

## Projection Pursuit Regression (PPR) on Statistical Downscaling Modeling For Daily Rainfall Forecasting\*

R P Putra<sup>1</sup>, D Anggraeni<sup>2</sup>, A F Hadi<sup>3</sup>

<sup>1,2,3</sup>Departemen of Mathematics, University of Jember, Indonesia

‡corresponding author: [putrario34@gmail.com](mailto:putrario34@gmail.com)

Copyright © 2021 R P Putra, D Anggraeni, and A F Hadi. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Rainfall forecasting has an important role in people's lives. Rainfall forecasting in Indonesia has complex problems because it is located in a tropical climate. Rainfall prediction in Indonesia is difficult due to the complex topography and interactions between the oceans, land and atmosphere. With these conditions, an accurate rainfall forecasting model on a local scale is needed, of course taking into account the information about the global atmospheric circulation obtained from the General Circulation Model (GCM) output. GCM may still be used to provide local or regional scale information by adding Statistical Downscaling (SD) techniques. SD is a regression-based model in determining the functional relationship between the response variable and the predictor variable. Rainfall observations obtained from the Meteorology Climatology and Geophysics Council (BMKG) are a response variable in this study. The predictor variable used in this study is the global climate output from GCM. This research was conducted in a place, namely Kupang City, East Nusa Tenggara because it has low rainfall. The Projection Pursuit Regression (PPR) will be used in this SD method for this study. In PPR modeling, optimization needs to be done and model validation is carried out with the smallest Root Mean Square Error (RMSE) criteria. The expected results must have a pattern between the results of forecasts and observations showing or approaching the observational data. The PPR model is a good model for predicting rainfall because The results of the forecast and observation show that the results of the rainfall forecast are observational data.

**Keywords:** daily rainfall forecasting, general circulation model, projection pursuit regression, statistical downscaling modeling.

---

\* Received: Jan 2021; Reviewed: Jan 2021; Published: Jun 2021

## 1. Introduction

Climate can be defined as weather habits that occur in a place or area which generally appear periodically or periodically and occur over a long period of time (Sucahyono & Ribudiyanto, 2013). The climate system can be known and understood using the General Circulation Model (GCM). GCM is the primary climate and weather prediction tool numerically and as a primary source of information for assessing the effects of climate change (Wigena, 2006). GCM is considered as an important model in the effort to understand the past, present, and future climate. Information on GCM is still on a global scale and the resolution is too low to predict local climate, so to obtain local or regional scale information, statistical downscaling techniques are used (Fernandez, 2005).

Statistical downscaling (SD) is a static downscaling process in which data on large-scale grids in a certain period and period of time is used as the basis for determining data on a smaller scale grid. The technique used in this study where the statistical downscaling approach uses a regression model in determining the functional relationship between the global scale climate variables GCM as a predictor variable with local climate variables as a predictor or response variable. The main advantage of this technique is that it is cheaper to compute and can easily be applied to the outputs of various GCM-based simulations and experiments (Sutikno et al., 2008).

Several statistical downscaling methods have been used, namely the Projection Pursuit Regression (PPR) method in the preparation of calibration models in the chemical field (Malthouse, 1995). Other than that Wigena (2006) examined rainfall forecasting in Indramayu with statistical downscaling using the Projection Pursuit Regression (PPR) method. In his research Anitawati (2010) shows that PPR has better results than PCR (Principal Component Regression). Based on the research that has been done, it can be concluded that PPR is able to overcome multicollinearity between predictor variables and the nature of the response variables which are nonlinear and have no standard distribution.

## 2. Method

### 2.1 General Circulation Model (GCM)

GCM is a computer-based model that consists of various numerical and deterministic equations that are integrated and follow the rules of physics. GCM is a major climate and weather prediction tool numerically and as a primary source of information to assess the effects of climate change (Wigena, 2006). GCM is able to simulate past, present, and future climates. Broadly speaking, GCM has been used and developed at the center of weather and climate forecasting in the world. The existing global circulation includes Goddard Institute for Space Studies (GISS) from National Aeronautics and Space Administration (NASA), United Kingdom Meteorological Office (UKMO), Geophysical Fluid Dynamic Laboratory (GFDL) from National Oceanic and Atmospheric Administration (NOAA), Commonwealth Scientific and Industrial Research Organization (CSIRO) from Australia, and National Centers for Environmental Prediction (NCEP).

## 2.2 Statistical Downscaling (SD)

Statistical downscaling (SD) is a static downscaling process in which data on large-scale grids in a certain period and period of time is used as the basis for determining data on a smaller scale grid. The technique used in this research is that the statistical downscaling approach uses a regression model to determine the functional relationship between climate variables on the global scale of GCM as a predictor variable with local climate variables as a predictor or response variable. The statistical downscaling model has the meaning of a transfer function that describes the functional relationship between the global atmospheric circulation and local climate elements (Sailor et al., 2000). In general, we can write a statistical downscaling model

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

where  $Y$  are local climate variables (eg rainfall) and  $X$  are GCM output variables (example precipitation). Whereas  $t$  is the time (for example, daily, monthly, yearly),  $p$  is the number of  $Y$  variables,  $q$  is the number of  $X$  variables,  $s$  is the number of atmospheric layers, and  $g$  is the number of GCM domain grids.

## 2.3 Projection Pursuit Regression (PPR)

Projection Pursuit Regression (PPR) is a regression model that has nonparametric and nonlinear properties in processing large-dimensional data that can provide information in small dimensions using projections. Projection Pursuit Regression can solve the problem of local averages, polynomial functions, and recursive partitioning. GCM data that is curse of dimensionality and multicollinearity can also be applied using this method.

The Projection Pursuit Regression method begins by maximizing the projection index, then empirically determining the single variable functions based on the optimum projections, and adding up the functions (Wigena, 2006). In their research Friedman & Stuetzle (1981) shows that the Projection Pursuit Regression method algorithm is as follows.

- a. Determine the initial residual value and the value of  $M$  or the number of functions

$$\begin{aligned} r_i &\leftarrow y_i, & i &= 1, 2, \dots, t \\ M &\leftarrow 0 \end{aligned} \quad (2)$$

where  $\sum y_i = 0$  (standardized response variable).

- b. Determine  $\alpha$  and function  $S_\alpha$  in the model

For linear combinations  $Z = \alpha_m X$ , determined smooth function  $S_\alpha = (Z)$  which conforms to the values  $Z$  by using a projection index  $I(\alpha)$  which can be written as follows.

$$I(\alpha) = 1 - \frac{\sum_{i=1}^t (r_i - S_\alpha(AX))^2}{\sum_{i=1}^t r_i^2} \quad (3)$$

- c. End of algorithm

If value  $I(\alpha)$  smaller than the threshold value, the Projection Pursuit Regression algorithm stops. However, if the opposite is the value  $I(\alpha)$  greater than the threshold value, the residual value and  $M$  are changed.

$$\begin{aligned} r_i &\leftarrow y_i - S_\alpha(Z), & i &= 1, 2, \dots, t \\ M &\leftarrow M + 1 \end{aligned}$$

(4)

In general, the Projection Pursuit Regression (PPR) model is as follows.

$$y_i = \sum_{m=1}^M S_{\alpha_m}(\alpha_m \mathbf{X}) = \bar{y} + \sum_{m=1}^M \beta_m f_m \left( \sum_{k=1}^p \alpha_{km} X_k \right) + \varepsilon_i \tag{5}$$

information :

- $y_i$  = response variable
- $\bar{y}$  = response variable mean
- $S_{\alpha_m}(\alpha_m \mathbf{X})$  = an unknown function
- $\alpha_m$  = unit vector (projection pursuit)  $(a_{1m}, a_{2m}, \dots, a_{km})$
- $\mathbf{X}$  = predictor variable  $(X_{i1}, X_{i2}, \dots, X_{ik})$
- $\varepsilon_i$  = random factor with  $E(\varepsilon_i) = 0$  dan  $var(\varepsilon_i) = \sigma^2$
- $m$  = many functions
- $i$  = unit of time
- $k$  = many predictor variable
- $X_i$  and  $\varepsilon_i$  free.

### 2.4 Validation Model

The criteria for the goodness of the model for validation can use the Root Mean Square Error Prediction (RMSE) value. RMSE is the square root of the mean squared error which has a minimum value of 0. The smaller the RMSE value, the smaller the difference between the estimated value of the modeling results and the actual data (Asyeifa, 2017). So that the best model is the model with the smallest RMSE value. The RMSE value is obtained from the formula:

$$RMSE = \sqrt{\frac{1}{n_1} \sum_{j=1}^{n_1} (y_j - \hat{y}_j)^2} \tag{6}$$

with  $y_i$  is a in-sample data to- $i$ ,  $\hat{y}_i$  is the estimated modeling result data in-sample data to- $i$ , and  $n_1$  is the sum in-sample.

### 3. Model Development

In this study using daily rainfall data for Kupang City, East Nusa Tenggara. Data at the Eltari observation station can be accessed through the BMKG website, namely <http://dataonline.bmkg.go.id> and data was taken from 1 January to 31 December 2019. GCM data is obtained through the official website of the Royal Netherlands Meteorological Institue (KNMI), namely <http://climexp.knmi.nl>. The variables in this study are predictor variables in the form of GCM data and response variables in the form of rainfall data for Kupang City. The analysis steps used in this study are in the following order.

1. Describe the data.
2. Forming a model of statistical downscaling using the Projection Pursuit Regression (PPR) method.

- a. Selection of the grid size starts from the lowest grid, which is  $3 \times 3$ , to the largest grid, which is  $10 \times 10$
  - b. Divide data into in-sample data and out-sample data. The proportion of division is 80:20, which means 80% of in-sample data and 20% of out-sample data.
  - c. Determine the number of functions in the Projection Pursuit Regression (PPR) model. The number of functions is determined from the simulation  $m = 1,2,3,4,5$ .
  - d. Creating a Projection Pursuit Regression (PPR) model based on in-sample data by simulating the number of functions  $m = 1,2,3,4,5$ .
3. Validate the statistical downscaling model.
  4. Identify relationships  $m$  with RMSE.
  5. Obtain the best statistical downscaling model.

#### 4. Result and Discussion

Modeling statistical downscaling with Projection Pursuit Regression (PPR) is performed by regressing data rainfall observation as response variable and output data GCM application as predictor variable. The following are the many function ( $m$ ) value for each grid presented in Table 1

Table 1: The Many Function ( $m$ ) Value For each Grid

Grid	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
$3 \times 3$	7.220980	6.618632	5.712351	5.971654	5.697723
$4 \times 4$	7.104985	6.398364	5.929931	5.552886	4.195765
$5 \times 5$	6.278040	5.798947	4.957509	4.987339	4.814503
$6 \times 6$	3.910466	2.861768	2.529859	1.931872	1.986956
$7 \times 7$	3.278983	3.009744	2.322622	2.052473	1.816421
$8 \times 8$	1.7335949	1.3403310	1.2827464	1.1657380	0.8765312
$9 \times 9$	1.1520267	0.8526537	0.7654405	0.6759548	0.6579353
$10 \times 10$	0.8175984	0.6872752	0.6291935	0.6136121	0.5878218

From here we can select the best grid by looking at the highest correlation (Table 2) based on the grid above, the  $3 \times 3$  grid is best. The last following are the results of the forecast (Figure 1). Based on the table obtained, it shows that the forecast results are close to the actual data with the RMSE value is 7.589626.

Table 2: The resulting correlation value.

Size	Correlation
3 × 3	0.369184141
4 × 4	0.001959057
5 × 5	0.192630016
6 × 6	0.129387519
7 × 7	0.200126340
8 × 8	0.093887779
9 × 9	0.201281292
10 × 10	0.030811146

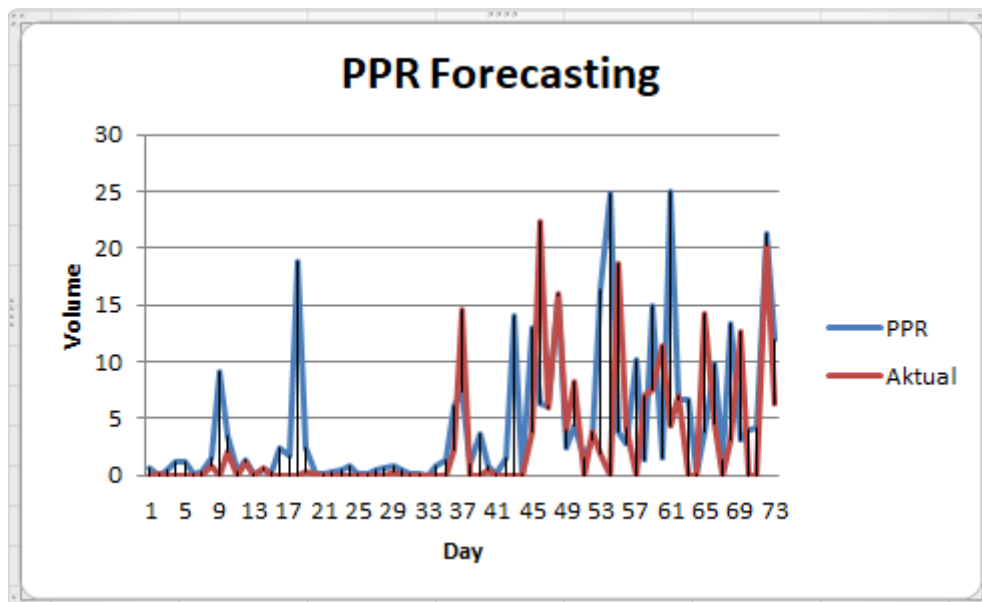


Figure 1: The Result Forecasting With PPR

### 5. Conclusion

The results of statistical model validation with the RMSE downscaling criteria show that the RMSE values in Kupang City, East Nusa Tenggara were obtained using several functions of  $m = 5$ . The pattern between the results of the forecast and observation shows that the results of rainfall forecasts in Kupang City, East Nusa Tenggara are close to the observational data with the RMSE value is 7.589626. So a model with many functions  $m = 5$  is a good model for forecasting rainfall in Kupang City, East Nusa Tenggara because it gets the smallest value.

### Acknowledgment

Researchers would like to thank the respondents at the University of Jember at the rectorate, faculty, departmental level as well as to LP2M who have provided their

support in this research.

## References

- Anitawati, M. (2010). *Pemodelan Statistical Downscaling Luaran GCM dengan Metode Principal Component Regression (PCR) dan Projection Pursuit Regression (PPR)* (PhD Thesis). Institut Teknologi Sepuluh Nopember.
- Asyeifa, V. R. (2017). *Pemodelan Statistical Downscaling Dengan Projection Pursuit Regression Untuk Meramalkan Curah Hujan Bulanan Di Sentra Produksi Padi Jawa Timur* (PhD Thesis). Institut Teknologi Sepuluh Nopember.
- Fernandez, E. (2005). On the influence of predictors area in statistical downscaling of daily parameters. *Norwegia Meteorological Institute*, 9: 1–21.
- Friedman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76(376): 817–823.
- Malthouse, E. C. (1995). *Nonlinear partial least squares*. Dissertation—Field of Statistics. Northwestern University.
- Sailor, D., Hu, T., Li, X., & Rosen, J. (2000). A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. *Renewable Energy*, 19(3): 359–378.
- Sucahyono, D., & Ribudiyanto, K. (2013). *Cuaca dan iklim ekstrim di Indonesia*. Penerbit Puslitbang, Badan Meteorologi, Klimatologi, dan Geofisika.
- Sutikno, B. R., Bey, A., Notodipuro, K., & Las, I. (2008). *Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi*. Desertasi—IPB.
- Wigena, A. (2006). *Pemodelan statistical downscaling dengan regresi projection pursuit untuk peramalan curah hujan bulanan* (Disertasi). IPB University, Bogor (ID).