BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2597–2608
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss4pp2597-2608

TRUNCATED SPLINE SEMIPARAMETRIC REGRESSION TO HANDLE MIXED PATTERN DATA IN MODELING THE RICE PRODUCTION IN EAST JAVA PROVINCE

Sri Sulistijowati Handajani[1*], Hasih Pratiwi[2], Respatiwulan[3], Yuliana Susanti[4], Muhammad Bayu Nirwana[5], Lintang Pramesti Nareswari[6]

[1,2,3,4,5,6]*Study Program of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret*
Jln. Ir Sutami No. 36A Surakarta, 57126, Indonesia

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

Article History:

Received: 28th January 2025
Revised: 4th March 2025
Accepted: 3rd June 2025
Available online: 1st September 2025

Keywords:

GCV;
Knot Point;
Mixed Pattern Data;
Rice Production;
Truncated Spline Semiparametric Regression.

ABSTRACT

Climate change can affect rice production through changes in temperature, precipitation patterns, extreme weather events, and atmospheric carbon dioxide levels. A statistical model can be used to understand the correlation between rice production and factors that affect it. The existence of some patterns that are formed from independent variables and others that do not show data patterns due to volatility in weather element data makes semiparametric regression modeling more appropriate. In forming a parametric model, the data pattern needs to be regular to make the model more precise. Irregular data patterns are more appropriately modeled with nonparametric regression models. The existence of several patterns formed from independent variables to their dependent variables, and several others, does not show a particular pattern due to the volatility in climate data, making truncated spline semiparametric regression modeling more appropriate to use. This research aims to model rice production in several regions in East Java Province in 2022 using a semiparametric regression model. The data used were from the Meteorology, Climatology, and Geophysics Agency and the Central Statistics Agency for East Java Province in 2022. The response variable is the rice production (tons) in 2022 in Tuban, Gresik, Nganjuk, Malang, Banyuwangi, and Pasuruan Regency (Y). The predictor variables are paddy harvested area (hectares), average temperature (℃), humidity (percent), and rainfall (mm). The semi-parametric spline truncated regression model is obtained by combining the parametric and non-parametric models based on truncated splines. The analysis showed a spline truncated semiparametric regression model with a combination of knot points (3,3,1) with a minimum GCV value of 12,642,272. The variables significantly affecting rice production were rice harvest area, temperature, air humidity, and rainfall, with an R^2adjusted value of 98.522%.

How to cite this article:

S. S. Handajani, H. Pratiwi, Respatiwulan, Y. Susanti, M. B. Nirwana, and L. P. Nareswari, "TRUNCATED SPLINE SEMIPARAMETRIC REGRESSION TO HANDLE MIXED PATTERN DATA IN MODELING THE RICE PRODUCTION IN EAST JAVA PROVINCE," *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2597-2608, December, 2025.

Copyright © 2025 Author(s)
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

Rice production in Indonesia is essential in the agricultural sector and the country's economy. Rice is one of Indonesia's leading food crops and a necessary source of carbohydrates for the Indonesian population. Indonesia has geographical conditions that support rice farming. The tropical climate, sufficient rainfall, and ample agricultural land make Indonesia one of the largest rice producers in the world. However, Indonesia no longer has sufficient rice production to meet domestic needs because paddy production in Indonesia fluctuates over time depending on various factors that influence it [1].

Indonesia is currently working towards achieving self-sufficiency in rice production, which is adequate for domestic needs. Ensuring that rice is available for the projected population increase will require increasing rice production [2], [3]. However, there are still challenges in increasing the productivity and efficiency of rice production, such as better agricultural land management and determining the right rice planting season by looking at the area's rainfall, temperature, and humidity.

In 2021, East Java was one of the largest rice producers in Indonesia [4]. In 2020, rice production in East Java reached around 14.9 million tons of dry grain rice. This amount accounts for about 19% of the total rice production throughout Indonesia. East Java has a significant agricultural land area and climatic and rainfall conditions that support the growth of rice plants [5]. Rice production in East Java is influenced by several factors, such as the agricultural technology used, rice varieties planted, irrigation methods, and government policies in the farm sector. East Java government, too, has attempted to increase the productivity and efficiency of rice production through programs that support farmers, such as providing superior seeds, agricultural training, and developing rural infrastructure. All of this has been attempted, but current conditions, such as El Nino, namely the lack of rain, have occurred in parts of the island of Java, which has been the primary national rice producer [6]. Hence, rainfall must be considered regarding its effect on rice production. This research aims to model rice production because it needs to be done by looking at climatic conditions such as rainfall, humidity, temperature, and land area to determine how much rice can be produced.

Regression modeling can explain the functional relationship between one or more variables [5], [7], [8]. Regression modeling begins by looking at the pattern of data between the variables of rice production, with each variable expected to influence the others, to determine a more appropriate model. Parametric, semi-parametric, and non-parametric regression modeling can be built using scatter diagrams based on data patterns and trends [9].

Besides pattern deep data trends from visible curves, deep regression parametric is also expected to have information on previous data patterns to obtain good modeling [10]. Liu and Li [11] have investigated a mixture model, i.e., a semiparametric mixture cure model, another model form of a mixture of the general linear model for the cure probability, and a general class of transformation models for the failure times of non-cured subjects.

In recent developments, there have been symptoms that show changes in behavior naturally that lead to no pattern like usual (as if abnormal). Several years ago, farmers could still estimate when the start and end seasons of drought and rain so that they could prepare themselves for when to start planting paddy and when to harvest it; however, now it is difficult to predict because it needs a methodical approach that is used to give modeling and more predictions. With the initiation of a semiparametric regression model built from the current data relationship pattern, it is hoped that it can predict rice production more accurately following current conditions, and help farmers in increasing rice production.

Nonparametric regression and semiparametric modeling can be used to get a good model with minor errors [12]. This approach has been used a lot, including histogram [13], kernel [14], [15], spline [15], [16], [17], and others. A nonparametric path analysis model has been developed [8] to identify changes in data behavior patterns that occur with corresponding knot points and polynomial order. Modeling of rice production in East Java, as a fairly large rice-producing center, has not been carried out using a semi-parametric approach, which provides better modeling.

2. RESEARCH METHODS

2.1 Types and Sources of Research Data

The research relies on secondary data acquired from the Meteorology, Climatology, and Geophysics Agency and the Central Statistics Agency for East Java Province in 2022. Data collection was only carried out in areas that produce quite large rice production in East Java, and also due to the completeness of climate data at observation stations in the area, it was incomplete in other regions. The response variable is the amount of rice production (tons) in 2022 in Tuban Regency, Gresik Regency, Nganjuk Regency, Malang Regency, Banyuwangi Regency, and Pasuruan Regency (Y). The predictor variables are paddy harvested area (hectares) (X_1), average temperature (°C) (X_2), humidity (percent) (X_3), and rainfall (mm) (X_4). The unit of observation used is 72 data points.

2.2 Research Steps

The research steps begin with collecting complete data in several areas in East Java and making scatterplots of each predictor variable against its response variable. By looking at the scatterplots, appropriate modeling can be done. Seeing the random patterns in several scatterplots, a semiparametric regression model was chosen using the truncated spline method. Building a truncated spline model requires the number of knot points that are considered optimal and the order of the function based on the smallest GCV. From the optimal knot points and the specified order, the model parameters are estimated, and the significance of the parameters obtained is tested. With the model obtained, the assumption of normality and homogeneity of variance is tested to determine the accuracy of the parameter significance test that has been carried out. The model's goodness is seen from the R^2_{adj} value and its RMSE value. The analysis steps above use the help of R software and finally interpret the model so that it can be helpful.

2.3 Analysis Method

Regression models are statistical techniques used to describe how predictor and response variables function together. Semiparametric, parametric, and nonparametric regression models are used frequently. Parametric regression models are used if the relationship between predictor variables and response variables is known. Furthermore, parametric regression needs to fulfill assumptions [18]. In contrast to the parametric regression approach, nonparametric regression does not require assumptions to be met [19]. The nonparametric approach is used if it is unknown where the predictor and response variables relate to each other [20]. The curve in nonparametric regression is assumed to be contained in a specific function depending on the data used [21], [22]. Semiparametric regression is a combination of parametric and nonparametric regression [23].

2.3.1 Truncated Spline Semiparametric Regression

Regression is divided into parametric, nonparametric, and semiparametric regression methods. The semiparametric regression approach is used if one of the regression curves is known while the other is unknown [14]. Truncated spline semiparametric regression can be used to estimate the regression curve. Truncated splines can overcome data patterns that show data fluctuations assisted by knot points [12], [15]. If there is paired data y_i, x_i, z_i, in semiparametric regression, it is assumed that the relationship between y_i, x_i, and z_i is in Equation (1).

$$y_i = f(x_i) + g(z_i) + \varepsilon_i, \qquad i = 1,2,\dots,n \tag{1}$$

$f(x_i)$ is the i^{th} parametric regression function, and $g(z_i)$ is the i^{th} nonparametric regression function. This function is described as follows:

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$
$$g(z_i) = \gamma_1 z_i + \gamma_2 z_i^2 + \cdots + \gamma_m z_i^m + \gamma_{1+m}(z_i - K_1)_+^m + \cdots + \gamma_{r+m}(z_i - K_r)_+^m$$
$$g(z_i) = \sum_{j=1}^{m} \gamma_j z_i^j + \sum_{k=1}^{r} \gamma_{k+m}(z_i - K_k)_+^m$$

Where:

$$(z_i - K_k)_+^m = \begin{cases} 0 & , \quad z_i < K_k \\ (z_i - K_k)^m & , \quad z_i \geq K_k \end{cases}$$

The regression model can be written in the following matrix form.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{y} = (\boldsymbol{X\beta} + \boldsymbol{Z\gamma}) + \boldsymbol{\varepsilon}$$
$$\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{k})\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the response vector, $\boldsymbol{X}$ is the parametric predictor variable matrix, $\boldsymbol{Z}$ is the nonparametric predictor variable matrix, $\boldsymbol{\beta}$ is the parametric variable parameter vector, $\boldsymbol{\gamma}$ is the nonparametric variable parameter vector, $M(k)$ is the composite matrix of the $\boldsymbol{X}$ matrix and $\boldsymbol{Z}$ matrix, $\boldsymbol{\delta}$ is the joint vector of the $\boldsymbol{\beta}$ vector and vector $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ is the error vector.

2.3.2 Parameter Estimation

A truncated spline regression model was obtained using the least squares method. We will obtain the following parameter estimates by minimizing the sum of the squared errors.

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{M}(\boldsymbol{k})\boldsymbol{\delta})^T(\boldsymbol{Y} - \boldsymbol{M}(\boldsymbol{k})\boldsymbol{\delta})$$

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{2\delta}^T\boldsymbol{M}^T(\boldsymbol{k}) + \boldsymbol{\delta}^T\boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{M}(\boldsymbol{k})\boldsymbol{\delta}$$

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \boldsymbol{Q}(\boldsymbol{\delta})$$

To obtain the parameter estimator $\boldsymbol{\delta}$, a partial derivative of $\boldsymbol{Q}(\boldsymbol{\delta})$ is carried out with respect to $\boldsymbol{\delta}$ and equated to zero, then **Equation (2)** is obtained:

$$-\boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{Y} + \boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{M}(\boldsymbol{k})\boldsymbol{\delta} = \boldsymbol{0}$$

$$\hat{\boldsymbol{\delta}} = \left(\boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{M}(\boldsymbol{k})\right)^{-1}\boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{y} \tag{2}$$

2.3.3 Optimal Knot Point Selection Method

The knot point is a convergence point that shows the behavior of the curve at certain sub-intervals [9]. One of the methods used in selecting knots is the Generalized Cross Validation (GCV) value. The optimal knot point is chosen based on the smallest GCV in **Equation (3)**.

$$GCV(\boldsymbol{k}) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\left[n^{-1}trace\left(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{k})\right)\right]^2} \tag{3}$$

$$\boldsymbol{A}(\boldsymbol{k}) = \boldsymbol{M}(\boldsymbol{k})\left(\boldsymbol{M}^T(\boldsymbol{k})\boldsymbol{M}(\boldsymbol{k})\right)^{-1}\boldsymbol{M}^T(\boldsymbol{k})$$

2.3.4 Parameter Significance Test

1. Simultaneous Test

Simultaneous testing was conducted to determine whether the predictor variable's parameters were jointly significant to the response variable [7]. The hypothesis used is as follows.

$H_0 : \beta_h = \gamma_{jl} = \gamma_{(j+k)l} = 0$
H_1: There is at least one $\beta_h \neq 0$ or $\gamma_{jl} \neq 0$ or $\gamma_{(j+k)l} \neq 0$

The test statistics used are

$$F = \frac{MS_{Regression}}{MS_{Error}}$$

The decision H_0 is rejected when $F > F_{\alpha;(p+q(r+m);n-(p+q(r+m))-1)}$ or $p_{value} < \alpha$, so it can be concluded that the predictor variables together have a significant effect on the response variable

2. Partial Test

Partial testing was conducted to determine whether the predictor variable parameters were jointly significant to the response variable [8]. The hypothesis used is as follows.

H_0 : $\beta_h = 0$ vs. H_1 : $\beta_h \neq 0; h = 1, \dots, p$
H_0 : $\gamma_{jl} = 0$ vs. H_1 : $\gamma_{jl} \neq 0; j = 1, 2, \dots, m \,; l = 1, 2\,, \dots, q$

The test statistics used are

$$t = \frac{\hat{\gamma}_{jl}}{se(\hat{\gamma}_{jl})};\ t = \frac{\hat{\beta}_h}{se(\hat{\beta}_h)}$$

The decision H_0 is rejected if $|t| > t_{\frac{\alpha}{2};df_{err}}$ or $p_{value} < \alpha$ so it can be concluded that the predictor variable significantly affects the response variable.

2.3.5 Model Goodness Criteria

1. The determination coefficient is a statistical measure that assesses the proportion of the variance in the response variable explained by the predictor variables in a regression model **[12]**. The regression model is better if it has a high coefficient of determination.

$$R^2_{adj} = 1 - \left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \times \frac{n-1}{n - p(m+k)}\right)$$

2. Root Mean Square Error (RMSE) in regression analysis to evaluate the accuracy of a predictive model. It measures the average magnitude of the errors between predicted and observed values.

The formula for RMSE is

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where:
n is the number of observations
y_i represents the observed values
$\hat{y}_i$ represents the predicted values.

2.3.6 Error Assumptions Test

1. Normality

The normality test is carried out to test whether the residuals are normally distributed. The normality assumption test uses the Kolmogorov-Smirnov normal distribution test **[12]**. The hypothesis used is as follows.

$H_0: F_n(\varepsilon) = F_0(\varepsilon)$

$H_1: F_n(\varepsilon) \neq F_0(\varepsilon)$

The test statistics used are

$$D = Sup_{\varepsilon}|F_n(\varepsilon) = F_0(\varepsilon)|$$

The decision H_0 is rejected if $|D| > D_{(1-\alpha)}$ or $p_{value} < \alpha$, so it can be concluded that the residuals are not normally distributed.

b. Heteroscedasticity

The heteroscedasticity test was carried out to determine the homogeneity of the variance of the regression model's deviation. This assumption is fulfilled if the variance between deviations is homogeneous. One method for testing the hypothesis of heteroscedasticity is the Glejser test **[15]**. The hypothesis used is as follows.

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$
H_1: There is at least one $\sigma_i^2 \neq \sigma^2; i = 1,2, \dots, n$

The test statistics used are

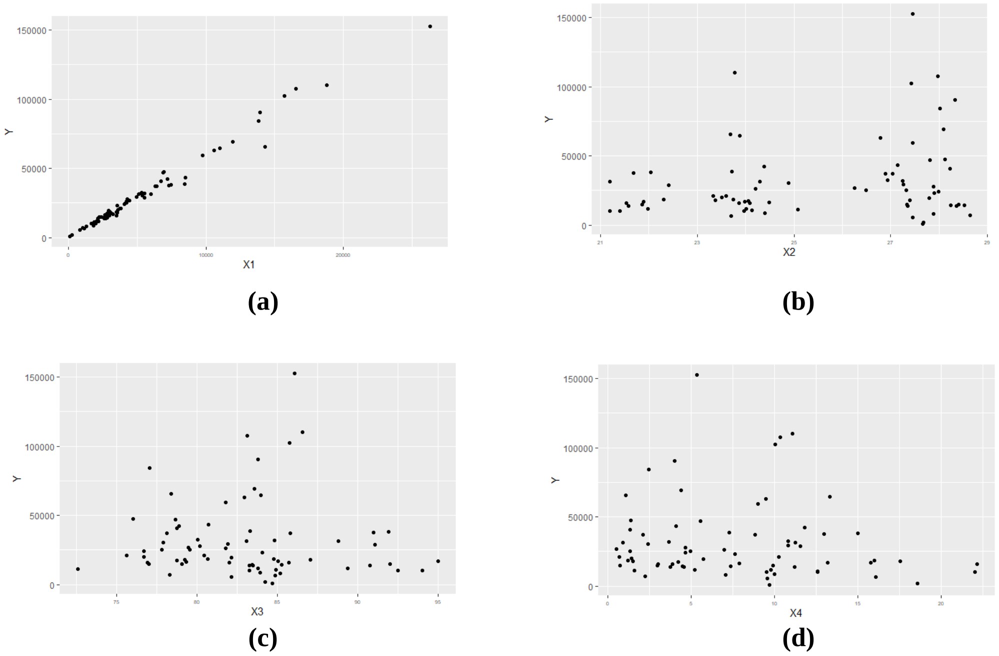
$$F = \frac{\dfrac{\sum_{i=1}^{n}(|\hat{\varepsilon}_i| - |\bar{\varepsilon}|^2)}{p+r}}{\dfrac{\sum_{i=1}^{n}(|\varepsilon_i| - |\hat{\varepsilon}_i|^2)}{n-(p+r)-1}}$$

The decision H_0 is rejected if $F > F_{\alpha;(v,n-v-1)}$ or $p-$value $< \alpha$, so it can be concluded that there is no indication of any homoscedasticity.

3. RESULTS AND DISCUSSION

3.1 Scatter Plots

The relationship pattern of each predictor variable to the response variable can be identified using a scatter plot. The scatterplot between the response variables and each variable $X_1, X_2, X_3,$ and X_4 can be seen in **Figure 1**.



(a)

(b)

(c)

(d)

Figure 1. (a) Scatter Plot between the Predictor Variable $X_1$and Y, (b) Scatter Plot between the Predictor Variables X_2 and Y, (c) Scatter Plot between the Predictor Variables X_3 and Y, (d) Scatter Plot between the Predictor Variables X_4 and Y

Figure 1 (a) shows that the scatter plot between the predictor variable X_1 and the response variable has a pattern. It can be seen that the pattern formed is linear, thus indicating a linear relationship between Y and the parameters of the linear regression model can be estimated. In contrast, the scatterplot between the predictor variables X_2 vs. Y, X_3 vs. Y, and X_4 vs. Y in **Figure 1**(b), **Figure 1**(c), and **Figure 1**(d), shows that there is no pattern, so they can be modeled in nonparametric regression. Because there are linear patterns and random patterns in the relationship between the Y variable and each variable X, the estimation of the regression model uses a truncated spline semiparametric approach.

3.2 Optimal Knot Point Selection

The selection of the optimum knot point needs to determine the order, the knot point, and the minimum GCV. The order used in this study is the linear order with a combination of several knot points. The results of selecting the knot point are shown in **Table 1**.

Table 1. GCV Value and Number of Knot Points

Point Knots	GCV
1	13143513
2	13038546
3	13413243
Combination (3,3,1)	12642272

Table 1 shows that the minimum GCV value is 12642272 using a combination of knot points. The optimal knot points are located at $X_2 = 22.4119;\ 22.56381;\ 22.71571,\ X_3 = 76.25714;\ 76.71429;\ 77.17143$, and $X_4 = 10.23145$.

3.3 Truncated Spline Semiparametric Model

The best model is obtained by selecting the optimal knot point with the smallest GCV value. The truncated spline model formed using a linear order and a combination of knot points is

$$\begin{gathered}\hat{y} = \beta_0 + \beta_1 x_{i1} + \gamma_{11} z_{i1} + \gamma_{21}(z_{i1} - k_{11})_+ + \gamma_{31}(z_{i1} - k_{21})_+ + \gamma_{41}(z_{i1} - k_{31})_+ + \gamma_{12} z_{12} \\ +\gamma_{22}(z_{i2} - k_{12})_+ + \gamma_{32}(z_{i2} - k_{22})_+ + \gamma_{42}(z_{i2} - k_{32})_+ + \gamma_{13} z_{i3} + \gamma_{23}(z_{i3} - k_{13})_+ \\ +\gamma_{33}(z_{i3} - k_{23})_+ + \gamma_{43}(z_{i3} - k_{33})_+\end{gathered}$$

From the results of parameter estimation and optimum knot points obtained, the model formed is

$$\begin{gathered}\hat{y} = 5.8748 x_{i1} - 2144.1090 z_{i1} + 317.1016(z_{i1} - k_{11})_+ + 1094.0594(z_{i1} - k_{21})_+ \\ +1870.9661(z_{i1} - k_{31})_+ + 609.3392 z_{i2} - 5288.0700(z_{i2} - k_{12})_+ - 981.7648(z_{i2} - k_{22})_+ \\ +5498.3889(z_{i2} - k_{32})_+ + 513.3825 z_{i3} - 483.1111(z_{i3} - k_{13})_+\end{gathered} \tag{4}$$

3.4 Parameter Significance Result

The parameter significance test was carried out in two ways, namely, the simultaneous test and the partial test. Simultaneous test results are shown in **Table 2**.

Table 2. Simultaneous Test Results

Source	DF	SS	MS	F	p-values
Regression	11	6,883,424,490	5,171,220,408	431.4116	9.261x10^{-53}
Error	60	719,204,717	119,867,45		
Total	71	5,760,269,207			

Table 2 shows a value F of 431.4116 and $p-\text{value} = 9.261 \times 10^{-53} < \alpha = 0.05$. From these results, it can be concluded that H_0 is rejected, which means that the predictor variables simultaneously significantly influence the model. The partial test results are shown in **Table 3**.

Table 3. Partial Test Results

Variable	Parameter	*Q*	*P*-values	Conclusion
X_1	β_0	1.9863	0.0515	H_0 is not rejected
	β_1	65.3685	1.7977×10^{-57}	H_0 is rejected
X_2	γ_{12}	1.2507	0.2158	H_0 is not rejected
	γ_{22}	0.8001	0.4267	H_0 is not rejected
	γ_{32}	1.8050	0.0760	H_0 is not rejected
	γ_{42}	2.0551	0.0442	H_0 is rejected
X_3	γ_{13}	1.1953	0.2366	H_0 is not rejected
	γ_{23}	2.1759	0.0335	H_0 is rejected
	γ_{33}	2.1248	0.0377	H_0 is rejected

Variable	Parameter	*Q*	*P*-values	Conclusion
	γ_{43}	2.0958	0.0403	H_0 is rejected
X_4	γ_{14}	2.9482	0.0045	H_0 is rejected
	γ_{24}	1.6883	0.0965	H_0 is not rejected

Table 3 shows that six parameters produce p_{value} less than 0.05, namely variables: paddy harvest area, average temperature, average humidity, and rainfall. It can be concluded that all predictor variables have a significant effect on rice production.

3.5 Error Assumptions Result

The normal distribution assumptions are tested using the Kolmogorov-Smirnov test. Based on the test results, $\alpha = 0.05$ obtained by the value $p - \text{value} = 0.096 > \alpha$, so H_0 failed to be rejected, which means that the error follows a normal distribution.

Testing the assumption of heteroscedasticity using the Glejser test. Based on the test results, $\alpha = 0.05$ obtained a value F of 7,767,333 and a value F_{table} of 1.5937 greater than α, which means H_0 failed to be rejected, meaning there is no heteroscedasticity error.

3.6 Goodness Model

Based on the analysis results, the adjusted R^2 value was 0.98522, which means that the predictor variables simultaneously affected rice production by 98.522%, and 1.478% was influenced by other variables outside the variables used in this study. The analysis produced an RMSE value of 3160.531.

3.7 Model Interpretation

Equation (4) becomes Equation (5) if the variables other than X_1 are constant, the effect of paddy harvest area on rice production is

$$\hat{y} = 5.8748\, x_{i1} \tag{5}$$

This means that increasing the paddy harvested area by one hectare will increase rice production by 5.8748 tons.

Equation (4) becomes Equation (6) if the variables other than Z_1 are constant, the effect of temperature on rice production is

$$\hat{y} = -2144.1090 z_{i1} + 317.1016(z_{i1} - 22.4119\ )_+ + 1094.0594(z_{i1} - 22.5638\ )_+ \\ +1870.9661(z_{i1} - 22.7157)_+ \tag{6}$$

$$\hat{y} = \begin{cases} -2144.1090 z_{i1} & , \quad z_{i1} < 22.4119 \\ -1827.0074 z_{i1} & , \quad 22.4119 \ \leq z_{i1} < 22.5638 \\ -732.948 z_{i1} & , \quad 22.5638 \leq z_{i1} < 22.7157 \\ 1138.0181 z_{i1} & , \quad z_{i1} \geq 22.7157 \end{cases}$$

If the temperature is less than 22.41 degrees Celsius, increasing the temperature by one unit will reduce rice production by 2144.1090 tons. Areas whose temperatures fall within the interval $z_{i1} < 22.4119$ are Malang. If the temperature is between 22.41 and 22.56 degrees Celsius, rising one temperature unit will reduce rice production by 1827.0074 tons. Areas whose temperatures fall within the interval $22.4119\ \leq z_{i1} < 22.5638$ are Malang. If the temperature is between 22.56 and 22.71 degrees Celsius, increasing one temperature unit will reduce rice production by 732.948 tons. Areas whose temperatures fall within the interval $22.5638 \leq z_{i1} < 22.7157$ are Malang. If the temperature is more than 22.71 degrees Celsius, increasing the temperature by one degree Celsius will increase rice production by 732.948 tons. Areas whose temperatures fall within the interval $z_{i1} \geq 22.7157$ are Tuban, Gresik, Nganjuk, Pasuruan, and Banyuwangi.

Equation (4) becomes Equation (7) if the variables other than Z_2 are constant, the effect of air humidity on rice production is

$$\hat{y} = 609.3392z_{i2} - 5288.0700\,(z_{i2} - 76.2571)_+ - 981.7648(z_{i2} - 76.7142)_+ + 5498.3889(z_{i2} - 77.1714)_+ \quad (7)$$

$$\hat{y} = \begin{cases} 609.3392z_{i2} & , \quad z_{i2} < 76.2571 \\ -4678.7308z_{i2} & , \quad 76.2571 \le z_{i2} < 76.7142 \\ -5660.4956z_{i2} & , \quad 76.7142 \le z_{i2} < 77.1714 \\ -162.1067z_{i2} & , \quad z_{i2} \ge 77.1714 \end{cases}$$

If the air humidity is less than 76.2571, a one percent increase will increase rice production by 609.3392 tons. Areas whose humidity falls within the interval $z_{i2} < 76.2571$ are Tuban, Nganjuk, and Pasuruan; if the air humidity is at 76.2571 up to 76.7142, a one percent increase, rice production will be reduced by 4678.7308 tons. Areas with humidity within the interval $76.2571 \le z_{i2} < 76.7142$ are Pasuruan. If the air humidity is at 76.7142 up to 77.1714, a one percent increase, it will reduce rice production by 5660.4956 tons. Areas with humidity within the interval $76.7142 \le z_{i2} < 77.1714$ are Tuban, Gresik, Nganjuk, Pasuruan, and Banyuwangi. If the air humidity exceeds 77.1714, a one percent increase will reduce rice production by 162.1067 tons. Areas with humidity within the interval $z_{i2} \ge 77.1714$ are Tuban, Gresik, Nganjuk, Pasuruan, Banyuwangi, and Malang.

Equation (4) becomes Equation (8) if the variables other than Z_3 are constant, the effect of rainfall on rice production is

$$\hat{y} = 513.3825z_{i3} - 483.1111(z_{i3} - 10.2314)_+ \quad (8)$$

$$\hat{y} = \begin{cases} 513.3825z_{i3} & , \quad z_{i3} < 10.2314 \\ 30.2714z_{i3} & , \quad z_{i3} \ge 10.2314 \end{cases}$$

If the rainfall is less than 10.2314, a one-unit increase will increase rice production by 513.3825 tons. Areas with rainfall within the interval $z_{i3} < 10.2314$ are Tuban, Gresik, Nganjuk, Pasuruan, Banyuwangi, and Malang. If rainfall exceeds 10.2314, a one-unit increase will increase rice production by 30.2714 tons. Areas with rainfall within the interval $z_{i3} \ge 10.2314$ are Tuban, Gresik, Nganjuk, Pasuruan, Banyuwangi, and Malang.

4. CONCLUSION

The best model for modeling rice production in East Java in 2022 is the truncated spline semiparametric regression model with a combination of knot points (3,3,1) with a minimum GCV value of 12642272. Variables significantly influencing rice production are paddy harvest area, temperature, air humidity, and rainfall. The model produces a value R^2_{adj} of 98.522% and RMSE of 3160.531. In order to meet food needs, the paddy harvest area can increase rice production, so the paddy harvest area can be a special concern for the Indonesian government. Climate conditions are important for rice plants in the process of producing quality rice. Increasing the area of the paddy can improve the quantity of rice production. Predictions of rice production in Indonesia can be used to support one of the SDGs points, Zero Hunger.

AUTHOR CONTRIBUTIONS

Sri Sulistijowati Handajani: Ideas; formulation or evolution of overarching research goals and aims; management and coordination responsibility for the research activity planning and execution; application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize the research data; preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages. Hasih Pratiwi: Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team; application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize the research data; acquisition of the financial support for the project leading to this publication. Respatiwulan: Conducting research and the investigation process, specifically performing the experiments, or data/evidence collection. Yuliana Susanti: Development or design of methodology; creation of models. Muhammad Bayu Nirwana: Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.

Lintang Pramesti Nareswari: Management activities to annotate (produce metadata), scrub data, and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This research was funded by a grant from Sebelas Maret University in 2023 with contract number 228/UN27.22/PT.0l.03/2023.

ACKNOWLEDGMENT

The first author acknowledges the financial support from the Fundamental Research in Universitas Sebelas Maret Indonesia. The authors are grateful to the anonymous referee for carefully checking the details and for the helpful comments that improved this paper.

CONFLICT OF INTEREST

The authors declare no conflicts of interest to report study.

REFERENCES

[1] R. S. Pirngadi and Rahmawaty, “THE IMPACT OF FLOODING ON RICE PRODUCTION IN THE KRUENG KLUET WATERSHED, ACEH PROVINCE, INDONESIA,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 977, no. 1, 2022, doi: https://doi.org/10.1088/1755-1315/977/1/012113.

[2] J. Cao *et al.*, “INTEGRATING MULTI-SOURCE DATA FOR RICE YIELD PREDICTION ACROSS CHINA USING MACHINE LEARNING AND DEEP LEARNING APPROACHES,” *Agric. For. Meteorol.*, no. 297, pp. 1–15, 2021.doi: https://doi.org/10.1016/j.agrformet.2020.108275

[3] E. Kamir, F. Waldner, and Z. Hochman, “ESTIMATING WHEAT YIELDS IN AUSTRALIA USING CLIMATE RECORDS, SATELLITE IMAGE TIME SERIES AND MACHINE LEARNING METHODS,” *J. Photogramm. Remote Sens.*, no. 160, pp. 124–135, 2020.doi: https://doi.org/10.1016/j.isprsjprs.2019.11.008

[4] P. J. T. Badan Pusat Statistik, “PRODUKSI PADI DAN BERAS MENURUT KABUPATEN/KOTA DI PROVINSI JAWA TIMUR, 2021 DAN 2022,” *BPS*, 2023. https://jatim.bps.go.id/id/statistics-table/1/MjUyMyMx/-produksi-padi-dan-beras-menurut-kabupaten-kota-di-provinsi-jawa-timur-2021-dan-2022-.html (accessed Mar. 16, 2023).

[5] V. D. Islami, R. Fitriani, and H. Pramoedyo, “SPATIALLY FILTERED RIDGE REGRESSION MODELING TO FIND OUT THE RICE PRODUCTION FACTORS IN EAST JAVA, INDONESIA,” *CommIT J.*, vol. 14, no. 2, pp. 95–102, 2020, doi: https://doi.org/10.21512/commit.v14i2.6665.

[6] D. Darmadi, A. Junaedi, D. Sopandie, Supijatno, I. Lubis, and K. Homma, “WATER-EFFICIENT RICE PERFORMANCES UNDER DROUGHT STRESS CONDITIONS,” *AIMS Agric. Food*, vol. 6, no. 3, pp. 838–863, 2021, doi: https://doi.org/10.3934/agrfood.2021051.

[7] W. Sanusi, R. Syam, and R. Adawiyah, “MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN SPLINE (STUDI KASUS: BERAT BADAN LAHIR RENDAH DI RUMAH SAKIT IBU DAN ANAK SITI FATIMAH MAKASSAR),” *J. Mat. dan Komputasi*, vol. 2, no. 1, pp. 70–81, 2017.doi: https://doi.org/10.35580/jmathcos.v2i1.12460

[8] M. Kaseside and S. B. Loklomin, “ANALISIS ANGKA KEMATIAN BAYI KABUPATEN HALMAHERA UTARA DENGAN METODE REGRESI NONPARAMETRIK SPLINE TRUNCATED,” *J. Stat.*, vol. 3, no. 1, pp. 1–5, 2021.doi: https://doi.org/10.30598/variancevol3iss1page1-5

[9] M. F. F. Mardianto, S. M. Ulyah, and E. Tjahjono, “PREDICTION OF NATIONAL STRATEGIC COMMODITIES PRODUCTION BASED ON MULTI-RESPONSE NONPARAMETRIC REGRESSION WITH FOURIER SERIES ESTIMATOR,” *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 1151–1176, 2019.

[10] A. Tsalasatul Fitriyah, N. Chamidah, and T. Saifudin, “PREDICTION OF PADDY PRODUCTION IN INDONESIA USING SEMIPARAMETRIC TIME SERIES REGRESSION LEAST SQUARE SPLINE ESTIMATOR,” *Data Metadata*, vol. 4, 2025, doi: https://doi.org/10.56294/dm2025527.

[11] Liu.Y. and L. Shuwey, “A SEMIPARAMETRIC MIXTURE CURE MODEL FOR PARTLY INTERVAL CENCORED FAILURE TIME DATA,” *J. Stat. Appl. Probablity*, vol. 10 No.1, pp. 1–10, 2021.doi: https://doi.org/10.18576/jsap/100101

[12] E. D. Igustin and I. N. Budiantara, “PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI TOTAL FERTILITY RATE DI INDONESIA MENGGUNAKAN REGRESI NONPARAMETRIK SPLINE TRUNCATED,” *J. Sains dan Seni*, vol. 9, no. 2, pp. 178–185, 2020.doi: https://doi.org/10.12962/j23373520.v9i2.56791

[13] J. Li, W. Zhang, P. Wang, Q. Li, K. Zhang, and Y. Liu, “NONPARAMETRIC PREDICTION DISTRIBUTION FROM RESOLUTION-WISE REGRESSION WITH HETEROGENEOUS DATA,” *J. Bus. Econ. Stat.*, vol. 41, no. 4, pp. 1157–1172, Oct. 2023, doi: 1 https://doi.org/10.1080/07350015.2022.2115498.

[14] F. Ubaidillah, A. A. . Fernande, A. Iriany, N. W. S. Wardhani, and Solimun., "TRUNCATED SPLINE PATH ANALYSIS MODELLING ON IN COMPANY X WITH THE GOVERNMENTS ROLE AS A MEDIATION VARIABLE," *J. Stat. Appl. Probab.*, vol. 11, no. 3, pp. 781–794, 2022.doi: https://doi.org/10.18576/jsap/110303

[15] A. F. D. Rositawati and I. N. Budiantara, "PEMODELAN INDEKS KEBAHAGIAAN PROVINSI DI INDONESIA MENGGUNAKAN REGRESI NONPARAMETRIK SPLINE TRUNCATED," *J. Sains dan Seni*, vol. 8, no. 2, pp. 113–120, 2020.doi: https://doi.org/10.12962/j23373520.v8i2.45160

[16] R. K. Sifriyani, Syaripuddin, M. Fathurahman, Nariza Wanti Wulan Sari, Meirinda Fauziyah, Andrea Tri Rian Dani, Raudhatul Jannah, S. Dwi Juriani, "NONPARAMETRIC SPATIO-TEMPORAL MODELING: CONTRUCTION OF A GEOGRAPHICALLY AND TEMPORALLY WEIGHTED SPLINE REGRESSION," *MethodsX*, vol. 14, no. December 2024, p. 103098, 2025, doi: https://doi.org/10.1016/j.mex.2024.103098.

[17] A. S. Suriaslan, I. N. Budiantara, and V. Ratnasari, "NONPARAMETRIC REGRESSION ESTIMATION USING MULTIVARIABLE TRUNCATED SPLINES FOR BINARY RESPONSE DATA," *MethodsX*, vol. 14, no. August 2024, 2025, doi: https://doi.org/10.1016/j.mex.2024.103084.

[18] Z. Zhang, "PARAMETRIC REGRESSION MODEL FOR SURVIVAL DATA: WEIBULL REGRESSION MODEL AS AN EXAMPLE," *Ann. Transl. Med.*, vol. 4, no. 24, 2016, doi: https://doi.org/10.21037/atm.2016.08.45.

[19] R. L. Eubank, *NONPARAMETRIC REGRESSION AND SPLINE SMOOTHING*. New York: CRC Press, 1999. doi: https://doi.org/10.1201/9781482273144.

[20] A. Caron, G. Baio, and I. Manolopoulou, "ESTIMATING INDIVIDUAL TREATMENT EFFECTS USING NON-PARAMETRIC REGRESSION MODELS: A REVIEW," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 185, no. 3, pp. 1115–1149, 2022, doi: https://doi.org/10.1111/rssa.12824.

[21] B. Lestari, I. N. Budiantara, S. Sunaryo, and M. Madhuri, "SPLINE ESTIMATOR OF TRIPLE RESPONSE NONPARAMETRIC REGRESSION MODEL," *J. Math. Stat.*, vol. 6, pp. 327–332, 2010.doi: https://doi.org/10.3844/jmssp.2010.327.332

[22] J. Yan, "ALMOST SURE CONVERGENCE FOR WEIGHTED SUMS OF WNOD RANDOM VARIABLES AND ITS APPLICATIONS TO NONPARAMETRIC REGRESSION MODELS," *Commun Stat - Theory Methods*, vol. 47, no. 16, pp. 3893–3909, 2018, doi: https://doi.org/10.1080/03610926.2017.1364390.

[23] L. Hidayati, N. Chamidah, and I. Nyoman Budiantara, "SPLINE TRUNCATED ESTIMATOR IN MULTIRESPONSE SEMIPARAMETRIC REGRESSION MODEL FOR COMPUTER BASED NATIONAL EXAM IN WEST NUSA TENGGARA," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: https://doi.org/10.1088/1757-899X/546/5/052029.