Data and Metadata. 2025; 4:527
doi: 10.56294/dm2025527

ORIGINAL

Prediction of Paddy Production in Indonesia Using Semiparametric Time Series Regression Least Square Spline Estimator

Predicción de la producción de arroz en Indonesia mediante un modelo de regresión de series temporales semiparamétricas basado en el estimador de spline de mínimos cuadrados

Any Tsalasatul Fitriyah[1,2], Nur Chamidah[3,4], Toha Saifudin[3,4]

[1]Airlangga University, Doctoral Study Program of Mathematics and Natural Sciences, Faculty of Science and Technology. Surabaya, Indonesia.
[2]Islamic University of Mataram, Department of Mathematics Education, Faculty of Teacher Training and Education. Mataram, Indonesia.
[3]Airlangga University, Department of Mathematics, Faculty of Science and Technology. Surabaya, Indonesia.
[4]Airlangga University, Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology. Surabaya, Indonesia.

Cite as: Tsalasatul Fitriyah A, Chamidah N, Saifudin T. Prediction of Paddy Production in Indonesia Using Semiparametric Time Series Regression Least Square Spline Estimator. Data and Metadata. 2025; 4:527. https://doi.org/10.56294/dm2025527

Submitted: 17-04-2024 **Revised:** 16-08-2024 **Accepted:** 30-11-2024 **Published:** 01-01-2025

Editor: Adrián Alejandro Vitón Castillo

Corresponding Author: Nur Chamidah

ABSTRACT

Support for one of the points of Sustainable Development Goals (SDGs), namely Zero Hunger, is by supporting sustainable agricultural empowerment. Indonesia is one of the countries with the fourth largest rice consumption according to the United States Department of Agriculture. 90 % of Indonesians consume rice as a staple food. In this study, we model paddy production in Indonesia using a semiparametric time series regression approach based on least square spline estimator (LSSE). Where spline is used to overcome data that tends to fluctuate in monthly paddy production data. Monthly data on paddy production in Indonesia over a certain period of time is used to build a model. The use of a semiparametric regression approach by combining parametric components and nonparametric components for analyzing factors that affect paddy production. In this study, the parametric component is paddy production in the previous period lag-1 and the nonparametric components are the potential area of crop failure and the generative area. For predicting paddy production in Indonesia using Semiparametric Time Series Regression Model (STSRM) approach based on LSSE, we determine the order and optimal knot points based on the smallest Generalized Cross Validation (GCV) value. The results of the study show that the Mean Absolute Percentage Error (MAPE) value of 18,05 % is less than 20 %. It means that prediction of paddy production in Indonesia using STSRM based on LSSE is a good prediction.

Keywords: Sustainable Agriculture; Time Series; Semiparametric Regression; Least Square Spline Estimator.

RESUMEN

El apoyo a uno de los puntos de los Objetivos de Desarrollo Sostenible (ODS), a saber, Hambre Cero, pasa por apoyar la potenciación de la agricultura sostenible. Indonesia es uno de los países con el cuarto mayor consumo de arroz según el Departamento de Agricultura de Estados Unidos. El 90 % de los indonesios consumen arroz como alimento básico. En este estudio, modelizamos la producción de arroz con cáscara en Indonesia utilizando un enfoque semiparamétrico de regresión de series temporales basado en el estimador spline de mínimos cuadrados (LSSE). El spline se utiliza para superar los datos que tienden a fluctuar en los datos mensuales de producción de arroz. Los datos mensuales sobre la producción de arroz en Indonesia durante un cierto periodo de tiempo se utilizan para construir un modelo. El uso de un enfoque de regresión semiparamétrica mediante la combinación de componentes paramétricos y componentes no paramétricos para analizar los factores que afectan a la producción de arroz. En este estudio, el componente paramétrico es la producción de arroz en el periodo anterior lag-1 y los componentes no paramétricos son el área potencial de fracaso de la cosecha y el área generativa. Para predecir la producción de arroz en Indonesia utilizando el enfoque del Modelo de Regresión de Series Temporales Semiparamétricas (STSRM) basado en LSSE, determinamos el orden y los puntos de nudo óptimos basándonos en el menor valor de Validación Cruzada Generalizada (GCV). Los resultados del estudio muestran que el valor del Error Medio Porcentual Absoluto (MAPE) del 18,05 % es inferior al 20 %. Esto significa que la predicción de la producción de arroz en Indonesia utilizando STSRM basado en LSSE es una buena predicción.

Palabras clave: Agricultura Sostenible; Series Temporales; Regresión Semiparamétrica; Estimador Spline de Mínimos Cuadrados.

INTRODUCTION

One of the SDGs points in order to support community welfare according to the United Nation (UN) is Zero Hunger. Hunger is one of the problems in the world that affects Public Health. This hunger is caused by food stock insecurity, poor food quality, and waste.[1] According to the World Food Programme (WFP), the projected increase in human population will reach two billions people in 2050. However, in recent years there has been a downward trend in food production and in amount 8,9 % of the world's population is in a state of hunger.[2] Therefore, food security has a very important role in achieving the Zero Hunger point. To achieve this, support is needed to empower sustainable agriculture.

Paddy is an important sector that dominates food production in many countries, including Indonesia. The majority of Indonesian people consume rice as a staple food.[3] According to the Badan Pangan Nasional (Bapanas), Indonesian people's rice consumption in 2023 reached 81,23 kg/capita/year. Meanwhile, the need for rice for household consumption reached 22,64 million tons per year in 2023. Ideally, rice consumption should be comparable to the paddy production produced in order to achieve food stock security. However, paddy production in Indonesia fluctuates over time depending on various factors that influence it.[4] One of the elements influencing the amount of paddy production is still a potential crop failure as an effect of climate change.[5] Apart from the potential for crop failure, the quality of the generative phase, including the size and health of the paddy plants, is also a factor that influences paddy production.[6] As the crucial nature of paddy production, precise and accurate predictions regarding paddy production are needed in order to plan and manage paddy production.

Regression models are statistical methods used to explain the functional relationship between predictor and response variables. Semiparametric, parametric, and nonparametric regression models are used frequently. Parametric regression models are used if the modeling of the relationship between predictor variables and response variables is known. Furthermore, parametric regression needs fullfiled assumptions.[7] Research conducted by Ansharifar et al.[8] using a parametric regression approach to predict corn and soybean yields in the US. The COVID-19 case fatality rate in Nigeria was also predicted using parametric regression approaches.[9] In contrast to the parametric regression approach, nonparametric regression does not require assumptions to be met.[10] If it is unknown where the predictor and response variables relate to each other, the nonparametric approach is used.[11] The curve in nonparametric regression is assumed to be contained in a certain function depending on the data used.[12,13] Semiparametric regression is a combination of parametric and nonparametric regression.[14] In the application of nonparametric and semiparametric regression, there are several smoothing techniques used, including: kernel,[15,16] local linear,[17,18,19,20] local polynomial,[21,22] spline,[23,24,25] penalized spline,[26,27] and fourier series.[28,29]

The application of regression analysis does not only cover cross-section data, but also time series data. Time series data is data taken at certain intervals periodically, such as annual, monthly, or even daily data.[30] Classical linear models such Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) are usually employed for modeling time series data. Previous research was conducted by Tripathi et al.[31] and Kasthuri et al.[32] using ARIMA analysis in building a prediction model for Food Production.

The development of research on time series data also applied to nonparametric and semiparametric approaches. Among them is research conducted by Ahmed et al.[33] and Fibiyani et al.[34] using local polynomial estimators to perform time series data analysis. By selecting of local linear estimators in time series regression analysis, both nonparametric and semiparametric, is also carried out by Li et al.[35] and Vilar et al.[36]. On research conducted by Gao et al.[37], Yan et al.[38] and Cen et al.[39] also, use the kernel estimator to analyze time series research.

Spline is a smoothing technique that is widely used in both nonparametric and semiparametric approaches.[40] Spline is considered capable of dealing with data patterns that tend to fluctuate by placing knot points according to the value of the Generalized Cross Validation (GCV).[41,42] According to Lestari[43], the smallest GCV value is chosen to produce an optimal model.

This study aims to estimate Indonesian paddy production through implementing a STSRM based on LSSE. The results of this study are expected to be used as a prediction model for paddy production in Indonesia.

METHOD

Data Set and Research Variables

The data in this study uses secondary data taken from the Badan Pusat Statistik (BPS) of Indonesia. The data in this study are monthly data from January 2018 to December 2023. In this study, the paddy production variable is used as the response variable. The predictor variables in this study are the paddy production in previous period lag-1 variable (y_{t-1}) as a parametric component and the area of potential crop failure variable (z_{1t}) and generative área variable (z_{2t}) as nonparametric component.

Semiparametric Time Series Regression Model (STSRM) Based on Least Square Spline Estimator (LSSE)

Given a STSRM as follows:[33]

$$Y_t = \mathbf{X_t\beta} + f(z_t) + \varepsilon_t \qquad ,t = 1 \le t \le n \quad (1)$$

Where:
Y_t = response variable on the t-th data.
X_t =(X_{1t},..., X_{pt}) is a (nxp)-dimensional matrix of predictor variables on the t-th data.
β = (px1)-dimensional vector of regression coefficient.
$f(Z_t)$ = the unknown t-th nonparametric function.
ε = the t-th error.

If the fuction of $f(Z_t)$ is estimated using function of spline with order p and knot points $\tau_1,\tau_2,...,\tau_k$, function $f(Z_t)$ can be expressed as follows:

$$f(z_t) = \sum_{j=0}^{p} \beta_j z_t^{\,j} + \sum_{k=1}^{K} \beta_{p+k}(z_t - \tau_k)_+^p, t = 1,\cdots,T \quad (2)$$

Where:

$$(z_t - \tau_k)_+^p = \begin{cases} (z_t - \tau_k)^p ; z_t \ge \tau_k \\ 0 \qquad\qquad ; z_t \ge \tau_k \end{cases} \quad (3)$$

The knot point is a typical fusion point where the function's or the curve's behavioral pattern transforms. To get the best model, it is necessary to select the optimal knot point based on the smallest GCV value. The formula of GCV is as follows:[44]

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left[\frac{1}{n} trace(\mathbf{I} - \mathbf{A}(\lambda))\right]^2} \quad (4)$$

Where λ=p,$\tau_1,\tau_2,...,\tau_k$ are smoothing parameter, *p* is order spline, τ is knot points, and $A=Z_\lambda(Z_\lambda^T Z\lambda)^{-1} Z_\lambda^T$.

The model in this research refers to the model described in equation (1) which has one nonparametric component, whereas in this research the model is developed by adding one nonparametric component and the parametric component is an autoregressive lag-1 of the response variable. The STSRM used in this research is as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + f(z_{1t}) + f(z_{2t}) + \varepsilon_t \quad (5)$$

Where y_t is response variable on the t-th data; y_{t-1} is a parametric components in the previous period; z_{1t} is the first nonparametric component on the t-th data; z_{2t} is the second nonparametric component on the t-th data; and ε_t are random errors.

This study produces a prediction model. Based on the MAPE value, an error level measurement is carried out to assess the model's goodness of fit:(45)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100 \quad (6)$$

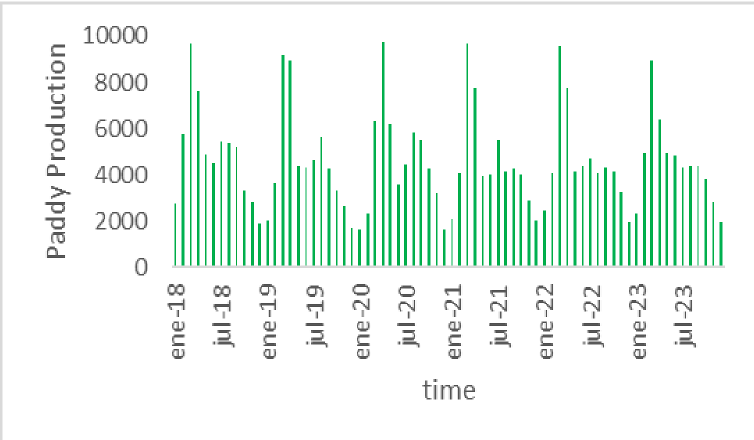
MAPE accuracy categories are:

Table 1. Interpretation of MAPE Value

MAPE Value	Interpretation
MAPE > 50	Inaccurate
20 < MAPE ≤ 50	Reasonable
10 < MAPE ≤ 20	Good Prediction
MAPE ≤ 10	High Accurate Prediction

RESULTS AND DISCUSSION

Paddy is a major commodity in many countries including Indonesia. Indonesia, which is an agricultural country, is one of the largest paddy producers. The development of paddy production in Indonesia tends to be stable every year. Paddy production in Indonesia tends to be high in certain months depending on the harvest period. Paddy production trend in Indonesia is depicted in figure 1.



Source: https://www.bps.go.id/
Figure 1. Trend of Paddy Production in Indonesia

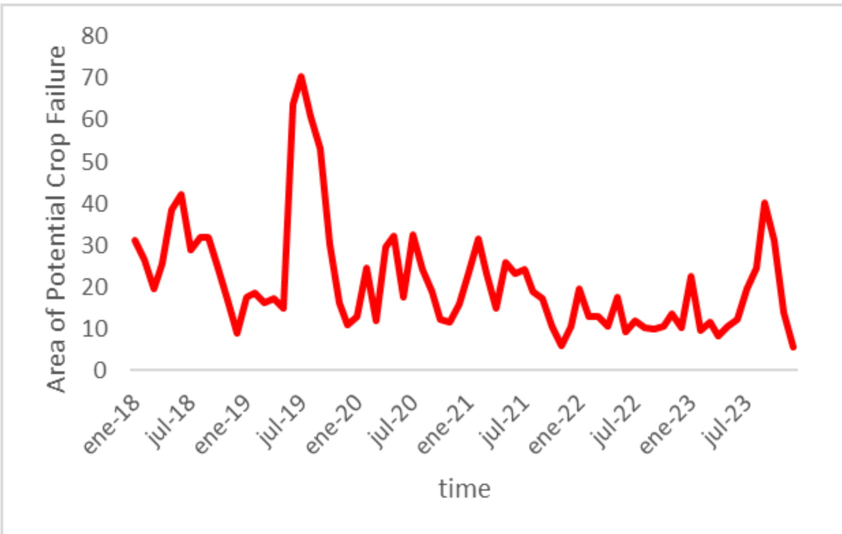


Figure 2. The Board Trend of Potential Crop Failure in Indonesia

https://doi.org/10.56294/dm2025527

Paddy production depends not only on the land area planted with paddy but also on other factors that can affect paddy production. Factors that affect paddy production include the area of potential crop failure and the generative area. The area of potential crop failure is obtained based on the area of paddy plants that have the potential to fail or harvest less than 11 % of normal conditions, usually characterized by damaged or unsuitable land. The broad trend of potential crop failure in Indonesia in the last six years has tended to decrease, only at certain times there has been an increase. This is caused by pest attacks, extreme weather, or other causes. The broad trend of potential crop failure is depicted in figure 2.

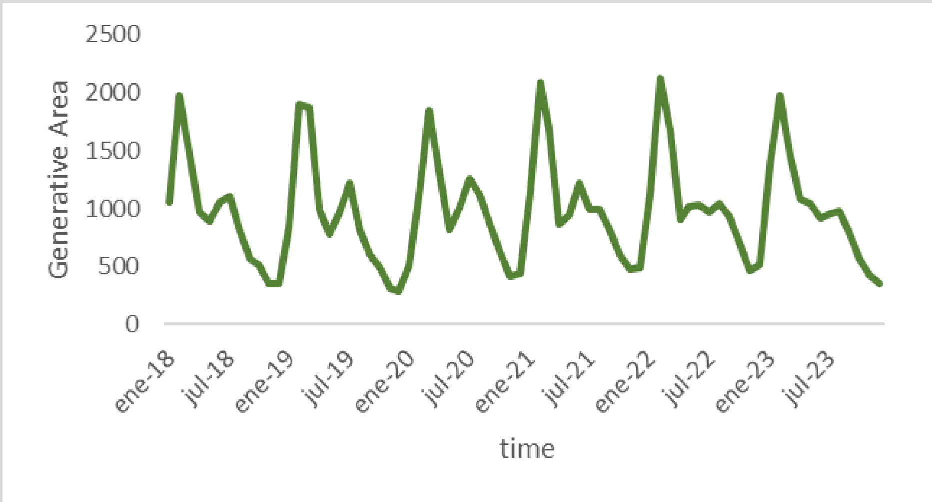


Figure 3. The Trend of Generative Area

Figure 3 depicts trends in the condition of the generative area of paddy plants in Indonesia. The generative phase can be used for potential harvest within the next one to three months. The generative area in Indonesia tends to be highest from January to April. This is in line with paddy production patterns which on average peak in March or April each year. The descriptive statistics of the variables used in this study are presented in the following table 2:

Table 2. Descriptive Statistic of Variables

Variable	Min	Max	Mean	Variance
Paddy Production	1620	9768	4605,514	4408349
Potential Area of Crop Failure	5,66	70,41	21,46792	171,0787
Generative Area	284	2122	974,9861	204652,1

In analyzing paddy production predictions, paddy production in the previous period lag-1 is used as a reference in studying seasonal patterns and long-term trends. In this section, we present a scatter plot for the variable of paddy production in the previous period against the variable of current paddy production, the variable of the area of potential crop failure against the variable of paddy production, and the variable of the area of the generative phase against paddy production. All these plots are given in the figure 4 - figure 6:

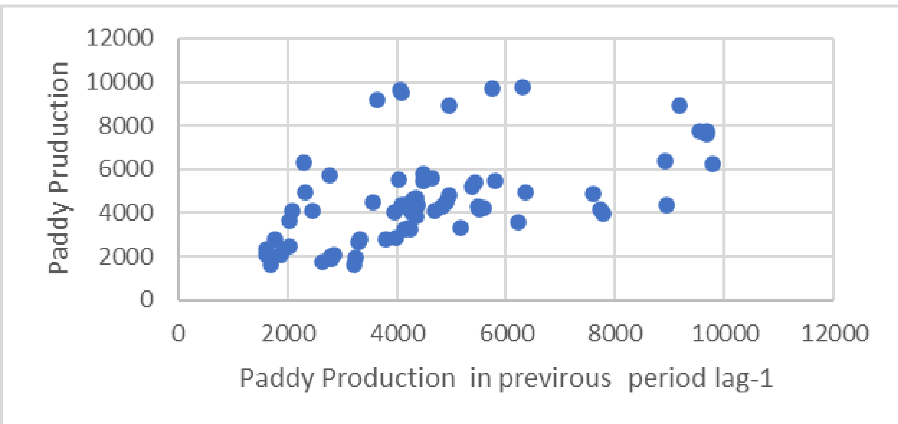


Figure 4. Scatterplot for Paddy Production in Previous Period versus Paddy Production

https://doi.org/10.56294/dm2025527

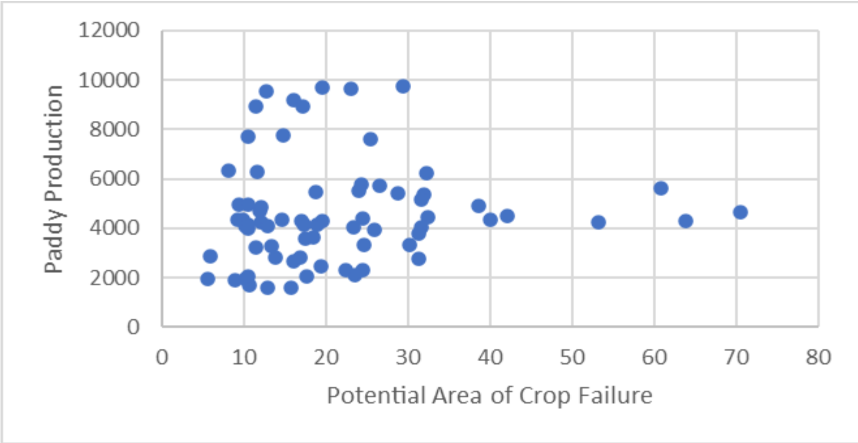


Figure 5. Scatter plot for Potential area of crop failure versus paddy production

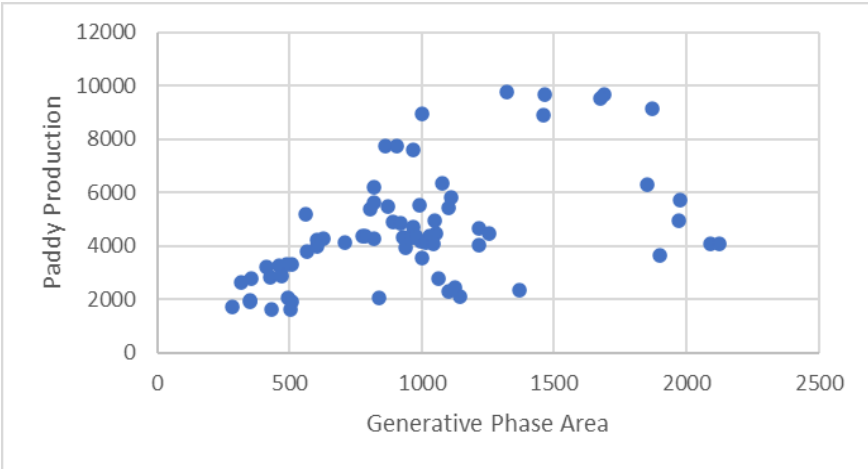


Figure 6. Scatter plot for generative phase area versus paddy production

In figure 4 it can be seen that the relationship between paddy production in the previous period lag-1 and current paddy production follows a linear pattern, so that the variable of paddy production in the previous period becomes a parametric component. While in figure 5 and figure 6 the relationship pattern of the variable area of potential crop failure and generative phase area to paddy production does not have a regular pattern so these two variables become nonparametric components.

In this study, to obtain the model STSRM was carried out based on LSSE. The model was built based on the order and knot points obtained from the smallest GCV value. The determination of the order and knots based on the GCV value is presented in the following table 3:

Table 3. Estimation results using STSRM based on LSSE

Knot of Potential Area of Crop Failure Variable			GCV	Knot of Generative Area Variable			GCV
12,55			3604436	604,5			1961573
18,065			3648827	951,5			1945256
27,085			3652928	1115,75			1971058
12,55	18,065		3734040	604,5	951,5		2015340
12,55		27,085	3720462	604,5		1115,75	2031887
	18,065	27,085	3772785		951,5	1115,75	1936159
12,55	18,065	27,085	3843402	604,5	951,5	1115,75	2000303

Based on table 3, the smallest GCV value for potential area of failure variable is 1 knot point, which is 18 065 with a GCV value of 3 648 827. For generative area variable, the smallest GCV value is 1 936 159 with 2 knot points, which are 951,5 and 1115,75. So to build the model, a combination of one knot point is chosen for potential area of Failure variable and 2 knot points for generative area variable. The STSRM based on LSSE is as follows:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{f}(z_{1t}) + \hat{f}(z_{2t})$$
$$= 106.256 + 0.7259 y_{t-1} - 27.628 z_{1t} + 0.715(z_{1t} - 12.55)_+ + 28.39 z_{2t}$$
$$+ 9.327(z_{2t} - 951.5)_+ - 7.376(z_{2t} - 1115.75)_+ \quad (7)$$

Based on equation equation (7), the relationship between paddy production and paddy production in the previous period lag-1 can be explained in the following equation (8):

$$\hat{y}_t = 106.256 + 0.7259 y_{t-1} + f(z_{1t}) + f(z_{2t}) \quad (8)$$

The relationship between paddy production is influenced by paddy production in the previous period lag-1. If $f(z_{1t})$ and $f(z_{2t})$ are considered constant, then y_t and y_{t-1} have a positive relationship. By looking at equation (8) it can be explained that paddy production in the previous period increased by one unit, then current paddy production will increase by 0,7259 if $f(z_{1t})$ and $f(z_{2t})$ are considered constant.

The nonparametric function that explains the potential area of crop failure variable can be expressed in equation (9):

$$\hat{f}(z_{1t}) = -27.628 z_{1t} + 0.715(z_{1t} - 12.55)_+ \quad (9)$$

Equation (9) can be explained as follows:

$$\hat{f}(z_{1t}) = \begin{cases} -27.628 z_{1t} & ;(z_{1t} \leq 12.55) \\ -8.973 - 26.913 z_{1t} & ;(z_{1t} > 12.55) \end{cases} \quad (10)$$

The paddy production variable is influenced by the potential for crop failure variable has negative relationship as explained in equation (10). If other variables are considered constant and the area of potential crop failure is less than or equal to 12,55 thousand hectares and the potential for crop failure increases by one unit, then paddy production will decrease by 27 628 if other variables are assumed constant. If the potential for crop failure variable is more than 12,55 thousand hectares and increases by one unit, then paddy production will decrease by 26 913. This result is reinforced by research conducted by Agustiarini et al.(46) which states that the potential for crop failure will have an impact on reducing the quantity and quality of paddy production and causing losses to agricultural businesses. The potential for crop failure should be a focus of attention for the Indonesian government so that it can realize sustainable agriculture in order to increase paddy production. Focusing on reducing the potential area for crop failure can be done by improving soil quality.

The nonparametric function that explains the generative area variable can be expressed in equation (11):

$$\hat{f}(z_{2t}) = 28.39 z_2 + 9.327(z_2 - 951.5)_+ - 7.376(z_2 - 1115.75)_+ \quad (11)$$

Equation (11) can be explained as follows:

$$\hat{f}(z_{2t}) = \begin{cases} 28.39 z_2 & ;(z_2 \leq 951.5) \\ -8874.64 + 37.717 z_2 & ;(951.5 < z_2 \leq 1115.75) \\ -644.868 + 30.341 z_2 & ;(z_2 > 1115.75) \end{cases} \quad (12)$$

In contrast to the relationship influenced by the potential crop failure variable, the generative area variable has a positive relationship to the paddy production variable. The relationship between variables is explained in equation (11). If other variables are assumed to be constant and the generative phase area is less than or equal to 951,5 thousand hectares and increases by one unit, then paddy production will increase by 28,39. If the generative area is between 951,5 to 1115,75, increases by one unit and other variables are considered constant, then paddy production will decrease by 37 717. Paddy production will increase by 30,41 if the generative area increases by one unit and is more than 1115,75 thousand hectares. The generative phase is an important phase in order to produce maximum rice production.(47,48) The generative phase is the paddy reproduction period, where important nutrients are needed that play a role in the health of paddy plants to produce maximum paddy harvests.

Figure 7 is the result of the estimation plot of the semiparametric regression model based on the least square spline estimator. Figure 4 shows the training data plot.

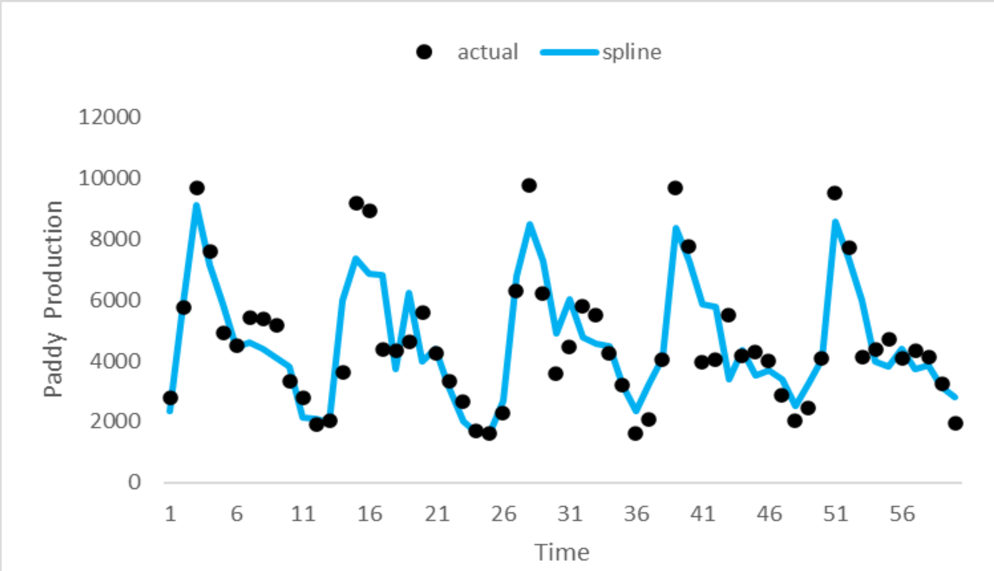


Figure 7. Plot of estimation result training data by STSRM based on LSSE

Based on figure 7, it can be seen that the plot produced by STSRM based on LSSE for training data described by the blue line is close to the actual value described by the black point. The same result is also explained in the plot of the STSRM based on LSSE for testing data depicted in figure 8.

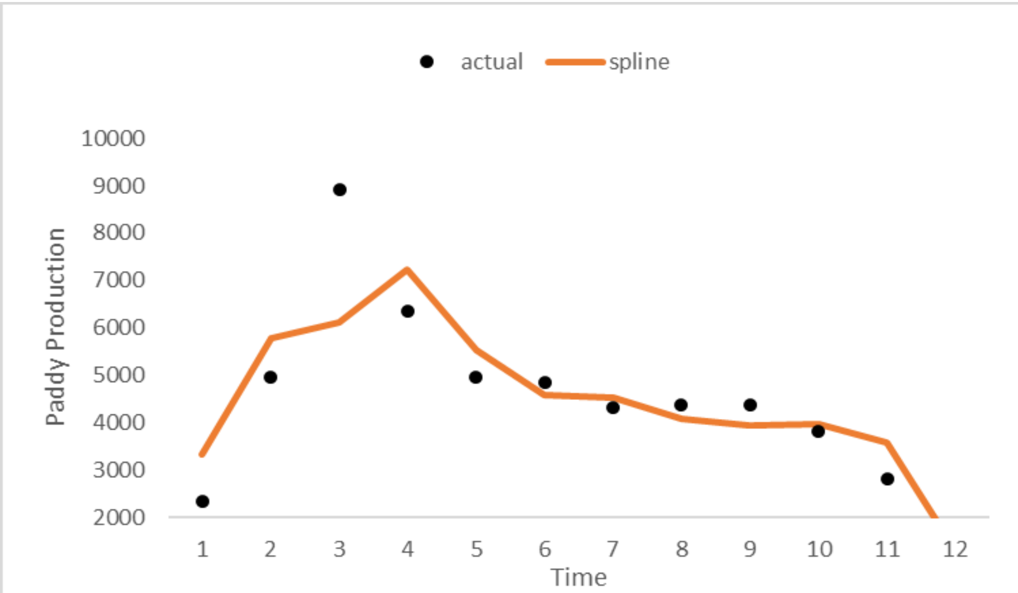


Figure 8. Plot of estimation result testing data by STSRM based on LSSE

The goodness of fit of the model in this study was done by calculating MAPE. The results of the MAPE value calculation for training, testing, and overall data are explained in table 4:

Table 4. MAPE value of Estimation using STSRM based on LSSE

	MAPE	Criteria
training	18,82 %	good prediction
testing	17,28 %	good prediction
overall	18,05 %	good prediction

https://doi.org/10.56294/dm2025527

The training data is 60 observations or 83,3 % of the total data used in this research. Based on the MAPE criteria, the MAPE value estimated from the training data is 18,82 %, less than 20 %, which means the model is good prediction. MAPE value on testing data (16,67 % of total data) is 17,28 % also good prediction. For overall data, the MAPE value is 18,05 %, which means that the model is a good prediction. Hasil dari penelitian ini menjelaskan bahwa model regresi semiparametric based on LSSE dapat menjadi suatu alat analisis untuk melakukan prediksi seperti yang penelitian yang dilakukan oleh Marbun et al.[(49)]. So, the model produced using STSRM based on LSSE in this research can be used as a reference for predicting paddy production in Indonesia. The government can determine policies that can be taken to increase paddy production in Indonesia in order to improve community welfare.

CONCLUSIONS

The prediction of paddy production in Indonesia using STSRM based on LSSE is a good prediction with MAPE value of 18,05 % is less than 20 %. Based on the estimated model, the relationship between paddy production and paddy production in the previous period lag-1 is positive. The variable potential area of crop failure has a negative influence on paddy production, while the area of generative area has a positive influence on paddy production. In order to meet food needs, the potential for crop failure can reduce paddy production, so the potential for crop failure can be a special concern for the Indonesian government. The generative phase is an important moment for rice plants in the process of producing quality paddy. Increasing the area of the generative phase can increase the quality and quantity of paddy production. Prediction of paddy production in Indonesia can be used as support for one of the SDGs points, Zero Hunger.

BIBLIOGRAPHIC REFERENCES

1. Kumar M, Choubey VK, Raut RD, Jagtap S. Enablers to achieve zero hunger through IoT and blockchain technology and transform the green food supply chain systems. J Clean Prod. 2023 Jun;405:136894. doi: 10.1016/j.jclepro.2023.136894.

2. Mumuni S, Aleer MJ. Zero Hunger by 2030 – Are we on track? Climate variability and change impacts on food security in Africa. Cogent Food Agric. 2023 Dec;9(1):2171830. doi: 10.1080/23311932.2023.2171830.

3. Ansari A, Pranesti A, Telaumbanua M, Alam T, Taryono, Wulandari RA, Nugroho BDA, Supriyanta. Evaluating the effect of climate change on rice production in Indonesia using multimodelling approach. Heliyon. 2023 Sep;9(9):e19639. doi: 10.1016/j.heliyon.2023.e19639.

4. Pirngadi RS, Rahmawaty. The impact of flooding on rice production in the Krueng Kluet Watershed, Aceh Province, Indonesia. IOP Conf Ser Earth Environ Sci. 2022 Jun;977(1):012113. doi: 10.1088/1755-1315/977/1/012113.

5. Rachman MI, Nuryartono N, Arifin B, Bakhtiar T. Optimizing crop insurance strategy as a protection tool from crop failure, due to climate change through private sector involvement. IOP Conf Ser Earth Environ Sci. 2021 Apr;739(1):012029. doi: 10.1088/1755-1315/739/1/012029.

6. Chen Y, Liu Y, Dong S, Liu J, Wang Y, Hussain S, Wei H, Huo Z, Xu K, Dai Q. Response of rice yield and grain quality to combined nitrogen application rate and planting density in saline area. Agriculture. 2022 Oct;12(11):1788. doi: 10.3390/agriculture12111788.

7. Zhang Z. Parametric regression model for survival data: Weibull regression model as an example. Ann Transl Med. 2016 Dec;4(24):484. doi: 10.21037/atm.2016.08.45.

8. Ansarifar J, Wang L, Archontoulis SV. An interaction regression model for crop yield prediction. Sci Rep. 2021 Sep;11(1):17754. doi: 10.1038/s41598-021-97221-7.

9. Suleiman AA, Suleiman A, Abdullahi UA, Suleiman SA. Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. Biosaf Health. 2021 Feb;3(1):4-7. doi: 10.1016/j.bsheal.2020.09.003.

10. Eubank RL. Nonparametric Regression and Spline Smoothing. 0th ed. CRC Press; 1999. doi: 10.1201/9781482273144.

11. Caron A, Baio G, Manolopoulou I. Estimating individual treatment effects using non-parametric regression models: a review. J R Stat Soc Ser A Stat Soc. 2022 Jul;185(3):1115-1149. doi: 10.1111/rssa.12824.

12. Lestari B, Chamidah N, Aydin D, Yilmaz E. Reproducing kernel Hilbert space approach to multiresponse smoothing spline regression function. Symmetry. 2022 Oct;14(11):2227. doi: 10.3390/sym14112227.

13. Yan J. Almost sure convergence for weighted sums of WNOD random variables and its applications to nonparametric regression models. Commun Stat - Theory Methods. 2018 Aug;47(16):3893-3909. doi: 10.1080/03610926.2017.1364390.

14. Hidayati L, Chamidah N, Budiantara IN. Spline truncated estimator in multiresponse semiparametric regression model for computer-based national exam in West Nusa Tenggara. IOP Conf Ser Mater Sci Eng. 2019 Jun;546(5):052029. doi: 10.1088/1757-899X/546/5/052029.

15. Utami TW, Lahdji A. Modeling of local polynomial kernel nonparametric regression for COVID daily cases in Semarang City, Indonesia. MEDIA Stat. 2022 Jan;14(2):206-215. doi: 10.14710/medstat.14.2.206-215.

16. Zhou B, Ma X, Luo Y, Yang D. Wind power prediction based on LSTM networks and nonparametric kernel density estimation. IEEE Access. 2019;7:165279-165292. doi: 10.1109/ACCESS.2019.2952555.

17. Ana E, Chamidah N, Andriani P, Lestari B. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression. J Phys Conf Ser. 2019 Dec;1397(1):012067. doi: 10.1088/1742-6596/1397/1/012067.

18. Kong L, Song X, Wang X. Nonparametric regression for interval-valued data based on local linear smoothing approach. Neurocomputing. 2022 Aug;501:834-843. doi: 10.1016/j.neucom.2022.06.073.

19. Sara L. Nonparametric local linear regression estimation for censored data and functional regressors. J Korean Stat Soc. 2022 Mar;51(1):25-46. doi: 10.1007/s42952-020-00080-7.

20. Nidhomuddin N, Chamidah N, Kurniawan A. Confidence interval of the parameter on multipredictor biresponse longitudinal data analysis using local linear estimator for modeling of case increase and case fatality rates COVID-19 in Indonesia: A theoretical discussion. Commun Math Biol Neurosci. 2022. doi: 10.28919/cmbn/6900.

21. Fibriyani V, Chamidah N. Prediction of inflation rate in Indonesia using local polynomial estimator for time series data. J Phys Conf Ser. 2021 Feb;1776(1):012065. doi: 10.1088/1742-6596/1776/1/012065.

22. Relan R, Tiels K, Timmermans J-M, Schoukens J. A local polynomial approach to nonparametric estimation of the best linear approximation of lithium-ion battery from multiple datasets. IEEE Control Syst Lett. 2017 Jul;1(1):182-187. doi: 10.1109/LCSYS.2017.2712364.

23. Setyawati M, Chamidah N, Kurniawan A. Modelling Scholastic Aptitude Test of State Islamic Colleges in Indonesia using least square spline estimator in longitudinal semiparametric regression. J Phys Conf Ser. 2021 Feb;1764(1):012077. doi: 10.1088/1742-6596/1764/1/012077.

24. Ramadan A, Chamidah N, Budiantara IN. Modeling the number of confirmed cases of COVID-19 in East Java using negative binomial regression based on least square spline estimator. Presented at: The International Symposium on Biomathematics (SYMOMATH); 2022; Bandung, Indonesia. 2024:040008. doi: 10.1063/5.0225160.

25. Ramadan A, Chamidah N, Budiantara IN. Modelling the number of HIV cases in Indonesia using negative binomial regression based on least square spline estimator. Commun Math Biol Neurosci. 2024. doi: 10.28919/cmbn/8582.

26. Hasanah N, Bahri S, Fitriyani N. The effect of human development index on poverty model in Indonesia using penalized basis spline nonparametric regression. IOP Conf Ser Mater Sci Eng. 2021 Mar;1115(1):012055. doi: 10.1088/1757-899X/1115/1/012055.

27. Chamidah N, Rifada M, Amelia D. A theoretical discussion on modeling the number of COVID-19 death cases using penalized spline negative binomial regression approach. Commun Math Biol Neurosci. 2022. doi: 10.28919/cmbn/7518.

28. Utami TW, Fauzi F, Yuliyanto E. Statistical downscaling using regression nonparametric of Fourier series-polynomial local of climate change. BAREKENG J Ilmu Mat Dan Terap. 2023 Sep;17(3):1411-1418. doi: 10.30598/barekengvol17iss3pp1411-1418.

29. Amri IF, Chamidah N, Saifudin T, Purwanto D, Fadlurohman A, Ningrum AF, et al. Prediction of extreme weather using nonparametric regression approach with Fourier series estimators. Data Metadata. 2024 Jan;3:319. doi: 10.56294/dm2024319.

30. Sharma C, Jha P, Anand P, Sharma M. Prediction and time series analysis of wheat, rice and maize yields using ARIMA models. In: 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS); 2023 Nov; Tashkent, Uzbekistan. Tashkent: IEEE; 2023. p. 1055-61. doi: 10.1109/ICTACS59847.2023.10389919.

31. Tripathi R, Nayak AK, Raja R, Shahid M, Kumar A, Mohanty S, et al. Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models. Adv Agric. 2014;2014:1-9. doi: 10.1155/2014/621313.

32. Kasthuri V, Selvakumar S. Forecasting foodgrains production using ARIMA model and neural network. Am J Neural Netw Appl. 2021;7(2):30. doi: 10.11648/j.ajnna.20210702.12.

33. Ahmed SE, Aydin D, Yilmaz E. Semiparametric time-series model using local polynomial: An application on the effects of financial risk factors on crop yield. J Risk Financ Manag. 2022 Mar;15(3):141. doi: 10.3390/jrfm15030141.

34. Fibriyani V, Chamidah N, Saifudin T. Modeling case fatality rate of COVID-19 in Indonesia using time series semiparametric regression based on local polynomial estimator. Commun Math Biol Neurosci. 2024. doi: 10.28919/cmbn/8379.

35. Li D, Chen J, Gao J. Non-parametric time-varying coefficient panel data models with fixed effects. Econom J. 2011 Oct;14(3):387-408. doi: 10.1111/j.1368-423X.2011.00350.x.

36. Vilar-Fernández J, Cao R. Nonparametric forecasting in time series - A comparative study. Commun Stat Simul Comput. 2007;36(2):311-34. doi: 10.1080/03610910601158377.

37. Gao J, Gijbels I. Bandwidth selection in nonparametric kernel testing. J Am Stat Assoc. 2008 Dec;103(484):1584-94. doi: 10.1198/016214508000000968.

38. Yan Y, Gao J, Peng B. A class of time-varying vector moving average models: Nonparametric kernel estimation and application. 2020. Available from: https://doi.org/10.48550/ARXIV.2010.01492

39. Chen SX, Gao J, Tang CY. A test for model specification of diffusion processes. Ann Stat. 2008 Feb;36(1):25-46. doi: 10.1214/009053607000000659.

40. Wang Y, Yu Z. A kernel regression model for panel count data with nonparametric covariate functions. Biometrics. 2022 Jun;78(2):586-97. doi: 10.1111/biom.13440.

41. Alwi W, Irwan, Ermawati, Nurman TA, Hernawati, Ibnas R. Modeling of poverty level in South Sulawesi Province through spline nonparametric regression approach. Commun Math Biol Neurosci. 2023. doi: 10.28919/cmbn/7946.

42. Maharani M, Saputro DRS. Generalized cross validation (GCV) in smoothing spline nonparametric regression models. J Phys Conf Ser. 2021 Mar;1808(1):012053. doi: 10.1088/1742-6596/1808/1/012053.

43. Lestari B, Fatmawati, Budiantara IN, Chamidah N. Estimation of regression function in multi-response nonparametric regression model using smoothing spline and kernel estimators. J Phys Conf Ser. 2018 Sep;1097:012091. doi: 10.1088/1742-6596/1097/1/012091.

44. Mariati NPAM, Budiantara IN, Ratnasari V. The application of mixed smoothing spline and Fourier series model in nonparametric regression. Symmetry. 2021 Nov;13(11):2094. doi: 10.3390/sym13112094.

45. Montaño Moreno JJ, Palmer Pol A, Sesé Abad A. Using the R-MAPE index as a resistant measure of forecast accuracy. Psicothema. 2013 Nov;25(4):500-6. doi: 10.7334/psicothema2013.23.

46. Agustiarini S. Study of developing models of crop failure risk information. J Pijar MIPA. 2024 Jan;19(1):136-44.

47. Kadyrova FZ, Klimova LR, Imatullina GI, Safin RI, Zakharov VG. The effectiveness of using new morphobiotypes of buckwheat in breeding for arid conditions of the Middle Volga region. Vestnik of Kazan State Agrarian University. 2024;2:12-17. doi: 10.12737/2073-0462-2024-12-17.

48. Abarajitha P, Rathika S, Avudaithai S, Balasubramaniam P, Ramesh T. Influence of nutrient management on growth, lodging and yield of transplanted rice. Int J Plant Soil Sci. 2022;34(22):519-24.

49. Marbun W, Suparti, Maruddani DAI. Modelling and estimation of solar energy output in Indonesia. J Phys Conf Ser. 2020;1524:012096.

FINANCING
This research was funded by Education Funding Institution (LPDP) and the Education Funding Serving Center (BPPT), grant number: 00198/J5.2.3./BPI.06/9/2022

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: Any Tsalasatul Fitriyah, Nur Chamidah.
Data curation: Any Tsalasatul Fitriyah, Nur Chamidah, Toha Saifudin.
Formal analysis: Nur Chamidah, Any Tsalasatul Fitriyah.
Research: Any Tsalasatul Fitriyah.
Methodology: Any Tsalasatul Fitriyah.
Project management: Any Tsalasatul Fitriyah.
Resources: Any Tsalasatul Fitriyah.
Software: Toha Saifudin.
Supervision: Nur Chamidah, Toha Saifudin.
Validation: Nur Chamidah, Toha Saifudin.
Display: Any Tsalasatul Fitriyah.
Drafting - original draft: Any Tsalasatul Fitriyah.
Writing - proofreading and editing: Nur Chamidah.