

Statistical Downscaling Model Based-on Support Vector Regression to Predict Monthly Rainfall: A Case Study in Indramayu District

Agus Buono

Department of Computer Science, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University
Center for Climate Risks and Opportunity Management in Southeast Asia and Pacific
pudasha@yahoo.com

Muhammad Asyhar Agmalaro

Department of Computer Science, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University

Mushthofa

Department of Computer Science, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
mush@ipb.ac.id

Muhammad Faqih

Department of Agro-meteorology, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University

ABSTRACT

Knowledge of weather and climate, especially rainfall is significantly needed in agricultural sector. Accurate information of rainfall can be used to determine the planting pattern and time appropriately, so that farmers can avoid crop failure caused by floods due to high rainfall and drought due to low rainfall. Techniques of statistical downscaling (SD) using a global circulation model output (GCM) are commonly used as a primary tool to learn and understand the climate system. The aim of this research was to develop an SD model using support vector regression (SVR) with GCM as input to predict monthly rainfall in the district of Indramayu. The research results showed that GCM can be used to predict the average value of monthly rainfall. The best result of prediction is at the Bondan Station having an average correlation of 0.766.

Keywords: Statistical downscaling, global circulation model, support vector regression, monthly rainfall

1 INTRODUCTION

Climate is an important natural phenomenon that influences human live. The knowledge about weather and climate pattern, especially rainfall is required by the agricultural sector. The models or tools that can simulate climate are required to investigate the climate can then be used to predict the amount of rainfall in one area. Global Circulation Model (GCM) is one of the recent models used to observe the impact and to predict the climate change. However, data resolution in GCM is relatively low, having large-sized grid. As a result, GCM output cannot be used to accurately model local impacts. One of the common methods used to overcome this problem is **statistical downscaling (SD)**, in which a statistical relationship is established from observations between large scale variables and a local variable at a particular site.

Previous studies have reported the implementation and management of impact and risk of climate by using SD. For instance, Wigena (2006) [1] has elaborated SD model to predict the rainfall in Indramayu and one of his analysis was to determine the best domain for GCM output by using projection pursuit regression (PPR). In addition, Cavazos and Hewitson (2002) [2] studied the

performance of GCM NCEP-NCAR output to find the potential combination of response variables by using artificial networks (ANN).

Support Vector Regression (SVR) is one the recent techniques developed in regression analysis ([3]). It has been known to have a very good performance in many applications of regression. This motivates the use of SVR in developing SD models in climate forecasting.

The main objective of this experiment was to develop SD models by using SVR in forecasting monthly rainfall (case study in Indramayu), in order to get the accurate climate information then can be used as a basic information to make a decision. The outcome of this study is to get the SD models that give the accurate climate information based on GCM data. This result can be used as a reference to forecast the rainfall more accurately. The data used for this study is the rainfall data from 13 rain stations in Indramayu taken during the years 1979-2002 and 6 models GCM data taken from years 1901-2002.

2 STATISTICAL DOWNSCALING

Statistical Downscaling (SD) is defined as an effort to relate between global-scale (explanatory variables) and local scale climate variables (response variables), [4]. Figure 1 illustrates the process of downscaling. We follow Sutikno [5] in applying GCM as the explanatory variable to perform statistical downscaling to predict the local climate response variables.

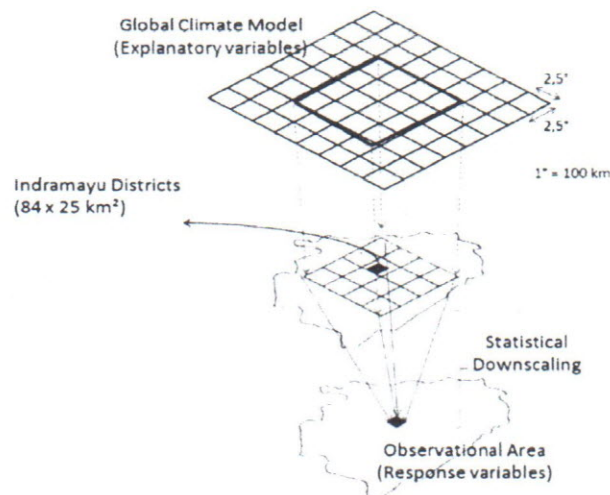


Figure 1: Statistical downscaling for climate predictions ([5])

There are two approaches for downscaling, using regional data (obtained from a regional climate model, RCM), or global data (obtained from the general circulation models, GCM). The first approach is known as statistical dynamical downscaling, while the second is known as statistical downscaling (SD). Statistical downscaling based on the relationship between coarse-scale grid (predictor) with local-scale data (response) is expressed with a statistical model that can be used to translate a global scale anomaly which became an anomaly of some variables of local climate. In this case the SD is a transfer function that describes the functional relationship of global atmospheric circulation with elements of the local climate, which is formulated in Equation (1).

$$Y_{t,p} = f(X_{t,q,s,g}) \quad (1)$$

where:

- Y : response climate variables
- X : global climate variables (provided by GCM)
- t : time period

- p : dimension of Y
- q : dimension of X
- s : layers in the atmosphere
- g : GCM domain

In general, a SD model involves time series data (t) and spatial data of GCM (g). Number of Y, X variables, the layer of the atmosphere in the model and the autocorrelation and co-linearity on the variables Y and X indicate the complexity of the model. In this research, we developed a support vector regression (SVR) model for statistical downscaling using precipitation data from the GCM as explanatory variables.

3 EXPERIMENTAL SETUP

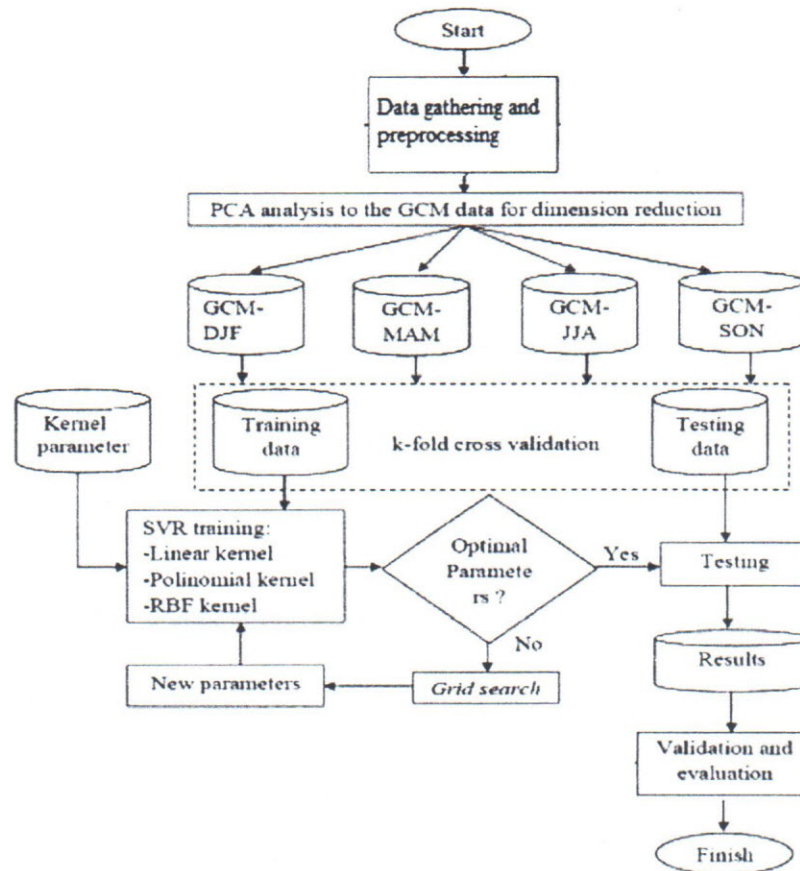


Figure 1: Experimental setup

Figure 1 shows the experimental setup used in this research. We first collect the observational data of the monthly rainfall in 13 stations in the district of Indramayu on the 22 years interval (1979-2000), and the GCM precipitations rate data covering those years from 6 different GCM models:

Pacific_20c3m_cgcm3.1_t47, Pacific_20c3m_cgcm3.1_t63, Pacific_20c3m_giss_model_er, Pacific_20c3m_gissaom, Pacific_20c3m_miub_echo_g and Pacific_20c3m_mri_cgcm2_3_2a. For each station, we determined a 5x5 GCM data grid on the surrounding coordinates of the station to be used as the predictor variables of the monthly rainfall on that station. For each of the GCM models, a different SD model will be created and evaluated.

To reduce the data dimension and eliminate redundancy, we perform the Principal Component Analysis (PCA) on the 25-variables GCM output collected from the 5x5 grid selected.

We then setup 5 different data division to create models for each of the periods: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). From these data and the observational data, we train 5 different SVR models for each month period using a 5-fold cross validation, and tested the result. To find the best combination of the SVR kernel and kernel parameters, we perform the standard grid search covering several different kernel and values of kernel parameters. Finally, we perform evaluation on the models generated and perform comparisons and analysis to draw conclusions.

4 RESULTS AND DISCUSSIONS

We evaluate the models by first looking at the comparison between the monthly averages of the models predictions versus the monthly averages of the observational data. In Figure 2, we see that the models trained have been able to follow the general pattern of the seasonal monthly rainfall.

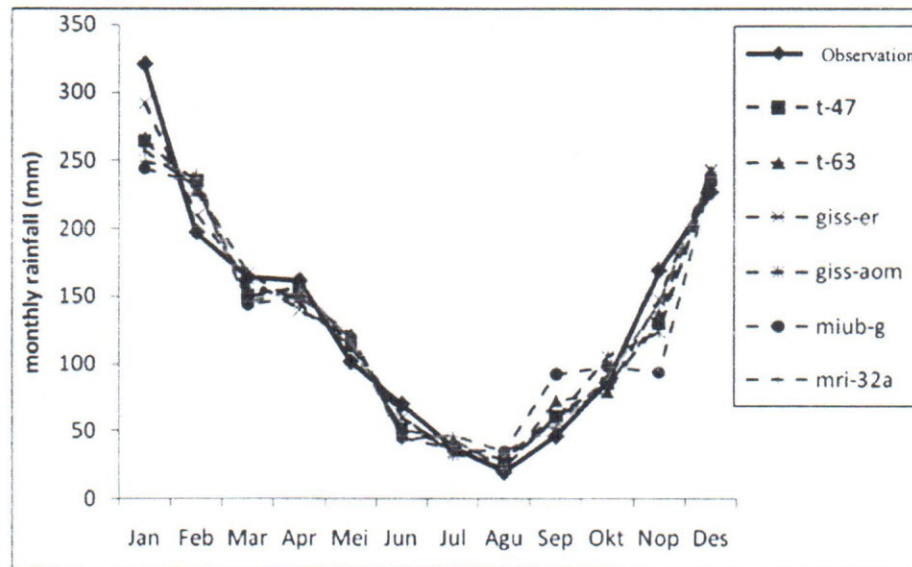


Figure 2: Comparison between the monthly averages of observational values and the monthly averages of the prediction values from the six models

We then compare the aggregate monthly prediction values (using the average, maximum and minimum aggregate operators) over the models trained using each of the GCM models to the monthly observational values. Figure 3 depicts such comparison. We can see that the observational values often fall between the maximum and minimum aggregate prediction values, while the averaged prediction values can, in general, follow the pattern of the observational values.

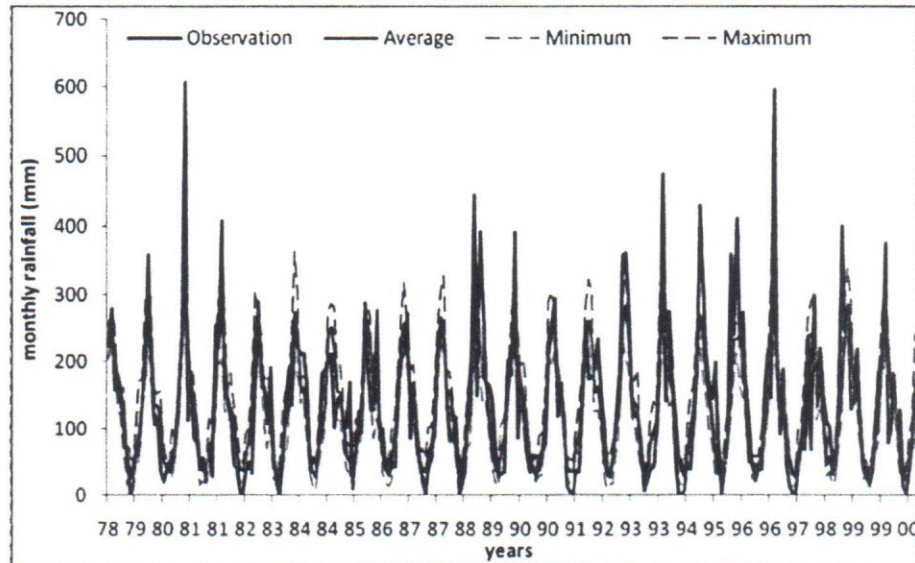


Figure 3: Comparison between monthly observational values and aggregated predictions using average, maximum and minimum aggregation

To quantitatively measure the performance of the predictions, we use the Normalized Root Mean Square Error (NRMSE) and the Mean Absolute Percentage Error (MAPE) as the measure of accuracy, and the linear correlation analysis to measure the statistical relationship between observational and predicted values. Figure 4 shows the average value of RMSE, MAEP, and correlation from the model validation based on the SVR kernel function used (for the entire GCM).

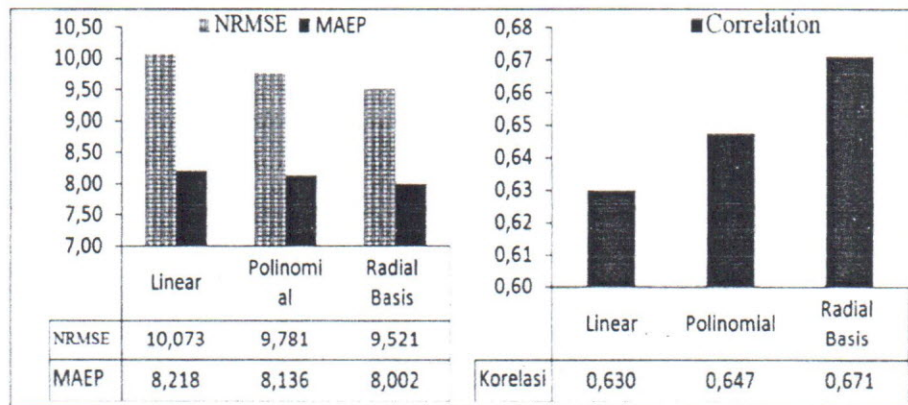


Figure 4: Average value of RMSE, MAEP, and correlation from the model validation based on the SVR kernel function used (for the entire GCM)

SVR models with linear kernel function have the largest error value and the smallest correlation value, whereas the radial basis function has the smallest error value and the greatest correlation value. In general, our results show that models which has lower error values tend to have higher correlation between observational and predicted values. The best averaged NMRSE and average MAEP are obtained using the Radial Basis kernel function, with a value of 9,521 and 8.002%, respectively. The best average correlation is also obtained using the Radial Basis kernel function, with a value of 0.671.

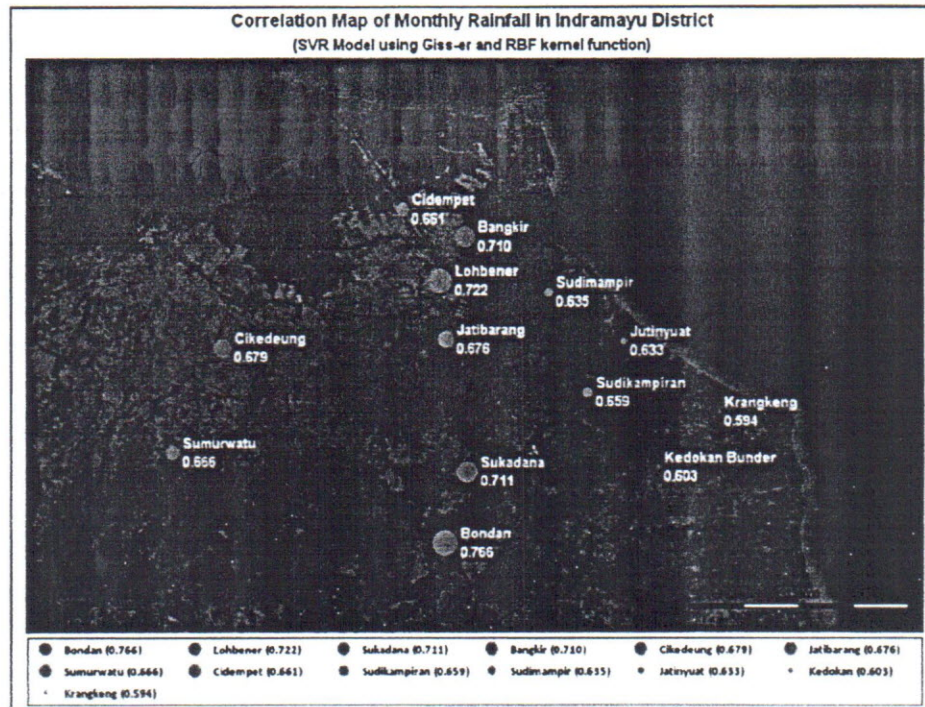


Figure 5: Map showing the correlation values for each of the station

We have also compared the performances of the models on each of the stations based on the geographical locations of the stations. Figure 5 shows the map of the 13 stations along with the correlation values associated with of each of the stations. Our results suggest that stations which are situated near shore lines and are directly affected by the sea weather conditions tend to have lower correlation values compared to the stations which are situated further from the sea. Specifically, the Bondan station, which are among the furthest stations from the sea in Indramayu, has the highest correlation value, 0.766. Meanwhile, the Krangkeng station, which is closest to the sea has the lowest correlation value, 0.594. However, deviations to this pattern exist.

5 CONCLUSION

We have performed an experiment to create a statistical downscaling model to predict the monthly rainfall using six GCM precipitations models, employing SVR as the regression method. The result shows a relatively good prediction values and promises a good prospects for further development of the method. Best prediction model results in a correlation value of 0.766. However, the prediction accuracy ranges between stations, and may be explained by the effects of their geographical locations and the surrounding atmospheric conditions.

ACKNOWLEDGEMENT

This research is part of the Project IM-HERE B2c IPB funded by the Government of the Republic of Indonesia through the contract No. 7/13.24.4/SPP/I-MHERE/2011.

REFERENCES

- [1] Wigena, A.H. 2006. Pemodelan Statistical Downscaling dengan Regresi Projection Pursuit untuk Peramalan Curah Hujan. Dissertation, Bogor Agricultural University, Bogor.

- [2] Cavazos, T., Hewitson, B. 2002. Relative performance of empirical predictors of daily precipitation. Proc. Of the 1st Biennial Meeting of the IEMSS, Lugano, Switzerland, 2:349:354.
- [3] V. Vapnik, S. Golowich, and A. Smola. Support vector method for function approximation, regression estimation, and signal processing. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 281–287, Cambridge, MA, 1997. MIT Press.
- [4] Von Storch, H., Zorita E., Cubash U., 1993. Downscaling of Global Climate Change Estimates to Regional Scales: An application to Iberian rainfall in wintertime. *J. Climate*: 6:1161-1171.
- [5] Sutikno. 2008. *Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi*. PhD Dissertation, Department of Agro-Meteorology, Graduate School, Bogor Agricultural University, Bogor.