ICoMathApp

INTERNATIONAL CONFERENCE
ON MATHEMATICS AND ITS APPLICATIONS 2020

# 1st International Conference on Mathematics and Its Application

**Strengthening Researches on Mathematics for the Challenge of Global Society**

Malang, East Java, Indonesia
September 30th, 2020
bit.ly/icomathapp

**Department of Mathematics**
**Universitas Negeri Malang**

PAPER • OPEN ACCESS

# Preface

To cite this article: 2021 *J. Phys.: Conf. Ser.* **1872** 011001

Journal of Physics: Conference Series **1872** (2021) 011001 doi:10.1088/1742-6596/1872/1/011001

# PREFACE

## The International Conference on Mathematics and its Applications (ICoMathApp) 2020

The International Conference on Mathematics and its Applications (ICoMathApp) 2020 is the first annual conference hosted by the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang. This conference was held virtually on September 30, 2020 using the platform Zoom. The aim of the conference is to provide a platform to the researchers, experts, and practitioners from academia, governments, NGOs, research institutes, and industries to meet and share cutting-edge progress in the fields of algebra, analysis, applied mathematics, combinatorics, computational sciences, geometry, and statistics. The ICoMathApp 2020 theme "Strengthening Researches on Mathematics for the Challenge of Global Society" is expected to give more contribution from the mathematical aspect as a response to the emerging of Covid-19 pandemic. This conference consists of a plenary and parallel session. The plenary session focuses on comprehensive reviews, concepts and perspectives. Specialized talks on recent developments are presented in the parallel session.

My highest appreciation for the four keynote speakers, Assoc. Prof. Arifah Bahar from Universiti Teknologi Malaysia, Malaysia; Prof. Purwanto, Ph.D from Universitas Negeri Malang, Indonesia; Prof. Hadi Susanto from Khalifa University, Abu Dhabi, & University of Essex, United Kingdom; and Assoc. Prof. Andrea Semaničová-Feňovčíková, Ph.D from Technical University of Kosice, Slovak Republic. My highest gratitude also goes to the invited speakers, Prof. Dr. Basuki Widodo, M.Sc from Institut Teknologi Sepuluh Nopember, Indonesia; Prof. Dr.rer.nat, Indah Emilia Wijayanti, S.Si., M.Si. from Universitas Gadjah Mada, Indonesia; Prof. Dr. Toto Nusantara, M.Si., from Universitas Negeri Malang, Indonesia; Dr. Swasono Rahardjo, M.Si, from Universitas Negeri Malang, Indonesia; and Dr. Desi Rahmadani, S.Si., M.Si. from Universitas Negeri Malang, Indonesia.

We would like to express our gratitude to all authors of contributed papers for participating excellently and eagerly. We hope that all participants in the ICoMathApp 2020 get a lot of insights and knowledge from this conference. We also would like to thank the reviewers for their positive contribution to maintain the quality of the articles presented in this conference. I want to also thank the committee members for their hard work, commitment, and dedication in organizing this conference. Our sincere gratitude also goes to the Journal of Physics: Conference Series IOP Publishing editors and coordinator for their helpful cooperation during the preparation of the proceedings.

**Mochammad Hafiizh, Ph.D**

Chairman

International Conference on Mathematics and its Application 2020

Journal of Physics: Conference Series **1872** (2021) 011001 doi:10.1088/1742-6596/1872/1/011001

# COMMITTEE

Scientific Committee

- Prof. Hadi Susanto (Khalifa University, Abu Dhabi, & University of Essex, UK)
- Assoc. Prof. Arifah Bahar (Universiti Teknologi Malaysia, Malaysia)
- Assoc. Prof. Andrea Semaničová-Feňovčíková, Ph.D (Technical University of Kosice, Slovak Republic)
- Prof. Purwanto, Ph.D (Universitas Negeri Malang, Indonesia)
- Prof. Dr. Toto Nusantara, M.Si (Universitas Negeri Malang, Indonesia)
- Prof. Dr. Basuki Widodo, M.Sc (Institut Teknologi Sepuluh Nopember, Indonesia)
- Prof.Dr. Agus Suryanto, M.Sc (Universitas Brawijaya, Indonesia)
- Prof. Dr. Surahmat, M.Si (Universitas Islam Malang, Indonesia)
- Prof. Dr. Indah Emilia Wijayanti, M.Si (Universitas Gajah Mada, Indonesia)
- Prof. Dr. Hadi Suwono, M.Si (Universitas Negeri Malang, Indonesia)
- Dr. Susiswo, M.Si (Universitas Negeri Malang, Indonesia)
- Santi Irawati, Ph.D (Universitas Negeri Malang, Indonesia)
- Tjang Daniel Candra, Ph.D (Universitas Negeri Malang, Indonesia)
- Dr. Makbul Muksar, M.Si (Universitas Negeri Malang, Indonesia)
- Dr. Hery Susanto, M.Si (Universitas Negeri Malang, Indonesia)

Chairman: Mochammad Hafiizh, Ph.D

Member:

1. Azizah, M.Si
2. Kridha Pusawidjayanti, M.Si
3. Mohammad Agung, M.Si
4. Dr. Desi Rahmadani
5. Vita Kusumasari, Ph.D

with the full support from all lecturers and staffs of Mathematics Department, Universitas Negeri Malang.

Journal of Physics: Conference Series

# Table of contents


## Volume 1872

2021

Previous issue Next issue

**1 st International Conference on Mathematics and its Applications (ICoMathApp) 2020 30 September 2020, Malang, Indonesia**

Accepted papers received: 23 March 2021
Published online: 14 May 2021






Open all abstracts

## Preface



## Combinatorics

## Algebra



## Analysis



## Statistics

## Applied Mathematics

Journal of Physics: Conference Series



# Projection pursuit regression and principal component regression on statistical downscaling using artificial neural network for rainfall prediction in Jember

To cite this article: C D Putri *et al* 2021 *J. Phys.: Conf. Ser.* **1872** 012023



# Projection pursuit regression and principal component regression on statistical downscaling using artificial neural network for rainfall prediction in Jember

**C D Putri, A F Hadi, D Anggraeni , and A Riski**

Data Science Research Group, Department of Mathematics, University of Jember, Jember, 68121, Indonesia.

E-mail: chandrikadesyanaputri@gmail.com, afhadi@unej.ac.id, dian_a.fmipa@unej.ac.id, riski.fmipa@unej.ac.id.

**Abstract**. Rainfall forecasting is essential for Indonesia, which is an agricultural country. Forecasting to see the rainfall needed to anticipate the danger of drought that will harm farmers. However, due to the complexity of topography and the interactions between the oceans, land, and atmosphere in Indonesia, it is difficult to predict rainfall. Therefore, Statistical Downscaling (SD) is needed to provide accurate rainfall predictions by considering the information about global atmospheric circulation obtained from the General Circulation Model (GCM). Statistics Downscaling (SD) modeling is a basic regression model based on the functional relationship between local scales, which is the response variable with the Global Circulation Model (GCM) global scale as a predictor variable. The Statistics Downscaling (SD) method used is Principal Component Regression (PCR) and Projection Pursuit Regression (PPR). The prediction of both methods was conducted by an Artificial Neural Network (ANN). The results showed that the prediction of rainfall in Jember using the PPR + ANN method (with the RMSE value of 79.58723) had better accuracy than the PPR, PCR, and PCR + ANN methods, which had RMSE values of 103.7539, 112.337 and 83.62029, respectively.

## 1. Introduction

Indonesia is known as a maritime country because most of Indonesia's territory is water. However, apart from being a maritime country, Indonesia is known as an agricultural country because the agricultural sector is the livelihood of most of its population. The agricultural sector is very vulnerable to climate change. Climate can be defined as weather habits in a place or area that generally appear periodically and occur over a long period[9]. Climate changes affect almost all agricultural aspects, such as cropping patterns, planting time, production yields, and yield quality. Climate change is dependent on several factors, one of which is rainfall. Rainfall is the most dominant climate factor in characterizing Indonesian climate conditions because rainfall in Indonesia has high diversity and fluctuation compared to temperature[8]. High and low rainfall significantly affect crop yields. Therefore, it is necessary to forecast rainfall to provide information about climate changes that affect the agricultural sector.

Indonesia is a tropical country that is crossed by the equator, so that Indonesia has a small diversity of temperatures and a high diversity of rainfall. Rainfall forecasting is a difficult challenge for researchers due to the complex topography and interactions between the oceans, land, and atmosphere. Therefore, we need an accurate rainfall forecasting model at the local scale by considering the global atmospheric circulation obtained from the output of the General Circulation Model (GCM). GCM is a climate model for understanding the climate through simulations in order to make a reasonable projection of future climate change based on emission change scenarios. GCM simulates global climate variables on each grid (measuring ± 2.5 ° or ± 300 km) for each layer of the atmosphere, predicting climate patterns in an annual period[11]. Information about the climate from the GCM output is still on a global scale, and we do need to obtain climate information on a local scale with high resolution, namely Statistical Downscaling[10].

Statistical downscaling is a rainfall forecasting model that utilizes information about the global atmospheric circulation from the GCM output to obtain information on a local scale. Statistical downscaling determine the functional relationship between local rainfall and the global climate of GCM[3]. However, the common problem in downscaling statistical is the dimensional size of the GCM output data, which is very large. So it is necessary to make a dimensional reduction. The relationship between the predictor variables on each grid shows a linear relationship, which indicates multicollinearity that must be avoided in statistical downscaling. The response variables and predictor variables in the rainfall study in Jember showed a non-linear relationship, so that an appropriate statistical downscaling method was needed to solve all these problems. The statistical methods used in the statistical downscaling model include Projection Pursuit Regression (PPR), Principal Component Regression (PCR), and Artificial Neural Network (ANN).

Projection Pursuit Regression (PPR) is a scaling down method that combines regression analysis with projection pursuit (PP). In a previous study[7], monthly rainfall forecasting was carried out using the PPR method at the center of rice production areas in East Java. Principal component regression is a reduction method that combines regression analysis with Principal Component Analysis (PCA). Previous research has conducted rainfall predictions in Jember Regency using the Principal Component Regression (PCR) method combined with ARIMA[1]. Another method used to estimate the SD model is an Artificial Neural Network (ANN). The research aims to see which model has the optimal performance in predicting rain in the district. Based on the research before, the researcher wants to develop a research model regarding rainfall prediction in Jember using a statistical downscaling model with several methods, namely PPR and PCR with ANN as prediction models. Then will be compared which method will give the best performance in predicting rainfall in Jember.

## 2. Material and Method

### 2.1. Study Region

The research located was in Kabupaten Jember. Kabupaten Jember is one of the districts in East Java which has an astronomical location with $113° \, 16' \, 28''$ E to $114° \, 3' \, 42''$ E longitude and $7° \, 59' 6''$ S to $8° \, 33' 56''$ S latitude. Jember Regency has an area of 3,293.34 km2 with a topographical character of fertile canyon plains in the middle and south and surrounded by mountains that extend the western and eastern borders. Kabupaten Jember was a tropical climate with a temperature range between 23ºC – 32ºC. The southern part of Jember is lowland, with the outermost point is Barong Island. Temperature figures range from 23ºC – 31ºC, with the dry season occurring from May to August and the rainy season from September to January. Simultaneously, the rainfall is quite a lot, which ranges from 1.969 mm to 3.394 mm[2].

### 2.2 Data Description

Monthly rainfall in Jember was surveillance at 77 observation station points with the period 2005-2018. Following the monthly local rainfall data, the year period for GCM output data is a predictor variable is January 2005 to December 2017. The variable used in the GCM output data is the variable precipitation. In this GCM output data, we need to take several domain sizes, which would be used as spatial data. The domain size was incrementally from the smallest to larger, namely from the smallest 3 × 3 grid size to the largest 10 × 10 grid size with a latitude range of -21.25 ° to 3.75 ° N and longitude ranging from 101.25 ° to 126.25 ° E. GCM output data was downloaded from http://climexp.knmi.nl.

*2.3 Method Description*

*2..3.1. Principal Component Analysis (PCA)*

Multiple variable analysis often requires dimensional reduction due to large data sets, so that PCA (Principal Component Analysis) is required. PCA is a multivariate analysis that transforms the original correlated variables into new, uncorrelated variables by reducing the number of these variables to have smaller dimensions but can explain most of the diversity of the original variables. Here is a linear combination of the principal components.

$$\begin{aligned} Y_1 &= e_1'X = e_{11}'X_1 + e_{21}'X_2 + \cdots + e_{p1}'X_p \\ Y_2 &= e_2'X = e_{12}'X_1 + e_{22}'X_2 + \cdots + e_{p2}'X_p \\ &\vdots \\ Y_p &= e_p'X = e_{1p}'X_1 + e_{2p}'X_2 + \cdots + e_{pp}'X_p \end{aligned} \quad (1)$$

with $Var(Y_i) = e_i' \sum e_i$ dan $Cov(Y_i, Y_k) = e_i' \sum e_i$ . The requirement to form the main component which is a linear combination of the variable $X$ in order to have a maximum variance is to choose a feature vector (eigen vector), namely $e = (e1, e2, \dots, ep)$ such that Var($Yi$) = $ei$'Σ$ei$ is maximum with diversity cumulative reaches 75% and $ei'ei = 1$

*2.3.2 Principal Component Regression (PCR)*

PCR is a method usually use to solve multicollinear problems. This method will produce uncorrelated main components. Here we note that if all the main components are included in the regression model, the resulting model will be the same as that obtained by the least-squares method. If only a few principal components are included, the regression coefficients will be a biased estimator but still has minimum variance[5]. The use of principal component analysis will produce new variables, which are linear combinations of the original predictor variables, and these new variables are mutually independent. The general equation for PCR is:

$$Y_{R_{(t)}} = \div \beta_0 + \sum_{i=1}^{m} \beta_i C_i \quad (2)$$

with
$Y_{R(t)}$ = the response variable
$\beta_0$ = intercept value
$\beta_i$ = the coefficient for the i-th component
$C_i$ = the i-th principal component
$m$ = the number of main components

*2.3.3 Projection Pursuit (PP)*

Projection Pursuit (PP) is a dimensional reduction method underdispersived based on searching for a projection of primary information from large-dimensional data[11]. The dimension reduction procedure using PP is carried out based on the optimum projection index, in contrast to the dimensional reduction using PCA, which is carried out based on the largest variance.

If $X = \{x_1, x_2, \dots, x_p\}$ is a matrix of predictor variables with dimension p, then the linear projection $\Re^p \rightarrow \Re^k$ is as follows.

$$Z^t = AX^t, X \in \Re^p, Z \in \Re^k, k < p \quad (3)$$

Dimensional reduction with PP uses the projection index $\boldsymbol{I(F_A)}$ ) to obtain the maximum matrix **A** through numerical optimization of the parameters.

$$I(Z) = I(AX^T) = I(A) \tag{4}$$

The projection index $\boldsymbol{I(A)}$ can be obtained from the following equation.

$$I(A) = 1 - \frac{\sum_{i=1}^{t}\left(r_i - S_\alpha(AX)\right)^2}{\sum_{i-=1}^{t} r_i^2} \tag{5}$$

*2.3.4 Projection Pursuit Regression (PPR)*

Projection Pursuit Regression (PPR) is a nonparametric and non-linear regression method for processing data in large dimensions that can describe information in small dimensions through a projection process. Based on Friedman and Stuetzle[4], the PPR algorithm is as follows.

1. Determine the initial residual value and the value of M (number of functions).

$$r_i \leftarrow y_i,\ i = 1,2,\dots,t \qquad M \leftarrow 0 \tag{6}$$

   where $\sum y_i = 0$. The number of functions is determined based on the optimization of the function m = 1, 2, 3, 4, and 5. The determination of the number of functions is selected based on the best validation results.

2. Determine $\boldsymbol{\alpha}$ and $S_\alpha$ in the model.
   For the linear combination $\boldsymbol{Z} = \boldsymbol{\alpha}_m\boldsymbol{X}$ , the smoothing function $S\_\alpha(Z)$ is determined according to the **Z** values using the projection index $\mathbf{I}(\boldsymbol{\alpha})$. The projection index $\mathbf{I}(\boldsymbol{\alpha})$ can be calculated using the following formula.

$$I(\alpha) = 1 - \frac{\sum_{i=1}^{t}\left(r_i - S_{\alpha(\alpha x_i)}\right)^2}{\sum_{i=1}^{t} r_i^2} \tag{7}$$

   Determine the coefficient vector $\alpha_{M+1}$ that maximizes $I(a)$ (*projection pursuit*), $\alpha_{M+1} = max_\alpha^{-1}(I(\alpha))$ and the smoothing function is $S_{\alpha_{M+1}}(Z)$.

3. End of algorithm
   If $\mathbf{I}(\boldsymbol{\alpha})$ is smaller than the threshold value, then stop. However, if $\mathbf{I}(\boldsymbol{\alpha})$ is greater than the threshold value, the residual value and M value are changed as follows:

$$r_i \leftarrow r_i - S_\alpha(Z),\ i = 1,2,\dots,t \qquad M \leftarrow M + 1 \tag{8}$$

   And return to the previous step. The threshold value is obtained from the linear combination boundary in the scatterplot between the response variable and the predictor variable. The following is the end of the PPR algorithm:

$$y_i = \sum_{m=1}^{M} S\_\alpha_m(a_m X) \tag{9}$$

$$= \beta_0 + \sum_{m=1}^{M} \beta_m f_m\left(\sum_{k=1}^{n} \alpha_{km} X_{ik}\right) + \varepsilon_i$$

$S\_\alpha_m(a_m X)$ is an unknown function and $\alpha_m = \alpha_{1m}, \alpha_{2m}, \dots, \alpha_{km}$ is a unit vector where m is the basis of the function. $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ is the k-th predictor variable and the i-th, $y_i$ is the response variable, $\varepsilon_i$ is a random factor with $E(\varepsilon_i) = 0$ and $var(\varepsilon_i = \sigma^2)$.

*2.3.5 Artificial Neural Network (ANN)*

Artificial Neural Network Artificial (ANN) is an information processing technique that works like a biological nervous system in human brain cells in processing information. This processing technique has a crucial element, the information processing system, which is unique and varies for each application. Neural Network consists of many information processing elements (neurons) connected and works together to find solutions to particular problems such as classification or prediction problems.

Figure 1 is an illustration of Artificial Neurons and modeling of a multilayered neural network. Based on Figure 1, the signal flow from the input $x_1, x_2, x_3, \dots x_n$ is considered to be in the direction indicated by the arrow, and $w_1, w_2, w_3 \dots w_n$ are the weights related to the corresponding input.

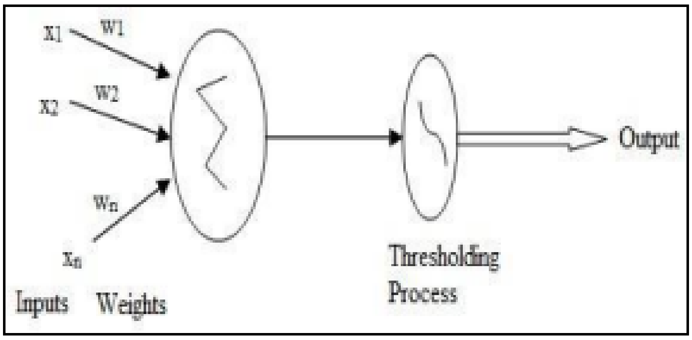


**Figure 1.** Artificial Neural Network Architecture[6].

The acceleration or deceleration of the input signal is modeled by the weight used to calculate its strength. Therefore, the total input received is:

$$I = w_{1\times 1} + w_{2\times 2} + \dots + w_{n\times n} \tag{10}$$

The sum is passed on to non-linear filter called Activation Function Φ and forms a thresholding process. The number will then be compared with the threshold value $\theta$, if $I > \theta$ then the output is 1 else 0.

$$y = \phi(I)$$

$$y = \sum_{i=0}^{n} w_i x_i - \theta \tag{11}$$

## 3. Model Development

The predictor variables for the GCM output data follow from local rainfall data; the GCM output data used is 2005-2018 (168 months). For the local rainfall data of Kabupaten Jember, it is taken from the 2005-2018 period. These data are divided into out-sample and in-sample data. The test bulk data from 2005-2016 (144 months) were used as in-sample data, and for out-sample data, data from the 2017-2018 year period (24 months) were taken. The domain size used in this study were each 3x3, 4x4, 5x5, 6x6, 7x7, 8x8, 9x9, 10x10, and then be adjusted by the astronomical location of the Jember. Each domain size has a total of 9, 16, 25, 36, 49, 64,81, and 100 predictor variables.

In statistical downscaling modeling with the PPR method, there are several stages in the modeling. The first stage determines the number of optimum functions used in determining the optimal domain size. The optimum number of functions will be selected for each grid by looking at the RMSE value

**Table 1**. Result of optimum function on PPR models.

| Domain size | Validation models | Number of function $(m)$ | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| $3 \times 3$ | RMSE | 78.70853 | 69.96927 | 55.95357 | **42.25480** | 51.50790 |
| $4 \times 4$ | | 73.10434 | 69.80424 | 65.32772 | 56.01909 | **23.95734** |
| $5 \times 5$ | | 63.14675 | 57.55973 | 36.55114 | 32.48607 | **18.83486** |
| $6 \times 6$ | | 49.90116 | 30.97867 | 21.27827 | 18.98777 | **15.72641** |
| $7 \times 7$ | | 41.388234 | 22.464495 | 16.153983 | 12.255301 | **6.099832** |
| $8 \times 8$ | | 45.571702 | 18.718511 | 6.905575 | 3.176746 | **1.418566** |
| $9 \times 9$ | | 48.992543 | 38.821391 | 2.294679 | 1.366185 | **0.928577** |
| $10 \times 10$ | | 17.7852544 | 6.9505160 | 2.1768314 | 0.4319677 | **0.2164227** |

After each grid's function has been determined, the next step is to determine the optimum domain size by looking at the highest correlation value on each grid. A high correlation value indicates that the grid size is the optimum grid size.

**Table 2.** Result of performance (correlation) statistic in different grid size on PPR model.

| Domain Size | Corelation |
|---|---|
| 3x3 | 0,6580940 |
| 4x4 | 0,6712881 |
| 5x5 | 0,6523241 |
| **6x6** | **0,6556798** |
| 7x7 | 0,6341951 |
| 8x8 | 0,6210702 |
| 9x9 | 0,6107696 |
| 10x10 | 0,5853008 |

Based on the correlation table above, it can be seen that the optimum grid in the PPR method is the 6x6 grid size with the highest correlation value, namely 0.6556798. Meanwhile, in modeling the statistical downscaling using PCR, the number of variable components used depends on the eigenvalues; if the eigenvalues have met the 75% target, then the principal components used are sufficient. The number of components used in PCR modeling will be processed for the next stage, namely determining the optimum domain size as seen from the correlation value in each grid.

\

**Table 3**. Result of performance (correlation) statistic in different grid size on PCR model.

| Domain Size | Corelation |
|---|---|
| 3x3 | 0,6580940 |
| **4x4** | **0,6712881** |
| 5x5 | 0,6523241 |
| 6x6 | 0,6556798 |
| 7x7 | 0,6341951 |
| 8x8 | 0,6210702 |
| 9x9 | 0,6107696 |
| 10x10 | 0,5853008 |

Based on the table above, it can be seen that the grid size that has the highest correlation value is the 4x4 grid so that the 4x4 grid size is the optimum domain size for PCR modeling.

## 4. Result and Discussion

In statistical downscaling modeling using the Projection Pursuit Regression (PPR) method, the optimum domain is obtained on a $6 \times 6$ grid with the optimum number of functions $m = 5$, which has an RMSE value of 0.7374709. Furthermore, after determining the optimum domain size, a PPR model is built between the rainfall observation data and the GCM output data. The results of the rainfall forecast in Jember were obtained for the $2017-2018$ period from the previous stage. Not only modeling using PPR, but also modeling statistical downscaling using PPR and ANN methods that function as a smoothing function. Then obtained the results of the rainfall forecast for the $2017-2018$ year period using the PPR + ANN method. The next step is validating the PPR model PPR + ANN funds using out-sample data by comparing the RMSEP value. In statistical downscaling modeling using the PPR method, the RMSE value is 103.7539. For modeling using a combination of the PPR and ANN methods, there is an increase in model performance seen from the RMSE value, which is smaller than the RMSE value from the PPR method, namely 79.58723. Based on Figure 2, it can be seen that the plot of the PPR + ANN method is closer to the plot of the actual data so it can be said that the PPR + ANN method is better than the PPR method.

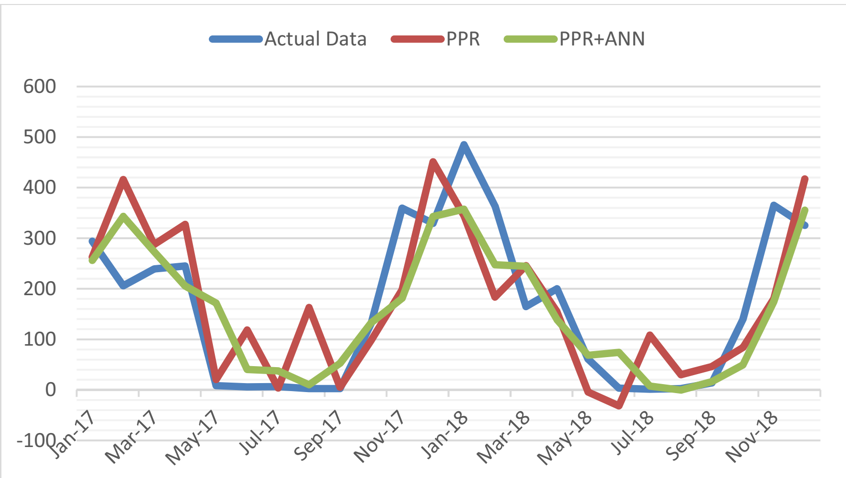


**Figure 2.** The plot of rainfall forecasting using PPR and PPR+ANN models.

The optimal domain size in modeling with the PCR method is a 4 × 4 domain with 16 predictor variables. The PCA method was used to extract orthogonal principal components by retaining more than 75% variance. So that in the next stage the optimum grid size is obtained 4 × 4 by looking at the highest correlation value. The next stage is to build a model using the PCR method between local rainfall observation data and GCM output data, and the results of rainfall forecasts are obtained in Jember Regency for the period 2017-2018. In addition to modeling with PCR, modeling is also carried out by combining PCR and ANN methods. The next step was validating the PCR model and the PCR + ANN using out sample data by comparing the RMSEP value. The PPR method's RMSEP value is 112,337, while the RMSEP value of the PPR + ANN method is 83.62029 so that the PPR + ANN method has more optimal performance in rainfall forecasting compared to the PCR method. Based on Figure 3, the plot of the PCR + ANN method is closer to the plot of the actual data.

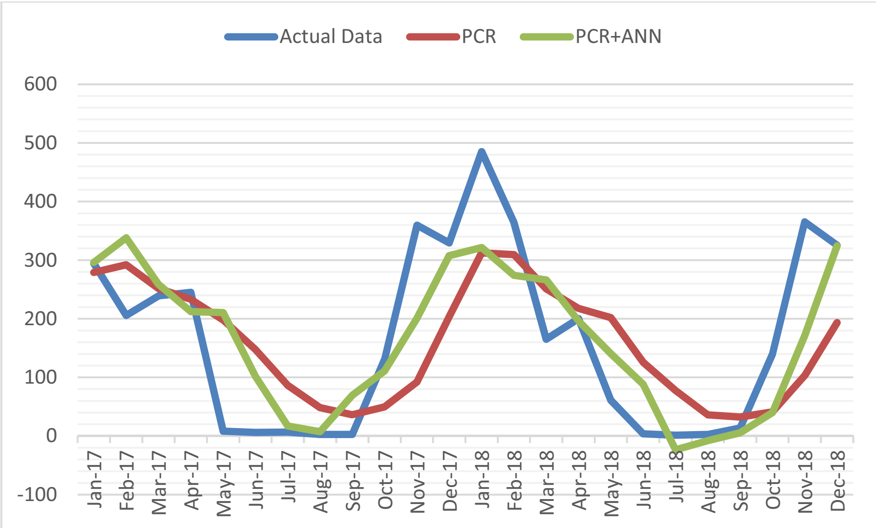


**Figure 3.** The plot of rainfall forecasting using PCR and PCR+ANN model.

In Table 4, the forecast results of each statistical downscaling model are obtained, namely PPR, PPR + ANN, PCR, and PPR + ANN from January 2017 to December 2018.

**Table 4.** The result of rainfall forecasting on each model of Statistical Downscaling.

| Periode test | Actual Data | PPR | PPR_ANN | PCR | PCR_ANN |
|---|---|---|---|---|---|
| Jan-17 | 294,1818 | 262,662470 | 256,2014645 | 279,16122 | 296,31896 |
| Feb-17 | 205,7403 | 415,996499 | 343,2100269 | 291,99043 | 338,21080 |
| Mar-17 | 239,1039 | 287,368591 | 273,2710707 | 251,54448 | 258,47976 |
| Apr-17 | 245,1429 | 327,473135 | 205,5668609 | 233,40631 | 212,09383 |
| Mei-17 | 8,24545 | 20,564512 | 171,8240208 | 197,97161 | 210,40413 |
| Jun-17 | 6,01948 | 118,471452 | 40,4452132 | 147,61427 | 102,02642 |
| Jul-17 | 6,6364 | 3,874141 | 37,5464200 | 86,54778 | 17,08004 |
| Agu-17 | 2,5714 | 163,136201 | 10,0807324 | 48,31236 | 7,48314 |
| Sep-17 | 2,67792 | 5,202689 | 52,6427973 | 36,42148 | 69,50363 |
| Okt-17 | 131,3766 | 97,601821 | 132,8618990 | 49,67560 | 111,17679 |

| | | | | | |
|---|---|---|---|---|---|
| Nov-17 | 359,4935 | 198,073421 | 181,0487558 | 92,23755 | 200,98741 |
| Des-17 | 329,4935 | 451,215129 | 342,9339555 | 203,65853 | 307,85284 |
| Jan-18 | 485,0128 | 344,834636 | 357,3399022 | 312,47126 | 321,47973 |
| Feb-18 | 363,8718 | 183,620014 | 247,4912141 | 309,50899 | 274,08332 |
| Mar-18 | 164,9231 | 245,898806 | 244,2280915 | 251,03967 | 266,35879 |
| Apr-18 | 200,1667 | 154,458823 | 138,9676265 | 217,58718 | 196,28155 |
| Mei-18 | 60,9359 | -4,073062 | 68,5261666 | 201,90690 | 140,55802 |
| Jun-18 | 3,5859 | -31,418982 | 74,3438076 | 125,77258 | 88,11286 |
| Jul-18 | 1,4487 | 108,523183 | 7,4233488 | 77,83181 | -23,45501 |
| Agu-18 | 2,5513 | 30,176993 | -0,1807121 | 36,16295 | -7,86042 |
| Sep-18 | 13,7179 | 46,379388 | 16,3349645 | 32,27656 | 5,49928 |
| Okt-18 | 139,7692 | 83,947006 | 49,3318250 | 41,18409 | 39,80318 |
| Nov-18 | 365,2308 | 180,402859 | 174,6601490 | 103,26893 | 170,94785 |
| Des-18 | 325,359 | 417,240272 | 355,3202809 | 193,57789 | 324,34774 |

The RMSE value of each model can be seen in table 4. In the table, it can be seen that the smallest RMSE value is obtained by the PPR + ANN method with an RMSE value of 79.58723.

**Tabel 5.** The value of RMSE on each model of Statistical Downscaling.

| No | Models | RMSE |
|---|---|---|
| 1 | PPR | 103.7539 |
| 2 | PPR+ANN | 79.58723 |
| 3 | PCR | 112.337 |
| 4 | PCR+ANN | 83.62029 |

Figure 4 shows that the plot of the PPR + ANN method is closer to the plot of the actual data than the other methods.

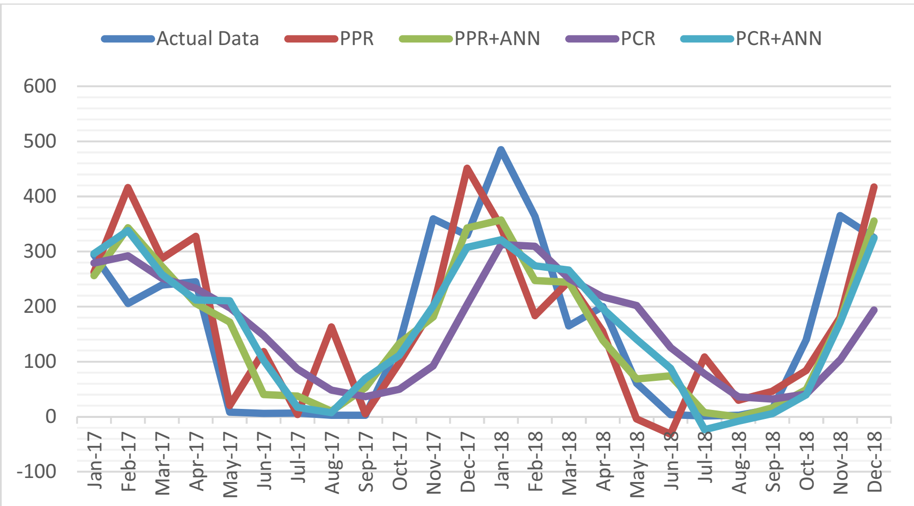


**Figure 4.** Comparison plot of each rainfall forecast model with actual data.

## 5. Conclussion

The best performing model was PPR + ANN with the lowest RMSE value of 79.58723 compared to the PPR, PCR, and PCR + ANN models. So it is expected that the PPR + ANN model be the promising hybrid forecasting model for mountly rainfall in Jember. However, ANN still experiences several weaknesses, such as the need for a large number of controlling parameters, difficulty obtaining a stable solution, difficulty in excluding insignificant inputs, and the danger of over-fitting. Further research is needed using other predictive models such as SVR to obtain the best model in rainfall forecasting.

## Acknowledgment

We would like to thank the Ministry of Research and Technology/National Research and Innovation Agency of Indonesia, especially the Directorate of Research and Community Empowerment (DRPM). The DRPM funded this research with ID contract No. 175/SP2H/AMD/LT/DRPM/2020. We also thank all Data Science Research Group members at the Department of Mathematics, University of Jember.

## References

[1] Asyeifa V R 2017 Pemodelan Statistical Downscaling Dengan Projection Pursuit Regression Untuk Meramalkan Curah Hujan Bulanan Di Sentra Produksi Padi Jawa Timur *Thesis* (Surabaya: Institut Teknologi 10 Nopember).

[2] Badan Perencanaan Pembangunan Kabupaten Jember [BAPPEDA] and Badan Kerja Politeknik Negeri Jember [POLIJE] 2015 *Laporan Akhir Penyusunan Masterplan Pertanian Jember* (Jember: Badan Perencanaan Pembangunan Kabupaten Jember).

[3] Fernandez E 2005 *On the Influence of Predictors Area in Statistical Downscaling of Daily Parameters* (Norwegia: Norwegia Meteorological Institute) chapter 9 pp 1–21.

[4] Friedman J H and Stuetzle W 1981 Projection Pursuit Regression. *Journal of The American Statistical Association.* **376** pp 817-823.

[5] Jollife I T 2002 *Principal components analysis* 2nd ed (New York: Springer-Verlag).

[6] Patil N S, Laddimath R S, Pooja and Hooli S 2016 Downscaling of Precipitation Data from GCM outputs using Artificial Neural Network for Bhima basin *Appl. Envirnmental Sci* **10(4)** p 1493–1508.

[7] Salsabila I 2019 Principal Component Regression dan Arima pada Teknik Statistical Downscaling untuk Peramalan Curah Hujan di Kabupaten Jember *Thesis* (Jember: Jember Universty).

[8] Sipayung S 2005 Dampak Variabilitas Iklim Terhadap Produksi Pangan di Sumatera *J. Sains Dirgant* **2(2)** p 111–126.

[9] Sucahyono D and Ribudiyanto K 2013 *Cuaca Dan Iklim Ekstrim* (Jakarta:Pusat Penelitian dan Pengembangan BMKG) p 275.

[10] Sutikno 2008 Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi *Disertation* (Bogor: Bogor Agricultural Institute).

[11] Wigena A H 2006 Pemodelan Statistical Downscaling Dengan Regresi Projection Pursuit Untuk Peramalan Curah Hujan Bulanan *Disertation* (Bogor: Bogor Agricultural Institute).