International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 6 June - 2025, Pages: 173-186

Modelling of Rice Production in Indonesia Using a Semi-Parametric Regression Approach on Panel Data

Suliyanto[1], Dita Amelia[2], Gabriella Agnes Budijono[3], Khansa Azizah[4], Raja Van Den Bosch Sihotang[5]

[1,2,3,4,5]Statistics Study Program, Department of Mathematics, Faculty of Science and Technology
Airlangga University
Surabaya, Indonesia
Corresponding author: suliyanto@fst.unair.ac.id

Abstract: *This study models rice production dynamics across Indonesia's 34 provinces (2022-2024) using a semi-parametric panel data approach with Generalized Additive Models (GAM). Rice (Oryza sativa L.) is Indonesia's primary staple crop, critical for national food security, yet production declined by 2.05% between 2022 (54.75 million tons GKG) and 2023 (53.63 million tons GKG). To address complex, potentially non-linear relationships, longitudinal data from the Central Bureau of Statistics (BPS) modeled provincial rice production (Y, tons) against key predictors: productivity (X_1, quintal/ha), percentage of informal agricultural labor (X_2, %), average temperature (X_3, °C), average humidity (X_4, %), and rainfall (X_5, mm). Descriptive analysis revealed significant regional disparities: average production was 1,582,867 tons (SD = 2,609,910 tons), with Java provinces dominating nationally. A 2024 decline occurred in most provinces. Productivity and informal labor percentages decreased in 2024, while humidity peaked (87.76%) and rainfall showed high volatility (max 4950.5 mm in 2024). Correlation analysis indicated positive links between production and productivity (r = 0.55) and labor (r = 0.19), but negative associations with temperature (r = -0.21) and humidity (r = -0.32). The optimal GAM, selected via the lowest Quasi Information Criterion (QIC), utilized an Independence working correlation structure. Results identified significant non-linear effects of productivity, informal labor, and temperature on production. Humidity exerted a significant negative linear effect (coefficient: -196,600), while rainfall had a positive linear effect (coefficient: +397.4). This underscores the complex interplay of socio-economic and climatic factors, particularly the constraining role of high humidity and the beneficial, yet variable, role of rainfall. The findings support targeted, climate-resilient agricultural policies. Future research should incorporate variables like harvested area and irrigation and explore Generalized Additive Models (GAM) for extended timeframes.*

Keywords: Rice Production; Generalized Additive Model; Cubic Spline Regression; Spline Truncated; Semi-Parametric Panel Regression

1. INTRODUCTION

Rice (Oryza sativa L.) is the primary food crop in Indonesia, with its main product, rice grains, serving as a staple food for over 270 million people. The per capita rice consumption in Indonesia reaches 111.58 kg annually (Indonesian Ministry of Agriculture, 2019), making rice a strategic commodity for national food security. According to data from the Central Statistics Agency (BPS), rice production in 2023 is estimated at 53.63 million tons of milled dry grain (GKG), a 2.05% decrease compared to the 2022 production of 54.75 million tons GKG. When converted to consumable rice, the 2023 rice production is projected to be approximately 30.90 million tons, a 2.05% decline from the 2022 production of 31.54 million tons (BPS, 2023). Meanwhile, in 2022 Indonesia's Global Food Security Index (GFSI) score was recorded at 60.2 points. Based on the Economist Impact report, this score increased by 1.7 percent compared to the previous year which amounted to 59.2 points (Maun & Kaleka, 2024). The index score puts Indonesia's food security in 2022 in the moderate category (score 55-69.9 points). Indonesia is ranked 63rd out of 113 countries. Although Indonesia's food security index shows an increase, rice production in Indonesia is still decreasing. Therefore, challenges in rice production remain a major concern, mainly due to various factors such as productivity, labor, as well as climatic and environmental conditions that may affect the stability of national food production in the future. The decline in rice production is influenced by interrelated factors, including productivity, labor, and environmental conditions. Research by Rochdiani (2022) highlights that national rice production is predominantly contributed by wetland rice productivity, while lower upland rice productivity contributes minimally to total output. Climate factors, such as temperature and humidity, also influence regional rice yields. Rainfall significantly impacts rice production and productivity. Optimal rainfall enhances production, but excessive rainfall can cause flooding, while water shortages during growth phases reduce harvest yields. Studies in regions such as Percut Sei Tuan District, West Kalimantan, and Winong District, Pati Regency, demonstrate that rainfall variability affects rice productivity, emphasizing the need for improved water resource management to stabilize agricultural outputs (Chaniago, 2023; Aditya et al., 2021; Qudriyah et al., 2022). In addition to environmental factors, informal labor in the agricultural sector plays a crucial role in rice production. Most informal workers are small-scale farmers who face limited access to technology and financial resources, resulting in low productivity. In Central Java, the agricultural sector relies heavily on informal labor, such as unpaid family workers and casual farm laborers, within subsistence systems. Low

education levels further exacerbate agricultural productivity challenges (Faiz, 2021). To address these challenges, a comprehensive analytical approach is essential. Longitudinal data analysis, which observes production trends over time, is critical for identifying patterns and factors influencing rice production. However, relationships between rice production and predictor variables are often nonlinear and complex, rendering conventional statistical methods like linear regression insufficient (Wood, 2017). The Generalized Additive Model (GAM) offers a solution to overcome these limitations. GAM enables the modeling of nonlinear relationships between response and predictor variables through flexible smoothing functions (Wood, 2017). Results revealed nonlinear relationships between rainfall and rice production, demonstrating GAM's superiority over conventional linear regression. The model provided accurate and consistent three-year forecasts, establishing its relevance for food security planning. Similarly, GAM was used to model rice productivity in Banten Province, revealing significant nonlinear effects of cropping systems, pest attacks, climate impacts, water availability, and fertilizer use on productivity, while seed quantity showed no significance. GAM proved valid for modeling nonlinear relationships, supporting evidence-based strategies to enhance productivity (Manurung et al., 2020). Beyond accurate estimations, GAM also excels in facilitating data-driven policymaking (Aly, 2021). By identifying nonlinear relationships between production variables and determinants, GAM enables precise decision-making in agricultural input allocation, climate impact mitigation, and farmer support policies (Matematika et al., 2023). This study aligns with Sustainable Development Goal (SDG) 2: Zero Hunger, which aims to achieve food security, improve nutrition, and promote sustainable agriculture. The research seeks to model rice production in Indonesia using GAM, incorporating productivity, informal labor, temperature, humidity, and rainfall. While prior studies have applied GAM to regional rice production analyses (e.g., Central Sulawesi and Banten), comprehensive national-scale longitudinal studies across provinces remain limited. Thus, further research is needed to capture Indonesia-wide production dynamics, providing a scientific basis for national food security policies. By leveraging longitudinal data from 34 provinces between 2022 and 2024, this study aims to deliver a holistic overview of national rice production trends. The findings are expected to contribute to statistical science through GAM applications and offer data-driven policy recommendations to enhance productivity and national food resilience.

2. METHODOLOGY

2.1 Assumption Test

The test assumptions discuss the testing criteria for linear variable testing and multicolinearity of variables.

2.1.1 Linearity Test

The linearity test is performed to determine whether a linear relationship exists between one variable and another. Generally, linearity testing is used as an analytical tool when data analysis is performed using simple linear regression or multiple linear regression (Widana & Muliani, 2020). Curve estimation, also known as residual analysis, bivariate plots, and linearity tests, are methods that can be used to verify linear relationships. By default, linear statistics will create linear regression or linear correlation based on variable-by-variable analysis of verified linear data (Widana & Muliani, 2020) The linearity test is performed to determine whether a linear relationship exists between one variable and another. Generally, linearity testing is used as an analytical tool when data analysis is performed using simple linear regression or multiple linear regression (Widana & Muliani, 2020). Curve estimation, also known as residual analysis, bivariate plots, and linearity tests, are methods that can be used to verify linear relationships. By default, linear statistics will create linear regression or linear correlation based on variable-by-variable analysis of verified linear data (Widana & Muliani, 2020).

2.1.2 Multicolinearity Test

Multicollinearity is a condition where there is correlation between independent variables or where independent variables are not mutually independent. The measure (quality) that can be used to detect multicollinearity is the variance inflation factor (VIF). VIF is used as a criterion to detect multicollinearity in linear regression involving more than two independent variables. A VIF value greater than 10 indicates a serious multicollinearity problem (Ryan, 2008). The VIF value for the j-th predictor variable is calculated using the formula:

$$VIF_j = \frac{1}{1-R_j^2} \quad (1)$$

with R_j^2 is the coefficient of determination from the regression results of predictor variable X_j against other predictor variables ., where $j = 1,2,\dots,p$.

2.2 Panel Data

Panel Data is a special case of longitudinal data in which each individual is observed periodically over time that is observed and measured repeatedly at specific intervals. The advantage of longitudinal data is its ability to identify the effect of measurement time on responses. Panel Data assumes that subjects are independent of one another, but observations within the same subject are not independent because they tend to be correlated (Hulin & Jin-Ting, 2006). The main characteristics and scope of panel research include:

a. Data is collected for each variable at two or more specific time periods.
b. The subjects or cases being analyzed are the same.
c. The analysis involves comparing the same data within a single period.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 6 June - 2025, Pages: 173-186

2.3 Spline Truncated

Spline regression is a commonly used technique in nonparametric regression. Splines effectively handle fluctuating data behavior across specific subintervals (Budiantara, 2005). Spline truncated regression involves segmented polynomials that allow for flexibility and the handling of changing data patterns within specific intervals (Häardie et al., 2000). In it there is a specification of the correlation between two different responses, or often referred to as working correlation, which is denoted as, namely a × matrix that depends on a parameter α so that the correlation matrix is often written as ($\boldsymbol{\alpha}$) (Danardono, 2018). Knot points are referenced in spline regression, marking the locations where the function's behavior changes across intervals (Härdle, 1994). The truncated power basis function used in spline truncated nonparametric regression with polynomial degree R and d knots points are $\{(z-\tau_1)_+^R, \dots, (z-\tau_d)_+^R\}$. By assuming each predictor variables are not correlated can be written as an additive model (Ramli, 2024). On panel data with q predictor variables, where subject $i = 1, 2, .., N$ and each subject observed $t = 1, 2, \dots, T$ times, the multi-predictor truncated spline regression model is generally an extension of the single predictor truncated spline model for panel data. This model can be written as follows:

$$y_{it} = \sum_{s=1}^{q} f_s(z_{sit}) + \varepsilon_{it} \;\; ; \;\; i = 1, 2, \dots, N; \;\; t = 1, 2, \dots, T \tag{2}$$

Where

$$f_s(z_{sit}) = \sum_{r=0}^{R} \alpha_{sr} z_{sit}^r + \sum_{k=1}^{d} \theta_{sk} (z_{sit} - \tau_{sik})_+^R \tag{3}$$

$$(z_{sit} - \tau_{sik})_+^R = \begin{cases} (z_{sit} - \tau_{sik})^R, & z_{sit} > \tau_{sik} \\ 0 & , \; z_{sit} \leq \tau_{sik} \end{cases} \tag{4}$$

The spline truncated function in equation (3) can be described using the predictor variable z_s; $s = 1, 2, \dots, q$, where each individual $i = 1, 2, \dots, N$ is observed for T time periods. Furthermore, based on equation (2) is obtained a truncated spline nonparametric regression model on panel data as:

$$\boldsymbol{y} = \boldsymbol{V\delta} + \boldsymbol{\varepsilon} \tag{5}$$

Where $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_N} \end{pmatrix}, \boldsymbol{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix},$

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V_{11}} & \boldsymbol{V_{21}} & \dots & \boldsymbol{V_{q1}} \\ \boldsymbol{V_{12}} & \boldsymbol{V_{22}} & \dots & \boldsymbol{V_{q2}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{V_{1N}} & \boldsymbol{V_{2N}} & \dots & \boldsymbol{V_{qN}} \end{pmatrix}, \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta_1} \\ \boldsymbol{\delta_2} \\ \vdots \\ \boldsymbol{\delta_q} \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_N} \end{pmatrix},$$

$$\boldsymbol{V_{si}} = \begin{bmatrix} z_{si1} & z_{si1}^2 & \cdots & z_{si1}^R & (z_{si1} - \tau_{si1})_+^R & \cdots & (z_{si1} - \tau_{sid})_+^R \\ z_{si2} & z_{si2}^2 & \cdots & z_{si2}^R & (z_{si2} - \tau_{si1})_+^R & \cdots & (z_{si2} - \tau_{sid})_+^R \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_{siT} & z_{siT}^2 & \cdots & z_{siT}^R & (z_{siT} - \tau_{si1})_+^R & \cdots & (z_{siT} - \tau_{sid})_+^R \end{bmatrix}$$

, and $\boldsymbol{\delta_s} = [\alpha_{s1} \;\; \alpha_{s2} \;\; \alpha_{s3} \;\; \cdots \;\; \alpha_{sR} \;\; \theta_{s1} \;\; \cdots \;\; \theta_{sd}]'$ for $s = 1, 2, \dots, q$

2.4 Selection of Optimal Knot Points

An effective spline regression model relies on the precise selection of knot points, which mark where the function's behavior shifts (Eubank, 1999). A widely used technique for identifying optimal knot points is the Generalized Cross Validation (GCV) method, which exhibits asymptotic optimality, meaning it approaches the best possible performance as the dataset grows infinitely large. For longitudinal data, the optimal truncated spline regression model is determined by selecting the knot configuration that yields the smallest GCV value (Wahba, 1990). The GCV formula:

$$GCV(d) = \frac{MSE(d)}{\left[\frac{1}{NT} trace(\boldsymbol{I} - \boldsymbol{H}(d))\right]^2} \tag{6}$$

where $\boldsymbol{I}$ is the identity matrix, n is the number of observations, d represents the knot points, $MSE(d) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it})^2$, and $\boldsymbol{H}(d) = \boldsymbol{V}(\boldsymbol{V'V})^{-1}\boldsymbol{V'}$.

2.5 Semi-parametric Generalized Additive Model

Cubic spline regression is a special form of truncated spline regression, where the polynomial order (R) used is 3 (Imaslihkah et al., 2013). This technique is used to model non-linear relationships between response variables and predictors without having to specify the shape of the function a priori. The cubic spline divides the predictor space into intervals separated by knots. In each interval, a three-order polynomial is used which is arranged such that the regression function remains continuous, including its first and second derivatives, at all knots. This provides great flexibility in modeling, while maintaining the smoothness of the estimated curve, so the model for semi-parametric GAM is as follows:

$$\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{V\delta} + \boldsymbol{\varepsilon} \tag{7}$$

Where $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_N} \end{pmatrix}, \boldsymbol{x_i} = \begin{bmatrix} 1 & x_{1i1} & x_{2i1} & \cdots & x_{pi1} \\ 1 & x_{1i2} & x_{2i2} & \cdots & x_{pi2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1iT} & x_{2iT} & \cdots & x_{piT} \end{bmatrix}$, and

$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$

3. DATA DESCRIPTION

The data used in this study are secondary data obtained from the Indonesian Central Bureau of Statistics (BPS, 2022; BPS, 2023; BPS, 2024). The observation unit covers all 34 provinces in Indonesia from 2022 to 2024. The variables consist of predictor variables (X & Z) and response variables (Y).

Table 1: Research Variables with scale and unit of Predictor and Response

Variable	Name of Variable	Unit	Scale
Y	Rice Production	Tons	Ratio
X_1	Average Humidity	Percent	Ratio
X_2	Rainfall	Milimeters	Ratio
Z_1	Productivity	Quintals/Hectare	Ratio
Z_2	Precentage of Informal Labor in the Agricultural Sector	Percent	Ratio
Z_3	Average Temperature	°C	Interval

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

4. Results and Discussion

4.1 Descriptive Statistics

To model rice paddy production in 34 provinces in Indonesia in the period 2022 - 2024 along with its influencing factors, including rice paddy productivity, percentage of labor in agriculture, average temperature, average air humidity, and amount of rainfall, an overview with descriptive analysis is presented as follows:

Table 1: Descriptive Statistics.

Variable	Mean	StDev	Minimum	Maximum
Y	1582867	2609910	305	9710661
X_1	81.349	3.304	73.660	87.760
X_2	2566.4	805.3	830.8	4950.5
Z_1	45.993	8.270	226.920	62.070
Z_2	85.97	10.30	47	99.38
Z_3	27.676	0.869	24.990	29.410

Based on Table 2 shows that rice production (Y) has an average value of 1,582,867 tons with a standard deviation of 2,609,910. The minimum production value is 305 tons and the maximum reaches 9,710,661 tons, indicating a variation in production between provinces. Humidity (X_1) has an average value of 81.349%. Meanwhile, rainfall (X_2) showed an average of 2,566.4 mm with a standard deviation of 805.3 mm, indicating variations in rainfall between regions. Rice productivity (Z_1) has an average of 45,993 quintals per hectare with a standard deviation of 8,270, reflecting the differences in productivity between regions. The percentage of labor in the agricultural sector (Z_2) shows a high average of 85.97%, indicating that agriculture is the main sector of employment in many provinces. In addition, it can be noted that the average temperature value (Z_3) achieved is 27.676°C.

4.2 Assumption Test

In semi-parametric panel data analysis with GAM, assumption testing is an important step to ensure the validity of the model built. Assumption tests are conducted to determine whether the model has met the necessary statistical requirements, so that the estimation results can be interpreted appropriately. This test includes linearity test, multicollinearity test, and residual normality test.

4.2.1 Linearity Test

Determining whether a predictor variable belongs to the parametric or nonparametric component can be done through the Ramsey RESET linearity test. This test is conducted based on the following hypothesis: H_0: Linear Model; H_1: Non-linear Model. The results of linearity testing for each predictor variable on the response variable are obtained as follows.

Table 2: Linearity Test between Predictor Variables and Response Variables.

Relationship	P-Value	Decision
X_1 and Y	0.4	Linear Model
X_2 and Y	0.9	Linear Model
Z_1 and Y	0.0009	Non-linear Model
Z_2 and Y	0.005	Non-linear Model
Z_3 and Y	0.007	Non-linear Model

Based on Table 3, it can be seen that variables X_1 and X_2 have a p-value > 0.05 which indicates that there is no deviation of the linearity assumption, so they can be modeled linearly. While the other predictor variables, Z_1, Z_2, and Z_3, have a p-value < 0.05, indicating that the relationship between each predictor variable and the response variable Y is not linear.

4.2.2 Multicollinearity Test for Each Predictor Variable Linear

Multicollinearity testing is carried out to determine whether there is a high correlation between predictor variables in the regression model, which can cause instability in the coefficient estimates and make it difficult to interpret the model. One commonly used indicator is the Variance Inflation Factor (VIF) value, where VIF values > 10 indicate potential multicollinearity between predictor variables. The VIF values for each variable can be seen in Table 4 below.

Table 3: Multicollinearity Test for Each Predictor Variable Linear.

Variable	VIF
X_1	1.80
X_2	1.14

Based on Table 4, all VIF values generated are below 10. This indicates that there is no high correlation between predictor variables in the model, so it can be concluded that multicollinearity problems do not occur in the model.

4.2.3 Residual Normality Test

Residual normality testing is carried out to evaluate whether the residuals from the regression model follow a normal distribution. The statistical test used is Kolmogorov-Smirnov, with the criterion that the data is considered normally distributed if the null hypothesis (H_0) is accepted, namely if the p-value > 0.05. The test results show a p-value of 0.132, so H_0 is accepted and it can be concluded that the residuals in the model are normally distributed.

4.3 Some Common Mistakes

GAM is a semi-parametric method that is flexible in capturing nonlinear relationships between predictor and response variables. In this study, GAM is used to analyze rice production patterns in Indonesia with a focus on the three provinces with the lowest production numbers. The selection of these three provinces aims to see how the GAM model can describe variations in rice paddy production in areas with relatively low production levels and production challenges that differ from the main productive areas. The three provinces with the lowest rice production during the period 2022-2024 based on BPS data are Riau Islands and DKI Jakarta. The analysis was conducted separately in each province to obtain a more specific and accurate model estimation. The estimation process includes determining the number of knots in the spline function, applying the GAM model to each province, and evaluating the model prediction results.

4.3.1 Knot Determination of Spline Function in GAM Model

Determining the number and position of knots on the spline function in the GAM model is an important stage in capturing the non-linear relationship between the predictor and response variables. In this study, the number of knots was determined as three points for each non-linear predictor variable. The knot locations were determined based on the quantiles of the distribution of each predictor variable, namely the 25%, 50%, and 75% quantiles. This approach allows the model to be more adaptive to the actual data structure, especially in areas with high data density. The following are the knot locations used on each non-linear predictor variable.

Table 4: Knot Points on Each Non-linear Predictor Variable.

Province	Year	Variable	1st Knot (25%)	2nd Knot (50%)	3rd Knot (75%)
Riau Islands	2022	Z_1	38.99	47.43	51.87
Riau Islands	2022	Z_2	80.69	90.155	93.65
Riau Islands	2022	Z_3	27.43	28.18	28.59
Riau Islands	2023	Z_1	40.86	47.74	52.03
Riau Islands	2023	Z_2	80.6	89.755	94.11
Riau Islands	2023	Z_3	27.178	27.643	27.94833
Riau Islands	2024	Z_1	40.48	46.07	51.84
Riau Islands	2024	Z_2	77.67	88.41	92.37
Riau Islands	2024	Z_3	26.67	27.597	27.84
DKI Jakarta	2022	Z_1	38.99	47.43	51.87
DKI Jakarta	2022	Z_2	80.69	90.155	93.65
DKI Jakarta	2022	Z_3	27.43	28.18	28.59
DKI Jakarta	2023	Z_1	40.86	47.74	52.03

Province	Year	Variable	1st Knot (25%)	2nd Knot (50%)	3rd Knot (75%)
DKI Jakarta	2023	Z_2	80.6	89.755	94.11
DKI Jakarta	2023	Z_3	27.178	27.643	27.94833
DKI Jakarta	2024	Z_1	40.48	46.07	51.84
DKI Jakarta	2024	Z_2	77.67	88.41	92.37
DKI Jakarta	2024	Z_3	26.67	27.597	27.84

4.3.2 Model Estimation in 2 Provinces with the Lowest Rice Production in Indonesia

After the knot determination process was carried out, the next step was to estimate the GAM for each of the two provinces with the lowest rice production, namely Riau Islands and DKI Jakarta. Estimation was conducted separately to capture the characteristics of the relationship between the predictor variables and the amount of rice paddy production in each province. This model allows the identification of non-linear patterns that cannot be captured by the classical linear model.

1. Riau Islands
   Rice production in Riau Islands Province is the lowest in Indonesia which can be caused by the dominance of water areas (98%) that limit agricultural land, infertile soil quality, and low levels of mechanization.

Table 5: Parameter Estimation of GAM Riau Islands in 2022.

Parameters	Value	P-Value
β_0	-6.560e+10	0.0154
β_1	-2.130e+05	0.1881
β_2	4.705e+02	0.2303
α_{11}	1.437e+07	0.5946
α_{12}	-4.069e+05	0.5918
α_{13}	3.790e+03	0.5901
γ_{11}	-6.186e+03	0.6143
γ_{12}	1.745e+04	0.3705
γ_{13}	-5.014e+04	0.0882
α_{21}	-1.184e+07	0.1025
α_{22}	1.841e+05	0.0975
α_{23}	-9.305e+02	0.0929
γ_{21}	7.524e+03	0.0374
γ_{22}	-5.354e+04	0.0279
γ_{23}	9.715e+04	0.0688
α_{31}	7.445e+09	0.0152
α_{32}	-2.809e+08	0.0152
α_{33}	3.530e+06	0.0151
γ_{31}	-1.493e+07	0.0173
γ_{32}	4.287e+07	0.0198
γ_{33}	-5.606e+07	0.0292

Based on Table 6, it can be seen that several parameters in the model show significant p-values below the significance level of 0.05, such as parameters β_0, and several parameters α and γ. This shows that there is a significant non-linear relationship between the predictor variables and the amount of rice production in Riau Islands Province. Meanwhile, some other parameters have a high p-value, which means that their contribution to the model is statistically insignificant. However, since this is an additive model, the overall contribution of the variables can still capture the complex relationship patterns in the data. Based on the parameter estimation results above, the GAM model equation form for Riau Islands Province in 2022 is as follows.

$$
\begin{aligned}
y_{11} = {} & (-6.560 \times 10^{10}) \\
& - (2.130 \times 10^{5} \times 83.58) \\
& + (4.705 \times 10^{2} \times 2590.6) \\
& + (1.437 \times 10^{7} \times 28.24) \\
& - (4.069 \times 10^{5} \times 28.24^{2}) \\
& + (3.790 \times 10^{3} \times 28.24^{3}) \\
& - (6.186 \times 10^{3} \times (28.24 \\
& - 38.99))_{+}^{3} \\
& + (1.745 \times 10^{4} \times (28.24 \\
& - 47.44))_{+}^{3} \\
& - (5.014 \times 10^{4} \times (28.24 \\
& - 51.87))_{+}^{3} \\
& - (1.184 \times 10^{7} \times 80.69) \\
& + (1.841 \times 10^{5} \times 80.69^{2}) \\
& - (9.305 \times 10^{2} \times 80.69^{3}) \\
& + (7.524 \times 10^{3} \times (80.69 \\
& - 80.69))_{+}^{3} \\
& - (5.354 \times 10^{4} \times (80.69 \\
& - 90.155))_{+}^{3} \\
& + (9.715 \times 10^{4} \times (80.69 \\
& - 93.65))_{+}^{3} \\
& + (7.445 \times 10^{9} \times 27.53) \\
& - (2.809 \times 10^{8} \times 27.53^{2}) \\
& + (3.530 \times 10^{6} \times 27.53^{3}) \\
& - (1.493 \times 10^{7} \times (27.53 \\
& - 27.43))_{+}^{3} \\
& + (4.287 \times 10^{7} \times (27.53 \\
& - 28.18))_{+}^{3} \\
& - (5.606 \times 10^{7} \times (27.53 \\
& - 28.59))_{+}^{3}
\end{aligned} \tag{8}
$$

The GAM model estimated for Riau Islands Province in 2022 produced a GCV value = 5.344537×10^{12} and Adjusted R-squared = 0.7887. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

Table 6: Parameter Estimation of GAM Riau Islands in 2023.

Parameters	Value	P-Value
β_0	9.063e+10	0.2398
β_1	-4.139e+05	0.0709
β_2	6.829e+02	0.2673
α_{11}	-1.192e+07	0.3785
α_{12}	3.217e+05	0.3898
α_{13}	-2.879e+03	0.3989
γ_{11}	6.843e+03	0.4326
γ_{12}	2.570e+03	0.8765
γ_{13}	-3.240e+04	0.0973
α_{21}	-3.234e+06	0.8091
α_{22}	4.395e+04	0.8122
α_{23}	-2.069e+02	0.8069
γ_{21}	3.640e+03	0.2725
γ_{22}	-2.716e+04	0.1289
γ_{23}	2.146e+04	0.7796
α_{31}	-1.021e+10	0.2396
α_{32}	3.842e+08	0.2373
α_{33}	-4.820e+06	0.2351
γ_{31}	2.615e+07	0.1221
γ_{32}	-6.047e+07	0.0845
γ_{33}	4.707e+07	0.0666

Based on Table 7, it can be seen that most of the parameters in the model have relatively high p-values above the 0.05 significance level, making them statistically insignificant. However, there are some parameters that are close to significance, such as γ_{13} ($p = 0.0973$), γ_{33} ($p = 0.0666$), and β_1 ($p = 0.0709$), which indicates that the nonlinear effects of these variables on the amount of rice production in Riau Islands Province may be substantively important. Although not individually significant, in the context of an additive model, the combined contribution of all components can still describe the complex pattern of relationships between predictor and response variables. Therefore, the interpretation of the model should pay attention to the overall additive structure, not just the significance of each parameter. Based on the parameter estimation results above, the GAM model equation form for Riau Islands Province in 2023 can be written as follows.

$$
\begin{aligned}
y_{12} = {} & (9.063 \times 10^{10}) \\
& - (4.139 \times 10^{5} \times 83.03) \\
& + (6.829 \times 10^{2} \\
& \times 3085.017) \\
& - (1.192 \times 10^{7} \times 28.11) \\
& + (3.217 \times 10^{5} \times 28.11^{2}) \\
& - (2.879 \times 10^{3} \times 28.11^{3}) \\
& + (6.843 \times 10^{3} \times (28.11 \\
& - 40.86))_{+}^{3} \\
& + (2.570 \times 10^{3} \times (28.11 \\
& - 47.74))_{+}^{3} \\
& - (3.240 \times 10^{4} \times (28.11 \\
& - 52.03))_{+}^{3} \\
& - (3.234 \times 10^{6} \times 79.98) \\
& + (4.395 \times 10^{4} \times 79.98^{2}) \\
& - (2.069 \times 10^{2} \times 79.98^{3}) \\
& + (3.640 \times 10^{3} \times (79.98 \\
& - 80.6))_{+}^{3} \\
& - (2.716 \times 10^{4} \times (79.98 \\
& - 89.755))_{+}^{3} \\
& + (2.146 \times 10^{4} \times (79.98 \\
& - 94.11))_{+}^{3} \\
& - (1.021 \times 10^{10} \\
& \times 27.80833) \\
& + (3.842 \times 10^{8} \\
& \times 27.80833^{2}) \\
& - (4.820 \times 10^{6} \\
& \times 27.80833^{3}) \\
& + (2.615 \times 10^{7} \\
& \times (27.80833 - 27.178))_{+}^{3} \\
& - (6.047 \times 10^{7} \\
& \times (27.80833 - 27.643))_{+}^{3} \\
& + (4.707 \times 10^{7} \\
& \times (27.80833 \\
& - 27.94833))_{+}^{3}
\end{aligned} \tag{9}
$$

The GAM model estimated for Riau Islands Province in 2023 produced a GCV value = 2.864042×10^{12} and Adjusted R-squared = 0.8828. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

Table 7: Parameter Estimation of GAM Riau Islands in 2024.

Parameters	Value	P-Value
β_0	-6.799e+10	0.4545
β_1	-2.559e+05	0.2574
β_2	4.248e+02	0.3865
α_{11}	2.944e+06	0.8229
α_{12}	-9.069e+04	0.8031
α_{13}	8.955e+02	0.7875
γ_{11}	-2.966e+03	0.7329
γ_{12}	1.145e+04	0.3696
γ_{13}	-3.564e+04	0.0258
α_{21}	4.059e+07	0.4140
α_{22}	-5.638e+05	0.4131
α_{23}	2.591e+03	0.4130
γ_{21}	-3.865e+03	0.5615
γ_{22}	-5.822e+03	0.8203
γ_{23}	7.855e+03	0.9313
α_{31}	7.602e+09	0.4651
α_{32}	-2.874e+08	0.4676
α_{33}	3.620e+06	0.4701
γ_{31}	-5.712e+06	0.5294
γ_{32}	1.419e+07	0.5434
γ_{33}	-1.488e+07	0.5607

Based on Table 8, most of the parameters in the model have p-values above 0.05, so they are not statistically significant. However, the parameter γ_{13} is significant with a p-value = 0.0258, indicating a nonlinear influence of the variable X_3 on rice production in Riau Islands Province. Although many parameters were not significant, the additive model was still able to capture the overall pattern of complex relationships. Therefore, the interpretation of the model must look at the combined contribution of the variables. The general form of the GAM model for Riau Islands Province in 2024 is as follows.

$$
\begin{aligned}
y_{13} = & -(6.799 \times 10^{10}) \\
& - (2.559 \times 10^{5} \times 85.6) \\
& + (4.248 \times 10^{2} \\
& \times 2737.05) \\
& + (2.944 \times 10^{6} \times 26.92) \\
& - (9.069 \times 10^{4} \times 26.92^{2}) \\
& + (8.955 \times 10^{3} \times 26.92^{3}) \\
& - (2.966 \times 10^{3} \times (26.92 \\
& - 40.48))_{+}^{3} \\
& + (1.145 \times 10^{4} \times (26.92 \\
& - 46.07))_{+}^{3} \\
& - (3.564 \times 10^{4} \times (26.92 \\
& - 51.84))_{+}^{3} \\
& + (4.059 \times 10^{7} \times 69.84) \\
& - (5.638 \times 10^{5} \times 69.84^{2}) \\
& + (2.591 \times 10^{2} \times 69.84^{3}) \qquad (10) \\
& - (3.865 \times 10^{3} \times (69.84 \\
& - 77.67))_{+}^{3} \\
& - (5.822 \times 10^{3} \times (69.84 \\
& - 88.41))_{+}^{3} \\
& + (7.855 \times 10^{3} \times (69.84 \\
& - 92.37))_{+}^{3} \\
& + (7.602 \times 10^{9} \times 27.98) \\
& - (2.874 \times 10^{8} \times 27.98^{2}) \\
& + (3.620 \times 10^{6} \times 27.98^{3}) \\
& - (5.712 \times 10^{7} \times (27.98 \\
& - 26.67))_{+}^{3} \\
& + (1.419 \times 10^{7} \times (27.98 \\
& - 27.597))_{+}^{3} \\
& - (1.488 \times 10^{7} \times (27.98 \\
& - 27.84))_{+}^{3}
\end{aligned}
$$

The GAM model estimated for Riau Islands Province in 2024 produced a GCV value = 4.481186×10^{12} and Adjusted R-squared = 0.804. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

2. DKI Jakarta

DKI Jakarta Province is the second province with the lowest rice production in Indonesia in 2022-2024 due to several main factors, including the limited area of paddy fields which is only around 414 hectares, a decrease in harvest area due to land conversion, and extreme weather disturbances such as floods that damage crops and accelerate the harvest period. In addition, pest attacks also reduce productivity. Rapid urbanization exacerbates the situation by reducing agricultural land.

Table 8: Parameter Estimation of GAM DKI Jakarta in 2022.

Parameters	Value	P-Value
β_0	-6.560e+10	0.0154
β_1	-2.130e+05	0.1881
β_2	4.705e+02	0.2303
α_{11}	1.437e+07	0.5946
α_{12}	-4.069e+05	0.5918
α_{13}	3.790e+03	0.5901
γ_{11}	-6.186e+03	0.6143
γ_{12}	1.745e+04	0.3705
γ_{13}	-5.014e+04	0.0882
α_{21}	-1.184e+07	0.1025
α_{22}	1.841e+05	0.0975
α_{23}	-9.305e+02	0.0929
γ_{21}	7.524e+03	0.0374
γ_{22}	-5.354e+04	0.0279
γ_{23}	9.715e+04	0.0688
α_{31}	7.445e+09	0.0152
α_{32}	-2.809e+08	0.0152
α_{33}	3.530e+06	0.0151
γ_{31}	-1.493e+07	0.0173
γ_{32}	4.287e+07	0.0198
γ_{33}	-5.606e+07	0.0292

Based on Table 9, most of the parameters in the model have p values above 0.05, so they are not statistically significant. However, there are some parameters that are significant, including: γ_{21} ($p = 0.0374$), γ_{22} ($p = 0.0279$), α_{32} ($p = 0.0152$), α_{33} (p = 0.0151), γ_{31} (p = 0.0173), γ_{32} (p = 0.0198), and γ_{33} (p = 0.0292). This indicates that there is a significant nonlinear effect of X2 and X3 variables on rice production in DKI Jakarta Province. Although many other parameters are not individually significant, the additive model is still able to capture the general pattern

of the complex relationship. Therefore, the interpretation of the model should consider the combined contribution between variables. The general form of the Generalized Additive Model (GAM) for DKI Jakarta Province in 2022 reflects the combination of linear and nonlinear effects of the variables.

$$\begin{aligned} y_{21} = {} & (-6.560 \times 10^{10}) \\ & - (2.130 \times 10^{5} \times 77.21) \\ & + (4.705 \times 10^{2} \times 2136.3) \\ & + (1.437 \times 10^{7} \times 48.98) \\ & - (4.069 \times 10^{5} \times 48.98^{2}) \\ & + (3.790 \times 10^{3} \times 48.98^{3}) \\ & - (6.186 \times 10^{3} \times (48.98 \\ & - 38.99))_{+}^{3} \\ & + (1.745 \times 10^{4} \times (48.98 \\ & - 47.43))_{+}^{3} \\ & - (5.014 \times 10^{4} \times (48.98 \\ & - 51.87))_{+}^{3} \\ & - (1.184 \times 10^{7} \times 47) \\ & + (1.841 \times 10^{5} \times 47^{2}) \\ & - (9.305 \times 10^{2} \times 47^{3}) \\ & + (7.524 \times 10^{3} \times (47 \\ & - 80.69))_{+}^{3} \\ & - (5.354 \times 10^{4} \times (47 \\ & - 90.155))_{+}^{3} \\ & + (9.715 \times 10^{4} \times (47 \\ & - 93.65))_{+}^{3} \\ & + (7.445 \times 10^{9} \times 28.47) \\ & - (2.809 \times 10^{8} \times 28.47^{2}) \\ & + (3.530 \times 10^{6} \times 28.47^{3}) \\ & - (1.493 \times 10^{7} \times (28.47 \\ & - 27.43))_{+}^{3} \\ & + (4.287 \times 10^{7} \times (28.47 \\ & - 28.18))_{+}^{3} \\ & - (5.606 \times 10^{7} \times (28.47 \\ & - 28.59))_{+}^{3} \end{aligned} \quad (11)$$

The GAM model estimated for DKI Jakarta Province in 2022 produced a GCV value = 5.344537×10^{12} and Adjusted R-squared = 0.7887. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

Table 9: Parameter Esimation of GAM DKI Jakarta in 2023.

Parameters	Value	P-Value
β_0	9.063e+10	0.2398
β_1	-4.139e+05	0.0709
β_2	6.829e+02	0.2673
α_{11}	-1.192e+07	0.3785
α_{12}	3.217e+05	0.3898
α_{13}	-2.879e+03	0.3989
γ_{11}	6.843e+03	0.4326
γ_{12}	2.570e+03	0.8765
γ_{13}	-3.240e+04	0.0973
α_{21}	-3.234e+06	0.8091
α_{22}	4.395e+04	0.8122
α_{23}	-2.069e+02	0.8069
γ_{21}	3.640e+03	0.2725
γ_{22}	-2.716e+04	0.1289
γ_{23}	2.146e+04	0.7796
α_{31}	-1.021e+10	0.2396
α_{32}	3.842e+08	0.2373
α_{33}	-4.820e+06	0.2351
γ_{31}	2.615e+07	0.1221
γ_{32}	-6.047e+07	0.0845
γ_{33}	4.707e+07	0.0666

Based on Table 10, most of the parameters in the model have relatively high p-values above the 0.05 significance level, indicating that they are statistically insignificant. However, some parameters are close to the significance threshold, such as γ_{13} (p = 0.0973), γ_{33} (p = 0.0666), and β_1 (p = 0.0709), suggesting that the nonlinear effects of these variables on the amount of rice production in DKI Jakarta Province may still be substantively important. Although not individually significant, the overall additive structure of the model allows the combined contributions of all components to capture the complex relationships between the predictor and response variables. Therefore, interpretation should focus on the holistic behavior of the model rather than the statistical significance of individual parameters. Based on the parameter estimation results above, the GAM model

equation form for DKI Jakarta Province in 2023 can be written as follows.

$$\begin{aligned} y_{22} = {} & (9.063 \times 10^{10}) \\ & - (4.139 \times 10^{5} \times 76.125) \\ & + (6.829 \times 10^{2} \times 1551.2) \\ & - (1.192 \times 10^{7} \times 49.26) \\ & + (3.217 \times 10^{5} \times 49.26^{2}) \\ & - (2.879 \times 10^{3} \times 49.26^{3}) \\ & + (6.843 \times 10^{3} \times (49.26 \\ & - 40.86))_{+}^{3} \\ & + (2.570 \times 10^{3} \times (49.26 \\ & - 47.74))_{+}^{3} \\ & - (3.240 \times 10^{4} \times (49.26 \\ & - 52.03))_{+}^{3} \\ & - (3.234 \times 10^{6} \times 59.72) \\ & + (4.395 \times 10^{4} \times 59.72^{2}) \\ & - (2.069 \times 10^{2} \times 59.72^{3}) \\ & + (3.640 \times 10^{3} \times (59.72 \\ & - 80.6))_{+}^{3} \\ & - (2.716 \times 10^{4} \times (59.72 \\ & - 89.755))_{+}^{3} \\ & + (2.146 \times 10^{4} \times (59.72 \\ & - 94.11))_{+}^{3} \\ & - (1.021 \times 10^{10} \\ & \times 29.135) \\ & + (3.842 \times 10^{8} \\ & \times 29.135^{2}) \\ & - (4.820 \times 10^{6} \\ & \times 29.135^{3}) \\ & + (2.615 \times 10^{7} \times (29.135 \\ & - 27.178))_{+}^{3} \\ & - (6.047 \times 10^{7} \times (29.135 \\ & - 27.643))_{+}^{3} \\ & + (4.707 \times 10^{7} \times (29.135 \\ & - 27.94833))_{+}^{3} \end{aligned} \quad (12)$$

The GAM model estimated for DKI Jakarta Province in 2023 produced a GCV value = 2.864042×10^{12} and Adjusted R-squared = 0.8828. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

Table 10 : Parameter Estimation of GAM DKI Jakarta in 2024.

Parameters	Value	P-Value
β_0	-6.799e+10	0.4545
β_1	-2.559e+05	0.2574
β_2	4.248e+02	0.3865
α_{11}	2.944e+06	0.8229
α_{12}	-9.069e+04	0.8031
α_{13}	8.955e+02	0.7875
γ_{11}	-2.966e+03	0.7329
γ_{12}	1.145e+04	0.3696
γ_{13}	-3.564e+04	0.0258
α_{21}	4.059e+07	0.4140
α_{22}	-5.638e+05	0.4131
α_{23}	2.591e+03	0.4130
γ_{21}	-3.865e+03	0.5615
γ_{22}	-5.822e+03	0.8203
γ_{23}	7.855e+03	0.9313
α_{31}	7.602e+09	0.4651
α_{32}	-2.874e+08	0.4676
α_{33}	3.620e+06	0.4701
γ_{31}	-5.712e+06	0.5294
γ_{32}	1.419e+07	0.5434
γ_{33}	-1.488e+07	0.5607

Based on Table 11, most of the parameters in the model have p-values greater than 0.05, indicating that they are not statistically significant. However, the parameter γ_{13} is statistically significant with a p-value of 0.0258, suggesting a nonlinear effect of variable X_3 on rice production in DKI Jakarta Province. Although individual parameters are largely insignificant, the additive model still succeeds in capturing the complex overall pattern between predictors and the response variable. Therefore, interpretation should focus on the collective contribution of all components rather than isolated effects. The general form of the GAM model for DKI Jakarta Province in 2024 is as follows.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 9 Issue 6 June - 2025, Pages: 173-186

$$
\begin{aligned}
y_{23} = &-(6.799 \times 10^{10}) \\
&- (2.559 \times 10^{5} \times 77.78) \\
&+ (4.248 \times 10^{2} \\
&\times 2159.05) \\
&+ (2.944 \times 10^{6} \times 46.29) \\
&- (9.069 \times 10^{4} \times 46.29^{2}) \\
&+ (8.955 \times 10^{3} \times 46.29^{3}) \\
&- (2.966 \times 10^{3} \times (46.29 \\
&- 40.48))_{+}^{3} \\
&+ (1.145 \times 10^{4} \times (46.29 \\
&- 46.07))_{+}^{3} \\
&- (3.564 \times 10^{4} \times (46.29 \\
&- 51.84))_{+}^{3} \\
&+ (4.059 \times 10^{7} \times 74.37) \\
&- (5.638 \times 10^{5} \times 74.37^{2}) \\
&+ (2.591 \times 10^{2} \times 74.37^{3}) \\
&- (3.865 \times 10^{3} \times (74.37 \\
&- 77.67))_{+}^{3} \\
&- (5.822 \times 10^{3} \times (74.37 \\
&- 88.41))_{+}^{3} \\
&+ (7.855 \times 10^{3} \times (74.37 \\
&- 92.37))_{+}^{3} \\
&+ (7.602 \times 10^{9} \times 29.06) \\
&- (2.874 \times 10^{8} \times 29.06^{2}) \\
&+ (3.620 \times 10^{6} \times 29.06^{3}) \\
&- (5.712 \times 10^{7} \times (29.06 \\
&- 26.67))_{+}^{3} \\
&+ (1.419 \times 10^{7} \times (29.06 \\
&- 27.597))_{+}^{3} \\
&- (1.488 \times 10^{7} \times (29.06 \\
&- 27.84))_{+}^{3}
\end{aligned} \quad (13)
$$

The GAM model estimated for Riau Islands Province in 2024 produced a GCV value = 4.481186×10^{12} and Adjusted R-squared = 0.804. The relatively small GCV value and high R-squared value indicate that the model is able to explain the variability of the data quite well and provide stable estimates despite the non-linear pattern in the data.

4.3.3. Prediction Results in 2 Provinces with the Lowest Rice Production in Indonesia

Based on the GAM model in section 4.3.2., rice production predictions were obtained for the three provinces with the lowest rice production in Indonesia, namely Riau Islands and DKI Jakarta, in (8) to (13). These three provinces were selected because their production characteristics are relatively low compared to other provinces, with production even approaching zero in some years. The estimated results were then compared with actual data to assess how accurately the model captures the historical production patterns presented in the Table 12 and Fig.1 below.

Table 12: Comparison of Actual and Predicted Rice Production in The Two Lowest-Yielding Provinces.

Province	Years	Actual Data	Predicted Data
Riau Islands	2022	506.91	162287.1
	2023	324.01	130105.7
	2024	305.09	250396.7
DKI Jakarta	2022	2337.77	17938.48
	2023	2674.28	-109220.4
	2024	2306.54	191991.5

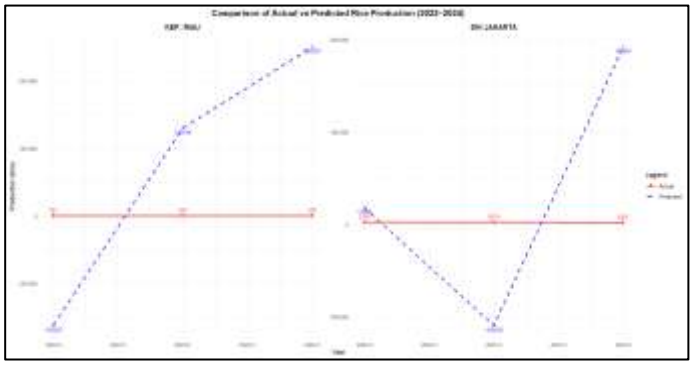

Fig. 1. Comparison of Actual Data and Predicted Rice Production.

Based on Table 12 and Fig. 1, rice production modeling using the Generalized Additive Model (GAM) with a truncated spline approach for the Riau Islands Province, DKI Jakarta, and Papua shows varying results. In Riau Islands, actual production decreased from 506.91 tons (2022) to 305.09 tons (2024), while the model predicts an increase from -162287.1 in 2022 to 250396.7 in 2024. In Jakarta, actual values remained relatively stable between 2300–2600 thousand tons, with model predictions ranging from 17938.48 (2022) to 191991.5 (2024), following a pattern closely aligned with the actual trend, particularly at the beginning and end of the period. Overall, the GAM approach with truncated splines provides an initial overview of rice production trends in each province and can be further optimized to enhance the precision of future estimates.

Significant differences between actual and predicted values, especially in provinces with low and fluctuating production levels such as Riau Islands and DKI Jakarta, could be due to structural limitations and environmental conditions. Limited agricultural land, rapid urbanization, and irregular production patterns pose challenges for the model to accurately capture year-to-year variations. In addition, the

unavailability of certain contextual variables, such as harvest size, irrigation access, and disaster events, may also contribute to the inaccuracy of the predictions. Nonetheless, the Adjusted R-squared values obtained from the model for each year in both provinces were within the range of 78% to 88%, indicating that the model was able to explain a good proportion of the variability in the data. Therefore, these differences in prediction values do not detract from the usefulness of the model, as the GAM framework remains effective in describing general production trends as well as identifying influential factors under different regional conditions

5. CONCLUSION

This study shows that the Generalized Additive Model (GAM) approach with truncated splines can be used to model and predict rice production in Indonesia, especially in three provinces: Riau Islands and DKI Jakarta. The estimation results show that the model is able to capture rice production trends despite variations in accuracy between provinces. In Riau Islands, the model predicts an upward trend despite the actual data showing a decline. Meanwhile, in DKI Jakarta, the model's predictions are quite consistent with the relatively stable actual trend. Overall, GAM modeling with truncated splines provides a useful initial look at annual rice production trends and can be further developed to improve prediction accuracy, especially if supplemented with additional variables or more complex modeling techniques.

6. ACKNOWLEDGMENT

The author extends sincere thanks to God Almighty and heartfelt appreciation to everyone who has contributed to the completion of this research. Special gratitude is directed to the supervisor for their valuable guidance and support throughout the process. It is the author's hope that this article will be beneficial and enhance the knowledge of its readers. Acknowledging the limitations in knowledge and capability, the author is fully aware that this article is not without flaws. Therefore, constructive feedback and suggestions for future improvement are warmly welcomed.

7. REFERENCES

[1] F. Aditya, E. Gusmayanti, and J. Sudrajat, "Pengaruh Perubahan Curah Hujan terhadap Produktivitas Padi Sawah di Kalimantan Barat," *J. Ilmu Lingkung.*, vol. 19, no. 2, pp. 237–246, 2021, doi: 10.14710/jil.19.2.237-246.

[2] G. F. Aly, *Keefektifan pengaruh hashtag, cuaca, dan mobilitas terhadap jumlah kasus harian covid-19 di Jawa dan Sumatera dengan generalized estimating equation*. 2021. [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/65030%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65030/1/GERY FERDIAN ALY-FST.pdf

[3] Badan Pusat Statistik (BPS), *Statistik Indonesia 2022: Statistical Yearbook of Indonesia*. Jakarta, Indonesia: BPS, 2022.

[4] BPS, *Luas Panen dan Produksi Padi di Indonesia 2023: Angka Sementara*. Jakarta, Indonesia: Badan Pusat Statistik, 2023.

[5] Badan Pusat Statistik (BPS), *Statistik Indonesia 2023: Statistical Yearbook of Indonesia*. Jakarta, Indonesia: BPS, 2023.

[6] BPS, "Persentase tenaga kerja informal sektor pertanian," *Badan Pusat Statistik*, 2023. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MTE3MSMy/persentase-tenaga-kerja-informal-sektor-pertanian.html

[7] BPS, "Luas panen, produksi, dan produktivitas padi menurut provinsi," *Badan Pusat Statistik*, Mar. 13, 2024. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen--produksi--dan-produktivitas-padi-menurut-provinsi.html

[8] Badan Pusat Statistik (BPS), *Statistik Indonesia 2024: Statistical Yearbook of Indonesia*. Jakarta, Indonesia: BPS, 2024.

[9] I. N. Budiantara, "Model keluarga spline polinomial truncated dalam regresi semiparametrik," *BIMIPA*, vol. 15, no. 3, pp. 55–61, 2005.

[10] N. Chaniago, "The Effect of Rainfall on Rice Production and Productivity in Percut Sei Tuan District, Deli Serdang Regency, North Sumatra," *Agril. J. Ilmu Pertan.*, vol. 11, no. 3, pp. 130–136, 2023, [Online]. Available: https://jurnal.uisu.ac.id/index.php/agriland

[11] Danardono, *Analisis Data Longitudinal*. Yogyakarta, Indonesia: Gadjah Mada University Press, 2018.

[12] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd ed. New York, NY, USA: Marcel Dekker, 1999. doi: 10.1201/9781482273144.

[13] A. Z. Faiz, "Kondisi Tenaga Kerja pada Sektor Pertanian, Kehutanan dan Perikanan Tahun 2020 di Provinsi Jawa Tengah," *J. Ketenagakerjaan*, vol. 16, no. 2, 2021, doi: 10.47198/naker.v16i2.101.

[14] W. Häardie and M. Müller, "Multivariate and Semiparametric Kernel Regression," pp. 357–391, 2000, doi: 10.1002/9781118150658.ch12.

[15] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to Statistical Science. ® www.jstor.org," *Stat. Sci.*, vol. 4, no. 4, pp. 409–435, 1989, doi: 10.2307/2246134.

[16] W. Hulin and Z. Jin-Ting, "Nonparametric regression methods for longitudinal data analysis," in *Choice Reviews Online*, vol. 44, no. 02, 2006. doi: 10.5860/choice.44-0978.

[17] S. Imaslihkah, M. Ratna, and V. Ratnasari, "Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya," *J. Sains dan Seni Pomits*, vol. 2, no. 2, pp. 177–182, 2013.

[18] Kementerian Pertanian Republik Indonesia, *Statistik Pertanian 2019*. Jakarta, Indonesia: Pusat Data dan Sistem Informasi Pertanian, Kementerian Pertanian Republik Indonesia, 2019.

[19] M. N. Lestari, "Perbandingan Model Regresi Logistik Dan Generalized Estimating Equation (Gee) Dalam Menganalisis Penyebab Tugas Akhir," 2018.

[20] M. Maun and M. U. Kaleka, "Penerapan pertanian ramah lingkungan guna menyediakan pangan sehat dan aman di Indonesia," *Jurnal Ilmu Pertanian Tropis*, vol. I, no. 1, pp. 14–17, 2024.

[21] P. S. Matematika, F. Sains, D. A. N. Teknologi, U. Islam, and N. Walisongo, "PERBANDINGAN GENERALIZED ADDITIVE MODEL (GAM) DAN GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) PADA PEMODELAN INVESTASI DAERAH DI," 2023.

[22] T. A. Prasanti, T. Wuryandari, and A. Rusgiyono, "Aplikasi regresi data panel untuk pemodelan tingkat pengangguran terbuka kabupaten/kota di Provinsi Jawa Tengah," *Jurnal Gaussian*, vol. 4

[23] R. A. Qudriyah, Y. Prasetyo, and M. A. Yusuf, "Analisis Pengaruh Curah Hujan Terhadap Estimasi Produktivitas Padi Berbasis Pemrosesan Citra Sentinel 2a Pada Subround I Dan Ii Tahun 2018-2021 (Studi Kasus : Kecamatan Winong, Kabupaten Pati)," *Elipsoida J. Geod. dan Geomatika*, vol. 5, no. 1, pp. 16–23, 2022, doi: 10.14710/elipsoida.2022.16859.

[24] M. Ramli, I. N. Budiantara, and V. Ratnasari, "Statistical inference on nonparametric regression model with approximation of Fourier series function: Estimation and hypothesis testing," *MethodsX*, vol. 13, no. August, 2024, doi: 10.1016/j.mex.2024.102922.

[25] D. Rochdiani, D. Sosial, E. Pertanian, and F. Pertanian, "Dinamika Produksi Padi Kaitannya Dengan Ketahanan Pangan Di Indonesia Rice Production Dynamics the Relationship," vol. 1, pp. 15–24, 2022.

[26] T. P. Ryan, *Modern Regression Methods*, 2nd ed. Hoboken, NJ, USA: Wiley, 2008.

[27] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: SIAM, 1990.

[28] W. Manurung and M. Fajar, "Pemodelan produktivitas padi menggunakan generalized additive models di Provinsi Banten," *J. Lebesgue: J. Ilmiah Pendidik. Mat., Mat. Statist.*, vol. 1, no. 2, pp. 105-112, 2020.

[29] Widana, I. W., & Muliani, N. P. L. (2020). Uji persyaratan analisis.

[30] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 1st ed. Boca Raton, FL, USA: CRC Press, 2017.

[31] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. New York, NY, USA: Chapman & Hall/CRC, 2017.