BAREKENG: Journal of Mathematics and Its Applications
December 2023 Volume 17 Issue 4 Page 1969–1984
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss4pp1969-1984

COMPARISON OF B-SPLINE AND TRUNCATED SPLINE REGRESSION MODELS FOR TEMPERATURE FORECAST

Sri Sulistijowati Handajani [1*], Hasih Pratiwi[2], Respatiwulan[3], Niswatul Qona'ah[4], Monica Ramadhania[5], Niken Evitasari[6], Nindya Eka Apsari[7]

[1,2,3,4,5,6,7]*Statistics Study Program FMIPA Universitas Sebelas Maret*
St. Ir. Sutami 36A, Surakarta, 57126, Indonesia

Corresponding author's e-mail: * **rr_ssh@staff.uns.ac.id**

ABSTRACT

The spline regression model is a nonparametric model and it is applied to data that do not have a certain curve shape and information about it. In this study, the results of the analysis of the B-spline regression model and the truncated spline model were compared on temperature data at several stations on Java Island to obtain the best model that can be used to forecast the temperature for the next few days. Daily temperature data were obtained from BMKG at Semarang, Juanda, Serang, Sleman, Bandung, and Kemayoran stations. The temperature data were modeled with the B-spline and truncated spline regression using the optimal knot point of the GCV, and the best model was obtained. The analysis shows that the B-spline regression models are better than the truncated spline models with a fairly small MSE value and a greater coefficient of determination than the truncated spline model.

Article History:

Received: 31st May 2023
Revised: 6th September 2023
Accepted: 5th October 2023

Keywords:

B-spline regression model
GCV;
R^2;
temperature;
truncated spline.

How to cite this article:

S. S. Handajani, H. Pratiwi, R., N. Qona'ah, M. Ramadhania, N. Evitasi and N. E. Apsari., "COMPARISON OF B-SPLINE AND TRUNCATED SPLINE REGRESSION MODELS FOR TEMPERATURE FORECAST," *BAREKENG: J. Math. & App.*, vol. 17, iss. 4, pp. 1969-1984, December, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Research on temperature has become something interesting to learn about current climate change. The temperature reflects the effect of climate change on Earth and its surrounding atmosphere [1]. Recently, climate change has caused extreme natural phenomena such as heat waves, severe winters, heavy snowfall, and droughts worldwide, leading to environmental and health crises [2]–[5]. Temperature prediction helps meteorologists to know the likelihood of hurricanes and floods in an area [6]. According to the BMKG, the high rainfall in Indonesia during the current dry season is influenced by La Nina. However, it is also affected by the continued warming of sea surface temperatures in Indonesian waters following the global warming trend. The contribution of warming sea surface temperatures is five times greater than the warming of sea surface temperatures due to La Nina in Indonesia [7]. It was further explained that the trend of continued warming of sea surface temperatures would supply more water vapor that can trigger cloud growth and heavier rain. This has implications for increasing the risk of hydrometeorological disasters in Indonesia. One of the sectors most affected by the increase in disasters is agriculture, specifically food crop production. High rainfall causes the potential for flooding to increase, which can damage agricultural land and the flowering process, so it happens to fail the fertilization process. Therefore, temperature modeling is important to forecast on Java Island. Various statistical approaches have been proposed recently, several of which have been developed in the econometric literature. The statistical models have applied i.e., cointegration approaches, which determine the relationship between non-stationary and stationary time series [8], and regression approaches, which evaluate the characteristics of time series for a given temperature data [9], [10]. However, since temperature data do not form a specific pattern, the non-parametric regression models, i.e., B-spline and truncated spline, are used to forecast the temperature in Java Island [11].

The regression models explain the relationship between the response variable and the predictor variables. However, if we have a time series variable, the model can be formed into a time series regression model or a time series model in the form of a forecasting model, one of which is the exponential smoothing model or the ARIMA model.

If the regression curve is a parametric model, according to [12], the parametric estimation will be efficient, but otherwise, it will give a misleading interpretation. Therefore, if the shape of the curve $g(Y_{t-k})$ is unknown and there is no information about it, a non-parametric approach will be more appropriate. Suppose the model is formed in the classical regression model for building a simpler model. In that case, the results obtained according to [13] will produce a biased model and cause a significant error. By non-parametric regression approach, it is hoped that the data will find their own estimated form of the regression curve without being influenced by the subjectivity of the researcher so that the prediction model is closer to the actual model. According to [14], non-parametric regression has better flexibility in modeling data patterns.

There are several estimation methods in nonparametric regression, such as spline estimation, kernel estimation, wavelet estimation, etc. The estimation of the spline approach, as in [15], has a base function, such as truncated spline and B-spline. There are two methods i.e., truncated spline and B-spline models, that are applied in this paper to forecast the temperature of some stations in Java Island. The nature of splines in that models is that they have high flexibility and can estimate changes in data behavior at different intervals, as indicated by the presence of slices attached to the parameters to be estimated [16].

2. RESEARCH METHODS

The data used are daily temperature sourced from the official website of the Meteorology, Climatology, and Geophysics Agency (BMKG). The temperature data were obtained from six stations on Java Island, including Semarang, Juanda, Serang, Sleman, Bandung, and Kemayoran stations. The data from Semarang station are from 2019 to 2021, while for the other stations, the daily data is from 2021. The response variable is temperature data (t time data) from the six stations, while the predictor variable is temperature data (time t-1).

Stages of analysis carried out to model temperature data begin with making a scatterplot between the response variable and predictor variable for each station, then determining the order and optimal knot point based on the minimum GCV value. The next step is to determine the parameter estimates, then build the model and select a model, and select the B-spline and truncated spline models based on the criteria for model goodness. The best model obtained is then interpreted and used to make predictions, for example the next ten days.

2.1 B-Spline Regression Model

The regression curve in the equation below, when approached with a B-spline function of order m with k knots, can be as **Equation (2) [17]**:

$$\mu(y_t) = \sum_{j=1}^{m+k} b_j N_{j-m,m}(y_t), \quad t = 1,2,\dots,n \tag{1}$$

where $\boldsymbol{N_{j-m,m}(y)}$ is the basis of B-spline and $\boldsymbol{b_j}$ is the regression parameter for B-spline. From **Equation (1)**, the B-spline non-parametric regression model is written in the form:

$$y_t = \sum_{j=1}^{m+k} b_j N_{j-m,m}(y_{t-1}) + \varepsilon, \quad i = 1,2,\dots,n$$

According to **[18]**, to build a B-spline function of order m with k knot points $\boldsymbol{\xi_1,\dots,\xi_k}$ where $\boldsymbol{a_0 \le \xi_1 \le \dots \le \xi_k \le a_1}$, first defined additional 2m knots, namely

$\boldsymbol{\xi_{-(m-1)},\dots,\ \xi_{-1},\ \xi_0,\ \xi_{k+1},\dots,\ \xi_{k+m}}$, where $\boldsymbol{\xi_{-(m-1)} = \dots = \xi_0 = a_0}$ and $\boldsymbol{\xi_{k+1} = \dots = \xi_{k+m} = a_1}$.

The basis of the B-spline function of order m with knot points in $\boldsymbol{\xi_i}$ where $\boldsymbol{i = -(m-1),\dots,k}$ is defined recursively as follows:

$$N_{i,m}(y_{t-1}) = \frac{x-\xi_i}{\xi_{i+m-1}-\xi_i} N_{i,m-1}(y_{t-1}) + \frac{\xi_{i+m}-x}{\xi_{i+m}-\xi_{i+1}} N_{i+1,m-1}(y_{t-1})$$

for $\boldsymbol{i = -(m-1),\dots,k}$, and

$$N_{i,1}(x) = \begin{cases} 1, & x \in |\xi_i, \xi_{i+1}| \\ 0, & other \end{cases}$$

If the basis of the B-spline function is categorized according to the order of m=2, it gives the basis of the linear B-spline function, which has the following function:

$$N_{i,2}(y_{t-1}) = \frac{x-\xi_i}{\xi_{i+1}-\xi_i} N_{i,1}(y_{t-1}) + \frac{\xi_{i+2}-x}{\xi_{i+2}-\xi_{i+1}} N_{i+1,1}(y_{t-1})$$

where $\boldsymbol{i = -1,\dots,k}$.

2.2 Parameter Estimation in B-Spline Model

The B-spline model in nonparametric regression of order m with k knot points can be written as:

$$y_t = b_1 N_{1-m,m}(y_{t-1}) + b_2 N_{2-m,m}(y_{t-1}) + \dots + b_{(m+k)} N_{k,m}(y_{t-1}) + \varepsilon_t$$

If the B-spline model is presented in the form of a matrix, we get:

$$\begin{pmatrix} y_1 \\ \dots \\ y_t \end{pmatrix} = \begin{pmatrix} N_{1-m,m}(y_0) & \cdots & N_{k,m}(y_0) \\ \vdots & \ddots & \vdots \\ N_{1-m,m}(y_{t-1}) & \cdots & N_{k,m}(y_{t-1}) \end{pmatrix} \begin{pmatrix} b_1 \\ \dots \\ b_{(m+k)} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_t \end{pmatrix}$$

Which can be written as:

$$\underline{\boldsymbol{y}} = \underline{\boldsymbol{N}}\,\underline{\boldsymbol{b}} + \underline{\boldsymbol{\varepsilon}}$$

Parameter Estimation $\underline{\boldsymbol{b_\lambda}} = \left(b_{\lambda 1}\ \ b_{\lambda 2}\ \ \dots\ \ b_{\lambda(m+k)}\right)^T$ obtained using the least squares spline. Estimator $\underline{\boldsymbol{b_\lambda}}$ obtained by minimizing the Sum of Squares (RSS). RSS minimum if the derivative partial RSS to $\underline{\boldsymbol{b_\lambda}}$ equal to zero.

$$\textbf{RSS} = \underline{\boldsymbol{y}}^T \underline{\boldsymbol{y}} - 2\, \underline{\boldsymbol{b}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{y}} + \underline{\boldsymbol{b}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda}\, \underline{\boldsymbol{b}_\lambda} \tag{2}$$

Next, **Equation (2)** is derivative to $\underline{\boldsymbol{b}_\lambda}$ so that obtained:

$$\frac{\partial \mathrm{RSS}}{\partial \underline{\boldsymbol{b}_\lambda}} = -2\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{y}} + 2\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda}\, \underline{\boldsymbol{b}_\lambda}$$

After RSS is differentiable against $\underline{\boldsymbol{b}_\lambda}$, then the result is equal to zero:

$$-2\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{y}} + 2\, \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda}\, \underline{\boldsymbol{b}_\lambda} = 0 \tag{3}$$

Equation (3) has the following solution:

$$\underline{\boldsymbol{b}_\lambda} = \left(\underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{N}_\lambda}\right)^{-1} \underline{\boldsymbol{N}_\lambda^T}\, \underline{\boldsymbol{y}}$$

with $\underline{\boldsymbol{b}_\lambda} = \left(b_{\lambda 1} \;\; b_{\lambda 2} \;\; \dots \;\; b_{\lambda(m+k)}\right)^T$

Estimator regression curve $\underline{\widehat{\boldsymbol{\mu}}_\lambda} = \left(\hat{\mu}_{\lambda 1} \;\; \hat{\mu}_{\lambda 2} \;\; \dots \;\; \hat{\mu}_{\lambda(m+k)}\right)^T$ is given by:

$$\underline{\widehat{\boldsymbol{\mu}}_\lambda} = \underline{\boldsymbol{N}_\lambda}\, \underline{\widehat{\boldsymbol{b}}_\lambda} = \underline{\boldsymbol{S}_\lambda}\, \underline{\boldsymbol{y}}$$

with matrix $\underline{\boldsymbol{S}_\lambda} = \boldsymbol{N}_\lambda (\boldsymbol{N}_\lambda^T \boldsymbol{N}_\lambda)^{-1} \boldsymbol{N}_\lambda^T$ symmetry and positive definite.

Estimator for regression curve can written by:

$$\hat{\mu}_\lambda(t) = \sum_{j=1}^{m+k} \hat{b}_{\lambda j} N_{j-m,m}(y_{t-1})$$

with $\hat{b}_{\lambda j}$from $\underline{\widehat{\boldsymbol{b}}_\lambda} = \left(\hat{b}_{\lambda 1} \;\; \hat{b}_{\lambda 2} \;\; \dots \;\; \hat{b}_{\lambda(m+k)}\right)^T$. The model estimation for the B-Spline function is:

$$\hat{y} = \textstyle\sum_{j=1}^{m+k} \hat{b}_{kj} N_{j-m,m}(x) \tag{4}$$

Equation (4) can be written as:

$$\hat{y} = \hat{b}_{\lambda 1} N_{1-m,m}(y_{t-1}) + \hat{b}_{\lambda 2} N_{2-m,m}(y_{t-1}) + \cdots + \hat{b}_{\lambda(m+k)} N_{k,m}(y_{t-1})$$

2.3 Optimal B-Spline Model Selection

Before using the least squares spline method for estimation, the location and number of knots for the estimator must first be selected. The location and number of knots can be determined by trial and error by using any value in $\lambda = \{\xi_1, \dots, \xi_k\}$, tested visuall provide a feasible toe estimation result. For linear splines $(m = 2)$, place knots at the point where the slope changes.

Choosing the optimal knot location is necessary to obtain the best B-spline model. Several methods are used to select the optimal, one of which is based on Generalized Cross Validation or GCV for short. [19] showed theoretically that GCV has asymptotic optimal properties that other methods do not have. In addition, the GCV method does not require knowledge of the population variance and the invariance to transformation [20]. This advantage makes GCV popular in nonparametric and semiparametric regression, and researchers in other spline estimators often generalize and modify it to select optimal smoothing parameters [21]. Assuming trace $[\underline{\boldsymbol{S}_\lambda}] < n$, the GCV criteria are defined by:

$$GCV(\lambda) = \frac{n^{-1}RSS(\lambda)}{\left(n^{-1}trace\left[\underline{\boldsymbol{I}} - \underline{\boldsymbol{S}_\lambda}\right]\right)^2}$$

2.4 Truncated Spline Regression Model

Spline regression has high flexibility and can handle data whose behavior changes at certain subintervals so that it can adapt to changes in data patterns with the help of knot points. The following model is a nonparametric truncated multivariable spline regression model in a spline space of order q and r knot points [22].

$$y_i = \beta_0 + \sum_{j=1}^{p}\sum_{u=1}^{q}\beta_{ju}X_{ji}^{u} + \sum_{j=1}^{p}\sum_{k=1}^{r}\beta_{j(q+k)}(x_j - K_{jk})_+^q + \varepsilon_i$$

The truncated function is formed [23]

$$(x_{ji} - K_{jk})_+ = \begin{cases} (x_{ji} - K_{jk})^q, & x_{ji} - K_k \geq 0 \\ 0, & x_{ji} - K_k < 0 \end{cases}$$

2.5 Truncated Spline Regression Model Estimation

The Least Square Method (LCM) estimates nonparametric truncated spline regression models. The following equation is the estimation of the truncated spline regression model in the form of a matrix [24].

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y}$$

2.6 Optimal Truncated Spline Regression Model Selection

The best truncated spline regression model is selected from the models that have the optimum knot point. The GCV method has optimal asymptotic properties; even though it is still optimal in a large sample, the best truncated spline regression model is obtained from the optimum knot point, as seen from the minimum GCV value [8]. [14] wrote a formula to find GCV.

$$GCV\ (\boldsymbol{k}) = \frac{MSE\ (k)}{[n^{-1}trace(\boldsymbol{I} - \boldsymbol{A})]^{\mathbf{2}}}$$

with $MSE(k) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$, $\boldsymbol{A} = \boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-\mathbf{1}}\boldsymbol{x}'$, I identity matrix, n number of observations, Mean Square Error (MSE), and $k = (k_1, k_2, \dots, k_r)$ are knot points, r is a number of knots.

The effect of the predictor variable on the response variable can be known by testing the significance of the parameter. The alternative hypothesis for the parameter significance test is that the predictor variable has a significant effect on the response variable, with a significance level of α; H_0 is rejected if the p-value is less than α.

2.7 Model Goodness Criteria

The criterion of the goodness of the model used in this study is the coefficient of determination, which can measure how much the predictor variable can explain the diversity of the response variables [25]. Here's the formula to get the coefficient of determination (R^2)

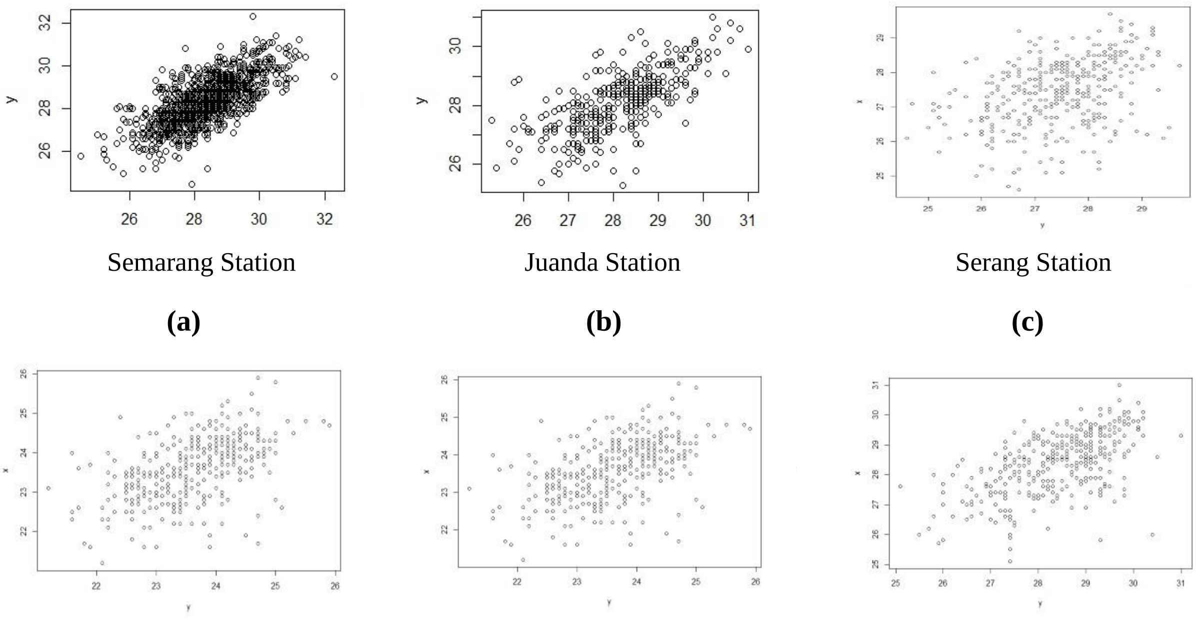
$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

The higher the value of R^2generated by a model, it means that the more independent variables better explain the diversity of the dependent variable [26].

3. RESULTS AND DISCUSSION

3.1 Scatterplot Variable

Temperature data from the six stations on the island of Java are time series data, so it is necessary to establish two variables, namely t-1 for the predictor variable and t for the response variable. The relationship between predictor variables and response variables is shown in Figure 1.



Semarang Station (a) Juanda Station (b) Serang Station (c)

Sleman Station (d) Bandung Station (e) Kemayoran Station (f)

Figure 1. Plot of the Relationship between Predictor Variables and Response Variables, (a) Semarang Station, (b) Juanda Station, (c) Serang Station, (d) Sleman Station, (e) Bandung Station, (f) Kemayoran Station

Figure 1 shows the pattern of annual temperature relationships from 2019 to 2021 in Semarang station and 2021 for the other five stations. The plots show that there are changes in the pattern that occur at several points. The point of change is called a knot, which will then be analyzed further using the B-spline and truncated spline methods.

3.2 Optimal Order and Knot Point Selection

The selection of the optimum knot point is based on the minimum GCV value. The determination of knot points starts from one to three knots with a linear to cubic order. Table 1 compares the minimum GCV values in the B-spline and truncated spline regression models' optimum order and knot points at each station.

Table 1. Orders, Knots, and GCV Value

Stations	B-Spline			Truncated		
	Order	Knot	GCV	Order	Knot	GCV
Semarang	Linear	28.5 29.6 30	0.5942	Linear	26.5 29.8 20.3	0.5958
Juanda	Linear	25.5 25.8 26.8	0.5245	Linear	27 28.3 28.5	0.5251
Serang	Linear	26.8 26.9	0.7413	Linear	27.08	0.7427

Stations	B-Spline			Truncated		
	Order	Knot	GCV	Order	Knot	GCV
Sleman	Linear	24.3 26 27.5	0.4833	Linear	24.3 26 27.5	0.4833
Bandung	Linear	22.2 22.3 22.4	0.4487	Linear	23.1 24.8 24.9	0.4554
Kemayoran	Linear	26.7 26.9 27.0	0.6572	Linear	26.6 27.3 27.9	0.6639

Table 1 explains that the B-Spline model produces a smaller GCV value when compared to the truncated spline. Six stations on the island of Java are optimum at three knots and linear order.

3.3 Parameter Estimation of B-Spline and Truncated Spline

The estimation results of non-parametric B-Spline and truncated spline regression are described in **Table 2**.

Table 2. Parameters Estimation B-Spline Regression and Truncated Spline Regression

	B-spline		Truncated	
Station	Parameter	Estimation	Parameter	Estimation
Semarang	β_1 β_2 β_3 β_4	37.3765 28.8469 29.9061 30.1153	β_0 β_1 β_2 β_3 β_4	16.5589 0.3955 0.3132 0.9064 -1.6962
Juanda	β_1 β_2 β_3 β_4	43.9272 26.9433 27.0549 30.3236	β_0 β_1 β_2 β_3 β_4	27.4996 -0.0135 0.9989 -1.6334 1.5527
Serang	β_1 β_2 β_3	33.7564 26.8982 28.8275	β_0 β_1 β_2	25.3835 0.0620 0.5954
Sleman	β_1 β_2 β_3 β_4	26.7960 25.9026 27.1470 26.5625	β_0 β_1 β_2 β_3 β_4	-58.1194 3.4465 3.3039 0.5395 -1.3502
Bandung	β_1 β_2 β_3 β_4	28.4821 22.2000 22.9039 24.8855	β_0 β_1 β_2 β_3 β_4	20.3468 0.1225 0.6438 -4.2121 4.5917
Kemayoran	β_1 β_2 β_3 β_4	34.4653 25.6826 27.4511 30.0780	β_0 β_1 β_2 β_3 β_4	37.0592 0.3776 1.8315 0.6591

3.4 Coefficient of Determination

The coefficient is used to select the best regression model to model temperature data at each station on the island of Java. In addition, it can provide an overview of the contribution of the t-1 temperature data to the t-th temperature data.

Table 3. R^2 **Value B-Spline Regression and Truncated Spline Regression**

Station	R^2 B-Spline	R^2 Spline Truncated
Semarang	0.9593	0.4943
Juanda	0.9966	0.4981
Serang	0.9748	0.2060
Sleman	0.9962	0.2421
Bandung	0.9923	0.2855
Kemayoran	0.9914	0.3743

Table 3 shows the value of the coefficient of determination from the B-spline regression for each station on the island of Java, which is higher than the truncated spline regression. It can be concluded that the B-spline method is better used to model the temperature of each station on the island of Java. The value of R^2 from the B-spline regression explains that about 95-99% of the temperature on the island of Java is influenced by data or previous days. Furthermore, modeling and interpreting the B-spline model for each station on the island of Java are carried out.

3.5 Best Models

Based on the estimated parameter values in **Table 2**, the B-spline non-parametric regression model for each station in Java is obtained as follows.

Semarang Station

$$\hat{y} = 37.3765B_{0,3} + 28.8469B_{1,3} + 29.9061B_{2,3} + 30.1153B_{3,3} \tag{5}$$

Equation (5) can be interpreted as when the temperature at Semarang Station has a minimum value of 24.5 when substituted in **Equation (5)**, the temperature on the following day is 37 degrees Celsius. When the temperature at Semarang Station has a maximum value of 32.3, it is substituted in **Equation (5)**, then the temperature for the next day is 30 degrees Celsius. **Equation (5)** is described by considering the value of the B-spline basis; the model is obtained as follows.

$$B_{0,3}(x) = \begin{cases} \dfrac{29.5 - x}{5}, & 24.5 < x < 29.5 \\ 0, & \text{for others} \end{cases}$$

$$B_{1,3}(x) = \begin{cases} \dfrac{x - 24.5}{5}, & 24.5 < x < 29.5 \\ \dfrac{29.5 - x}{0.1}, & 29.5 < x < 29.6 \\ 0, & for\ others \end{cases}$$

$$B_{2,3}(x) = \begin{cases} \dfrac{x - 29.5}{0.1}, & 29.5 < x < 29.6 \\ \dfrac{29.6 - x}{0.4}, & 29.6 < x < 30.0 \\ 0, & for\ others \end{cases}$$

$$B_{3,3}(x) = \begin{cases} \dfrac{x - 29.6}{2.7}, & 29.6 < x < 30.0 \\ \dfrac{30 - x}{2.3}, & 30.0 < x < 32.3 \\ 0, & for\ others \end{cases}$$

with an MSE of 0.54.

Juanda Station

$$\hat{y} = 43.9272B_{0,3} + 26.9433B_{1,3} + 27.0549B_{2,3} + 30.3236B_{3,3} \quad (6)$$

Equation (6) can be interpreted as when the temperature at Juanda Station has a minimum value of 25.3 when substituted in **Equation (6)**, the temperature for the next day is 44 degrees. When the maximum value is 31, if it is covered in **Equation (6)**, then the temperature for the next day is 30 degrees. **Equation (6)** is described by considering the value of the B-Spline basis; the model is obtained as follows.

$$B_{0,3}(x) = \begin{cases} \dfrac{25.5 - x}{0.2}, & 25.3 < x < 25.5 \\ 0, & for\ others \end{cases}$$

$$B_{1,3}(x) = \begin{cases} \dfrac{x - 25.3}{0.2}, & 25.3 < x < 25.5 \\ \dfrac{25.5 - x}{0.3}, & 25.5 < x < 25.8 \\ 0, & for\ others \end{cases}$$

$$B_{2,3}(x) = \begin{cases} \dfrac{x - 25.5}{0.3}, & 25.5 < x < 25.8 \\ \dfrac{25.8 - x}{1}, & 25.8 < x < 26.8 \\ 0, & for\ others \end{cases}$$

$$B_{3,3}(x) = \begin{cases} \dfrac{x - 25.8}{5.2}, & 25.8 < x < 26.8 \\ \dfrac{26.8 - x}{4.2}, & 26.8 < x < 31.0 \\ 0, & for\ others \end{cases}$$

with an MSE of 0.32.

Serang Station

$$\hat{y} = 33.7564B_{0,2} + 26.8982B_{1,2} + 28.8275B_{2,2} \quad (7)$$

Equation (7) can be interpreted when the temperature at Serang Station has a minimum value of 24.6. If it is substituted in **Equation (7)**, then the temperature on the following day is 33.7 degrees Celsius. When the maximum is 29.7 when substituted in **Equation (7)**, the next day's temperature is 28.8 degrees Celsius. **Equation (7)** is described by considering the value of the B-spline basis; the model is obtained as follows.

$$B_{0,2}(x) = \begin{cases} \dfrac{26.8 - x}{2.2}, & 24.6 < x < 26.8 \\ 0, & for\ others \end{cases}$$

$$B_{1,2}(x) = \begin{cases} \dfrac{x - 24.6}{2.2}, & 24.6 < x < 26.8 \\ \dfrac{26.8 - x}{0.1}, & 26.8 < x < 26.9 \\ 0, & for\ others \end{cases}$$

$$B_{2,2}(x) = \begin{cases} \dfrac{x-26.8}{2.7}, & 26.8 < x < 26.9 \\ \dfrac{26.9-x}{2.8}, & 26.9 < x < 29.7 \\ 0, & for\ others \end{cases}$$

with an MSE of 1.37.

Sleman Station

$$\hat{y} = 26.796B_{0,3} + 25.9026B_{1,3} + 27.1470B_{2,3} + 26.5625B_{3,3} \qquad (8)$$

Equation (8) can be interpreted: when the temperature at Sleman Station has a minimum value of 23.9 when substituted in **Equation (8)**, the temperature on the following day is 26.8 degrees Celsius. When the maximum temperature is 28.3 when substituted in **Equation (8)**, the next day's temperature is 26.6 degrees Celsius. **Equation (8)** is described by considering the value of the B-spline basis; the model is obtained as follows.

$$B_{0,3}(x) = \begin{cases} \dfrac{24.3-x}{0.4}, & 23.9 < x < 24.3 \\ 0, & for\ others \end{cases}$$

$$B_{1,3}(x) = \begin{cases} \dfrac{x-23.9}{0.4}, & 23.9 < x < 24.3 \\ \dfrac{24.3-x}{1.7}, & 24.3 < x < 26.0 \\ 0, & for\ others \end{cases}$$

$$B_{2,3}(x) = \begin{cases} \dfrac{x-24.3}{1.7}, & 24.3 < x < 26.0 \\ \dfrac{26.0-x}{1.5}, & 26.0 < x < 27.5 \\ 0, & for\ others \end{cases}$$

$$B_{3,3}(x) = \begin{cases} \dfrac{x-26.0}{2.3}, & 24.3 < x < 26 \\ \dfrac{27.5-x}{0.8}, & 27.5 < x < 28.3 \\ 0, & for\ others \end{cases}$$

with an MSE of 0.58.

Bandung Station

$$\hat{y} = 28{,}4821B_{0,3} + 22{,}2B_{1,3} + 22{,}9039B_{2,3} + 24{,}8855B_{3,3} \qquad (9)$$

Equation (9) can be interpreted: when the temperature at Bandung Station has a minimum value of 21.2 when substituted in **Equation (9)**, the temperature on the following day is 28 degrees Celsius. When the maximum temperature is 25.9 when substituted in **Equation (9)**, the next day's temperature is 24 degrees Celsius. **Equation (9)** is described by considering the value of the B-spline basis; the model is obtained as follows.

$$B_{0,3}(x) = \begin{cases} \dfrac{22.2 - x}{1.0}, & 21.2 < x < 22.2 \\ 0, & for\ others \end{cases}$$

$$B_{1,3}(x) = \begin{cases} \dfrac{x - 21.2}{1}, & 21.2 < x < 22.2 \\ \dfrac{22.2 - x}{0,1}, & 22.2 < x < 22.3 \\ 0, & for\ others \end{cases}$$

$$B_{2,3}(x) = \begin{cases} \dfrac{x - 22.2}{0.1}, & 22.2 < x < 22.3 \\ \dfrac{22.3 - x}{0.1}, & 22.3 < x < 22.4 \\ 0, & for\ others \end{cases}$$

$$B_{3,3}(x) = \begin{cases} \dfrac{x - 22.3}{3.6}, & 22.3 < x < 22.4 \\ \dfrac{22.4 - x}{3.5}, & 22.4 < x < 25.9 \\ 0, & for\ others \end{cases}$$

with an MSE of 0.57.

Kemayoran Station

$$\hat{y} = 34.4653B_{0,3} + 25.6826B_{1,3} + 27.4511B_{2,3} + 30.0780B_{3,3} \qquad (10)$$

Equation (10) can be interpreted as when the temperature at Kemayoran Station has a minimum value of 25.1. If it is substituted in **Equation (10)**, then the temperature for the next day is 34 degrees Celsius. When the maximum temperature is 31 when substituted in **Equation (10)**, the temperature for the next day is 30 degrees Celsius. **Equation (10)** is described by considering the value of the B-spline basis; the model is obtained as follows.

$$B_{0,3}(x) = \begin{cases} \dfrac{26.7 - x}{1.6}, & 25.1 < x < 26.7 \\ 0, & for\ others \end{cases}$$

$$B_{1,3}(x) = \begin{cases} \dfrac{x - 25.1}{1.6}, & 25.1 < x < 26.7 \\ \dfrac{26.7 - x}{0.2}, & 26.7 < x < 26.9 \\ 0, & for\ others \end{cases}$$

$$B_{2,3}(x) = \begin{cases} \dfrac{x - 26.7}{0.2}, & 26.7 < x < 26.9 \\ \dfrac{26.9 - x}{0.1}, & 26.9 < x < 27.0 \\ 0, & for\ others \end{cases}$$

$$B_{3,3}(x) = \begin{cases} \dfrac{x - 26.9}{0.1}, & 26.9 < x < 27.0 \\ \dfrac{27.0 - x}{4.0}, & 27.0 < x < 31.0 \\ 0, & for\ others \end{cases}$$

with an MSE of 0.35.

3.6 Temperature Forecasting

The best B-spline regression model based on section 3.5 is used to forecast the air temperature at each station on Java Island from January 1, 2022, to January 10, 2022. The forecasting results are described in **Table 4**.

Table 4. Comparison of Forecasting Value with Actual Temperature Value in Java

Station	Predicted Date	Forecasted Value	Actual Value
Semarang	1 January 2022	27.2686	27.8
	2 January 2022	27.5949	27.6
	3 January 2022	27.8221	28.1
	4 January 2022	27.9804	28.2
	5 January 2022	28.0906	27.8
	6 January 2022	28.1673	27.1
	7 January 2022	28.2208	26.9
	8 January 2022	28.2580	27.7
	9 January 2022	28.2839	27.5
	10 January 2022	28.3019	27.3
Juanda	1 January 2022	27.3469	27.4
	2 January 2022	27.4630	26.9
	3 January 2022	27.5548	28.6
	4 January 2022	27.6273	27.9
	5 January 2022	27.6846	28.4
	6 January 2022	27.7299	28.1
	7 January 2022	27.7657	28.0
	8 January 2022	27.7940	28.6
	9 January 2022	27.8164	27.2
	10 January 2022	27.8341	28.3
Serang	1 January 2022	27.6561	27.5
	2 January 2022	27.4192	26.8
	3 January 2022	27.2559	27.9
	4 January 2022	27.1434	27.4
	5 January 2022	27.0659	28.0
	6 January 2022	27.0125	27.5
	7 January 2022	26.9757	29.0
	8 January 2022	26.9504	29.3
	9 January 2022	26.9329	28.1

Station	Predicted Date	Forecasted Value	Actual Value
	10 January 2022	26.9209	27.8
Sleman	1 January 2022	26.9305	27.3
	2 January 2022	26.6797	26.8
	3 January 2022	26.5095	26.6
	4 January 2022	26.3939	26.6
	5 January 2022	26.3155	27.5
	6 January 2022	26.2622	27.4
	7 January 2022	26.2261	27.4
	8 January 2022	26.2016	27.4
	9 January 2022	26.1849	26.2
	10 January 2022	26.1736	25.8
Bandung	1 January 2022	24.0362	24.8
	2 January 2022	23.8303	24.1
	3 January 2022	23.7137	24.4
	4 January 2022	23.6477	24.4
	5 January 2022	23.6103	24.5
	6 January 2022	23.5891	24.6
	7 January 2022	23.5772	23.6
	8 January 2022	23.5704	24.0
	9 January 2022	23.5665	25.0
	10 January 2022	23.5644	24.2
Kemayoran	1 January 2022	28.9615	28.9
	2 January 2022	28.7393	28.2
	3 January 2022	28.5933	29.2
	4 January 2022	28.4975	28.9
	5 January 2022	28.4345	29.4
	6 January 2022	28.3932	27.9
	7 January 2022	28.3660	29.3
	8 January 2022	28.3482	28.6
	9 January 2022	28.3365	29.1
	10 January 2022	28.3288	28.3

The visualization between forecasted and actual for the six stations on Java Island can be seen in **Figure 2**.

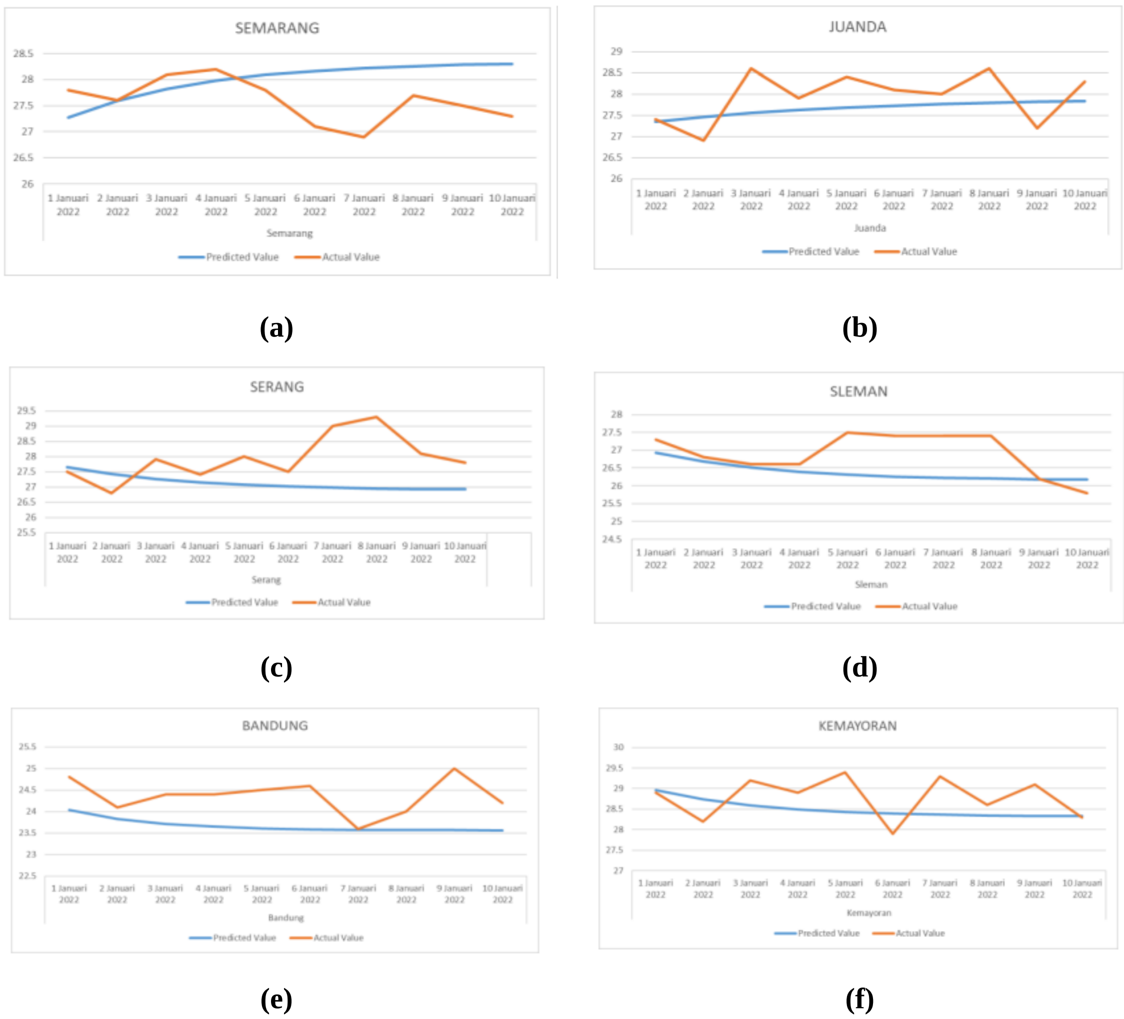


Figure 2. Forecasted and Actual Temperatures, (a) Semarang Station, (b) Juanda Station, (c) Serang Station, (d) Sleman Station, (e) Bandung Station, (f) Kemayoran Station

The forecast results are quite close to the actual temperature for the first 10 days in January. From the six stations above, it can be seen that the temperature forecasting results that are close to the actual value have the smallest MSE value and are obtained from the forecasting results at Juanda and Kemayoran stations because the forecasting points are not far above or below the actual value. This can be used as a reference for forecasting around the area and can be used as a policy consideration in fields that use temperature data.

4. CONCLUSIONS

Based on data analysis regarding the pattern of the relationship between the t-*th* temperature and the (t-1) temperature, it can be concluded that:

1) The analysis results of the B-spline linear model have a fairly small MSE value from temperature data from Semarang, Juanda, Serang, Sleman, Bandung, and Kemayoran stations so as to provide data predictions that are quite close to the next ten days.
2) The results of the analysis of the B-Spline model provide a greater coefficient of determination than the truncated spline model, so that the B-Spline model is better used to predict temperature data at the six stations in Java.

The analysis that has been done is still limited, considering the t-th and t-1 temperature data so that the model can be developed for the previous two periods. In addition, P Spline modeling can also be applied to compare with the current results by including a penalty factor in building the prediction model.

ACKNOWLEDGMENT

This research has produced beneficial analytical results, so the authors would like to thank Sebelas Maret University for funding this research. We also do not forget to thank BMKG for publishing published data so that it can be analyzed and obtain the desired model.

REFERENCES

[1] M.K. Alomar, F. Khaleel, M. M. Aljumaily, A. Masood, S. F. M. Razali, M. A. AlSaadi, N. Al-Ansari, and M. M. Hameed, "Data-driven Models for Atmospheric Air Temperature Forecasting at a Continental Climate Region," *JPlos ONE*, vol. 17, no. 11, 2022, [Online]. Available: https://doi.org/10.1371/journal.pone.0277079

[2] R.J. Vidmar, "On the Use of Atmospheric Pressure Plasmas as Electromagnetic Reflectors and Absorbers," *IEEE Trans Plasma Sci*, vol. 18, no. 4, pp. 733–741, 1990.

[3] O. A. Dombayci and M. Golcu, "Daily Means Ambient Temperature Prediction Using Artificial Neural Network Method: A Case Study of Turkey," *Renew Energy*, vol. 34, no. 4, pp. 1158–1161, 2009.

[4] X. Li, Z. Li, W. Huang, and P. Zhou, "Performance of Statistical and Machine Learning Ensembles for Daily Temperature Downscaling," *Theor Appl Clim.*, vol. 140, no. 1, pp. 571–588, 2020.

[5] I. Park, H. S. Kim, J. Lee, J. H. Kim, C. H. Song, and H. K. Kim, "Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network," *Atmosphere*, vol. 10, no. 11, 2019.

[6] B.A Smith, G. Hoogenboom, and R.W. McClendon, "Artificial Neural Networks for Automated Year-round Temperature Prediction," *Comput Electron Agric*, vol. 68, no. 1, pp. 52–61, 2009.

[7] A. Arif, "Mitigasi Peningkatan Curah Hujan," *Kompas*, May 31, 2022. [Online]. Available: https://www.kompas.id/baca/ilmu-pengetahuan-teknologi/2022/05/30/mitigasi-peningkatan-curah-hujan

[8] R.K. Kaufmann, H. Kauppi, M.L. Mann, and J.H. Stock, "Reconciling Anthropogenic Climate Change with Observed Temperature 1998–2008," *Proc Natl Acad Sci*, vol. 108, no. 29, pp. 11790–11793, 2011.

[9] D.A Stone and M.R. Allen, "Attribution of Global Surface Warming Without Dynamical Models," *Geophys Ress Lett*, vol. 32, no. 18, 2005, [Online]. Available: https://doi.org/10.1029/2005GL023682

[10] D.H. Douglass, E.G Blacman, and R.S Knox, "Temperature Response of Earth to the Annual Solar Irradiance Cycle," *Phys Lett A*, vol. 323, no. 3, pp. 315–322, 2004.

[11] M. Afzali, A. Afzali, and G. Zahedi, "The Potential of Artificial Neural Network Technique in Daily and Monthly Ambient Air Temperature Prediction," *Int J Env. Sci Dev*, vol. 3, no. 1, pp. 33–38, 2012.

[12] W. Hardle, *Applied Nonparametric Regression*. New York: Cambridge University Press, 1990.

[13] I.N. Budiantara, *Spline Dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang, Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu Matematika Statistika Dan Probabilitas, Pada Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember.* Surabaya: ITS Press, 2009.

[14] R.L. Eubank, *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker Inc., 1999.

[15] T.Lyche and K. Morken, *Spline Methods Draft*. 2004.

[16] M. Kaseside and S.B. Loklomin, "Analisis Angka Kematian Bayi di Kabupaten Halmahera Utara dengan Metode Regresi Nonparametrik Spline Truncated," *Variance*, vol. 3, no. 1, pp. 1–5, 2021.

[17] R.E.Caraka and A.R. Devi, "Application Of Non Parametric Basis Spline (B-SPLINE) In Temperature Forecasting," *J. Stat.*, vol. 4, no. 2, pp. 68–74, 2016.

[18] G. Farin, *Curves and Surfaces for CAGD A Practical Guide*, Fifth. United States of America: Academic Press, 2002.

[19] G. Wahba, "Spline Models for Observational Data," *CBMS-NSF Reg. Conf. Ser. Appl. Math.*, vol. 59, 1990.

[20] D. Fitriana, "Pengujian Hipotesis Parsial Pada Regresi Semiparametrik Truncated spline(Aplikasi Data Angka Harapan Hidup di Indonesia," Institut Teknologi Sepuluh November, Surabaya, 2017.

[21] P.S. Carmack, J.S. Spence, and W.R. Schucany, "Generalised correlated cross-validation," *J. Nonparametric Stat.*, vol. 24, no. 2, pp. 269–282, 2012.

[22] A.A.R. Fernandes, I N. Budiantara, B.W. Otok, and Suhartono, "Spline Estimator for Bi-Response and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data," *J. Math. Stat.*, vol. 11, no. 2, pp. 61–69, 2015.

[23] A.P. Sugiantari and I N. Budiantara, "Analisis Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline," *J. Sains Dan Seni ITS*, vol. 2, no. 1, pp. 37–41, 2013.

[24] A. Tripena and I N. Budiantara, "Fourier Estimator in Nonparametric Regression," in *International Conference On Natural and Applied Natural Scienes. Ahmad Dahlan Universiti*, 2006.

[25] D. N. Gujarati and D. C. Porter, *Basic Econometrics 4th Edition*. Graw Hill, New York, 2004.

[26] N. R. Draper and H. Smith, *Analisis Regresi Terapan Ekonomi*, Edisi Kedu. PT Gramedia Pustaka Utama, Jakarta, 1992.