simultaneously [12]. This condition causes high rainfall variability so that extreme rainfall (high or low) and irregular often occur.

Based on these conditions, a prediction model that can capture these extreme patterns is needed. ANFIS (Adaptive Neuro Fuzzy Inference System) is a regression method that can capture nonlinear patterns and chaos. This regression prediction model based on artificial intelligence used to capture the behavior of past rain anomalies that are used for the prediction model in the future. For the purposes of predictive capability analysis, 22 years of training data and 3-year testing data are used with the $gbellmf$ membership function and Sugeno type. Prediction with ANFIS uses the Sugeno type, with membership numbers ($numMFs$) are two, the membership function input pattern is 'gbellmf' and the membership function output is 'linear'. The ANFIS structure is presented in Figure 4. The figure shows two data inputs ($x$, $y$).

In this ANFIS structure, EBP (Error Back Propagation) method is carried out in the first layer and LSE (Least Square Estimator) [13]. In the ANFIS structure the process for updating the fuzzy membership parameters is done at the first layer by EBP and for updating the linear parameters of the system output at layer four by LSE. ANFIS has the ability to recognize past rainfall anomalies and reduce prediction errors.

![ANFIS multilayer structure](image)

2.2.4. Prediction results verification.

After prediction, to see the accuracy of the model, a model verification analysis is performed. The first analysis uses qualitative descriptive analysis using graphs by looking at the comparison between the rain signal patterns of prediction results and observation. Quantitatively, the reliability of the prediction model is calculated by the regression variance or RMSE (root mean square error) which is obtained from the square of the difference between the predictions $y_k$ and observations $o_k$ rainfall [15]. RMSE values are getting smaller (approaching zero) indicates better model. In addition, a correlation
is also made between the results of prediction and observation. High correlation value shows the model is getting better.

\[ MSE = \frac{1}{n-\mu} \sum_{i=1}^{n} (y_i - \mu) \]  

(3)

\[ RMSE = \sqrt{MSE} \]  

(4)

3. Results and Discussion

3.1. Correlation results between predictors and predictand

Correlation results between predictor variables and rainfall using Singular Value Decomposition (SVD) show strong correlation. The highest correlation is indicated by surface air temperature (2 meters) variable with a correlation on the Maritime Continental grid box (0.72) and local (0.71) as in Table 2. Predictions are made with multiple independent variables with predictors from the Maritime Continent and local grid boxes. This best variable becomes the input regression model using ANFIS.

Table 2. Correlation between predictor variables and rainfall

| No | Variables | Maritime Continent | Local |
|----|-----------|-------------------|-------|
| 1  | U850      | 0.70              | 0.42  |
| 2  | V850      | 0.68              | 0.57  |
| 3  | U200      | 0.52              | 0.54  |
| 4  | V200      | 0.67              | 0.60  |
| 5  | T2M       | **0.72**          | **0.71** |
| 6  | T850      | 0.70              | 0.65  |
| 7  | SLP       | 0.69              | 0.55  |
| 8  | Z500      | 0.30              | 0.29  |
| Best | T2M       | **0.72**          | **0.71** |

3.2. Rainfall pattern in Lombok Island and model verification

Scatter analysis of rainfall from 25 years observations between monthly rainfall. Monthly rainfall in Lombok shows a monsoonal pattern (one peak of rainfall). Rainfall with value is greater than 150 millimeters per month occur in November to April with the peak rainfall in December. Dry conditions begin to occur in April and the driest in July continued to August (Figure 4). This dry soil condition and cracking due to the earthquake and when filled with water will cause a subsequent landslide.

Figure 5. Monsoonal rainfall pattern in Lombok
The prediction result shows the rainfall signal between the prediction (thick line) and observation (dashed line) can capture the observation rainfall pattern (monsoon) as in Figure 6. In a few years the high variability between November and July has not fully captured the pattern local rainfall, but can quantitatively capture extreme rainfall that exceeds 250 millimeters per month.

![Figure 6. Comparison between prediction result and observation rainfall pattern](image)

Furthermore, based on quantitative analysis calculated the correlation value and RMSE value. The result of the correlation between prediction and observation by simulating the prediction of 2003-2005 shows that rainfall prediction can capture high rainfall patterns (Figure 6). Correlation values between prediction and observation is strongly related (0.70) and RMSE (Root Mean Square Error) is 69 millimeters.

4. Conclusions
Correlation analysis between predictors and predictand shows the best variable for being grid box predictor to be used for rainfall prediction in Lombok is surface air temperature (T2M) with relatively strong correlation (0.72). From the best predictor variable, it is obtained a regression model with predictor combination based on grid boxes in the Maritime Continent and local around West Nusa Tenggara which is able to capture extreme rainfall patterns of monsoon and inter-annual variability. Predictive model capability shows a strong relationship between prediction and observation results, shows high correlation values (0.70) and RMSE 69 millimetres of monthly rainfall which generally ranges from 100 millimetres per month.

References
[1] Marchiavelly M I, Narieswari L, Munajati S L, Sumaryono, Santoso W E and Martha S 2012 Hazard risk mapping on tourism region of West Lombok Regency, West Nusa Tenggara Majalah Ilmiah Globe 14 187-99.
[2] Sipayung S B, Cholianawati N, Susanti I, Aulia A and Maryadi E 2014 Development of empirical equation model in predicting the occurrence of landslide at watershed of Citarum (West Java) based on the TRMM satellite data Jurnal Sains Dirgantara 12 12-21.
[3] Utami A W, Jamhari and Hardyastuti S 2017 El Nino, La Nina, dan penawaran pangan di Jawa, Indonesia Jurnal Ekonomi Pembangunan 12 257-71.
[4] BPS-Statistics of Lombok Barat Regency 2016 Lombok Barat Regency in Figures (Gerung: BPS-Statistics of Lombok Barat Regency) p 239.
[5] Juneng L, Tangang F T, Kang H, Lee W J and Seng Y K 2010 Statistical downscaling forecasts for winter monsoon precipitation in Malaysia, using multimodel output variables J. Clim. 23 17-27.

[6] Safril A, Hadi T W, Hadi S and Tjasjono B H K 2013 Prediksi curah hujan bulanan menggunakan Adaptive Statistical Downscaling Jurnal Meteorologi dan Geofisika 14 25-31.

[7] Ramage 1968 Role of tropical maritim continent in the atmospheric circulation Mon. Weather Rev. 96 6.

[8] Chang C P, Wang Z, Ju J and Li T 2004 On the relationship between Western Maritim Continent monsoon rainfall and ENSO during northern winter J. Clim. 17 665-72.

[9] Aldrian E and Susanto R D 2003 Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature Int. J. Climatol. 23 1435-52.

[10] Björnsson H and Venegas S A 1997 A Manual for EOF and SVD Analyses of Climatic Data (Montreal: Department of Atmospheric and Oceanic Sciences and Centre for Climate and Global Change Research, Mc Gill University) p 23.

[11] Harijono S W B 2008 Interaksi Fenomena El Nino dan Dipole Mode Secara Simultan serta Monsun Musim Panas India Terhadap Variabilitas Curah Hujan di Sumatera Bagian Utara (Bandung: Institute Technology Bandung)

[12] Tjasyono B H K, Lubis A, Junaeni I., Ruminta and Harijono S W B 2008 Dampak variasi temperatur Samudera Pasifik dan Hindia Ekuatorial terhadap curah hujan di Indonesia Jurnal Sains Dirgantara 5 83-95.

[13] Fariza A, Helen A and Rasyid A 2007 Proc. Seminar Nasional Aplikasi Teknologi Informasi (SNATI) (Yogyakarta: Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia) pp 77-82.

[14] Jang J S R 1993 ANFIS: Adaptive-Network-Bace Fuzzy Inference System IEEE Trans Syst Man Cybern 23 665-685.

[15] Wilks D S 1995 Statistical Methods in the Atmospheric Sciences (New York: Cornell University) p 279.