CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 36-52
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Simulation Study and Development of Semiparametric Multiresponse Multigroup Using Truncated Spline Regression for Rice Pest Control

Laila Nur Azizah*, Adji Achmad Rinaldo Fernandes, Ni Wayan Surya Wardhani

Department of Statistics, University of Brawijaya, Indonesia

Email: laila.nurazizah52@gmail.com

ABSTRACT

This study aims to develop a multi-response semiparametric regression model using a truncated spline multi-group approach to understand the factors influencing rice pest control under light and dark conditions. The approach utilized is a multi-response semiparametric regression based on truncated splines, with analyses conducted on secondary and simulated data. The model was tested under various scenarios to identify the best model based on the coefficient of determination. The analysis revealed that the optimal model for secondary data is a semiparametric regression model with a linear order and a single knot point, achieving a coefficient of determination of 89.2%. For simulated data, the linear model with a single knot point also showed the best performance, with the highest coefficient of determination at 96.1%, particularly when error variance and multicollinearity levels were kept low to moderate. The study concludes that the multi-response semiparametric regression model using a truncated spline approach is effective in capturing relationships between variables in rice pest control, both in actual and simulated data. This model proves optimal for situations involving complex data variability. The research contributes methodologically to the development of more flexible multi-response regression models.

Keywords: Multi-Group; Multi-Responses Semiparametric Regression; Rice Pest; Truncated Spline; Weighted Least Square

INTRODUCTION

A statistical method that is often used to determine the relationship pattern between one or more variables is regression analysis. Regression analysis can identify the relationship between predictor variables and response variables [1]. Predictor and response variables must first be determined as the first step in performing regression analysis. After that, it is necessary to determine the relationship between predictor variables and response variables through the relationship curve displayed by the scatter diagram. Based on the shape of the relationship curve and the fulfillment of the linearity assumption tested using Ramsey RESET, there are three approaches used to estimate the regression curve, namely parametric, nonparametric, and semiparametric

Submitted: November 10, 2024 Reviewed: December 12, 2024 Accepted: January 10, 2025
DOI: http://dx.doi.org/10.18860/ca.v10i1.29773

approaches [2].

In parametric regression, a linear relationship between predictor and response variables. If this linearity assumption is not met, a nonparametric regression model can be used as an alternative. The parametric regression approach has rigid and strong assumptions, and the regression curve's shape is already known. Conversely, in the nonparametric regression approach, the regression curve's shape is assumed to be unknown. Nonparametric regression curves are considered to have high flexibility [3]. If part of the regression curve pattern is known and part is unknown, then a semiparametric regression approach is used [4]. The regression curve pattern can be identified by examining the scatter plot of each predictor and response variable.

One form of nonparametric regression is spline, which is continuous and segmented polynomial pieces [2]. Spline is divided into two models namely smoothing spline and truncated spline. Smoothing spline is used to predict functions based on model accuracy criteria and smoothing parameters that determine the size of the smoothing curve. While the truncated spline considers the presence of knot points in determining the optimal points that should be used. Truncated spline has knot points that show changes in data behavior patterns so that the spline has the advantage and finds the data pattern by itself. This advantage is because the spline can provide good flexibility so that it is possible to adjust to the local characteristics of the function or data [5]. Truncated spline has the main advantage in overcoming data that has significant changes in behavior at certain sub-intervals. Truncated Spline analysis using spline knot points can produce a more flexible regression curve because it is able to overcome data patterns that show sharp up and down patterns [6].

Truncated spline regression allows the model to adapt to data complexity without imposing a linear shape that may not fit the data. Therefore, it is necessary to ensure that the relationship between the predictor and response variables is linear or requires a more complex model, such as a truncated spline. The linearity assumption test used is the Ramsey RESET (Regression Equation Specification Error Test). Currently, there is a modification of the Ramsey RESET test developed by [2], namely the quadratic Ramsey RESET modification. This test is designed to capture quadratic patterns in the data, thereby identifying models that may require simple nonlinear elements like second-degree polynomials. However, the quadratic Ramsey RESET modification is limited to quadratic patterns and cannot detect more complex patterns, such as those captured by truncated spline models. To date, there has been no further development leading to a Ramsey RESET modification that can specifically detect the need for a truncated spline model. This indicates an opportunity to develop a more sophisticated test approach, which can directly evaluate whether the data requires a more complex spline approach.

A relationship in regression analysis is not always between predictor variables and one response variable. Multi-response regression is a regression model when the response variable is more than one and a response variable has a relationship with other response variables. Multirespon regression allows the relationship between variables that can be seen through the variance matrix [7]. The estimation used is Weighted Least Square. WLS allows more accurate estimation in truncated spline regression models for multirespon data, especially in capturing complex pattern changes in various data segments.

The development of the multirespon multigroup regression model in this study lies in the simultaneous multirespon multigroup analysis, where usually each group is analyzed one by one for each response variable. In the multigroup multiresponse regression approach, several response variables are analyzed simultaneously for each group and facilitated by the use of dummy variables to accommodate differences in characteristics between groups. Multigroup multiresponse semiparametric regression models have a wide range of potential applications, including in the agricultural field related to pests.

As an agricultural country, Indonesia aims to increase rice productivity to achieve food self-sufficiency. One of the biggest challenges in rice cultivation is pest attacks that can significantly reduce yields. According to [8], effective rice pest control involves the using of various methods, including chemical pesticides, pest-resistant varieties, and integrated ecosystem management. Pests such as brown planthoppers (*Nilaparvata lugens*), stem borers (*Scirpophaga incertulas*), and rice field rats often cause significant losses in rice production in Indonesia, even up to 30-50% of the potential harvest in some regions [9]. Farmers can experience increased yields and better welfare when pest control methods are applied correctly. The effectiveness of pest control is crucial to ensure the sustainability of rice production and farmers' welfare. Therefore, pest control strategies must be comprehensively designed to achieve optimal results by considering environmental and economic factors. This study divides rice pest control into two conditions, namely day and night, because pests in both conditions have different activities.

The research conducted aims to develop a model of the effectiveness of rice pest control in Indonesia through multirespon multigroup semiparametric regression analysis with a truncated spline estimator. This model is expected to help farmers and policy makers in designing effective pest control strategies in order to increase rice productivity so as to achieve food self-sufficiency. In addition, this model also conducts simulation studies to see the best multirespon multigroup semiparametric regression model scenario. This model is expected to assist farmers and policymakers in designing effective pest control strategies to improve rice productivity, ultimately achieving food self-sufficiency. The novel approach proposed for rice pest control utilizes a multi-response multi-group semiparametric regression model. This approach accounts for differences in pest activity under light and dark conditions, offering flexibility in analyzing complex and heterogeneous data, and supporting increased crop yields and the welfare of farmers in Indonesia. The multi-response multi-group approach has not been applied in previous research, making it an innovative contribution of this study.

METHODS

Data and Research Variables

This study used secondary data from research by Wardhani, et al (2024) to analyze the presence of pests and natural enemies in two temporal conditions, namely day and night. three pests analyzed in this study were *Oxya* sp. (X_1), *Nilaparvata* sp. (X_2), and *Chilo* sp. (X_3). Meanwhile, the three natural enemies analyzed were *Sympetrum* sp. (Y_1), *Pardosa* sp. (Y_2), and *Coccinella* sp. (Y_3). The relationship between predictor variables and response variables is as shown in Figure 1.

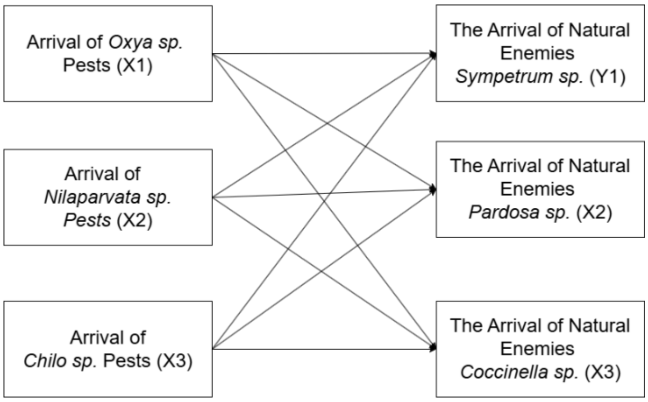

Figure 1. Research Model

A sample size of 60 observations was determined for each group (120 samples in total).

Stages of Data Analysis

The stages of data analysis carried out in this study are as follows:

1. Prepare secondary data
2. Test the linearity assumption with the modified Regression Specification Error Test (RESET).
3. Estimating the component functions of multiresponse semiparametric regression using Iterative Weighted Least Square. WLS has a diagonal matrix whose elements consist of the components of vector **W** called the weighting matrix, where **W** is equal to $\hat{\boldsymbol{\Sigma}}^{-1}$. Thus, the estimator form of the spline function using WLS is as follows [10].

$$\hat{\underset{\sim}{f}} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\underset{\sim}{y} \quad (1)$$

4. Build multigroup multiresponse semiparametric regression model with truncated spline function.
5. Selecting the optimal knot points in the truncated spline multirespon semiparametric regression based on the minimum Generalized Cross Validation (GCV). The GCV method can be expressed using the following formula [11].

$$GCV(K) = \frac{MSE(\underset{\sim}{k})}{\left(n^{-1}tr\left[\mathbf{I} - \mathbf{A}_{\underset{\sim}{k}}\right]\right)^2} \quad (2)$$

$$MSE\left(\underset{\sim}{k}\right) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$

6. Determining the best spline model by calculating the minimum GCV value of all models formed.
7. Estimating the best semiparametric multirespon multigrup truncated spline regression model in pest control modeling. The best model is determined based on the model with the highest coefficient of determination.
8. Calculating the model determination coefficient. The coefficient of determination represents the proportion of total variability around the mean value explained by the regression model [12]. The formula for the coefficient of determination is defined as follows [13].

$$R^2_{k.adjusted} = 1 - \left(\frac{\sum_{i=1}^{n} \left(\bar{y}_k - \hat{f}_{ki} \right)^2 / (n-k-1)}{\sum_{i=1}^{n} \left(y_{ki} - \bar{y}_k \right)^2 / (n-1)} \right); 0 \le R^2_k \le 1 \quad (3)$$

9. Generating simulated data from secondary data.
10. Form a model from the simulation data according to the scenario that has been determined from the simulation data.
11. Testing the accuracy of the results of each simulation data scenario using the GCV value.
12. Interpret the results of secondary data analysis and simulation data analysis results.

Data Simulation Steps

The stages of data simulation in multigroup multirespon semiparametric regression are as follows:

1. Assign data for predictor variables derived from secondary data.
2. Using regression coefficients obtained from secondary data for simulation data to be applied.
3. Generate multivariate normally distributed error values with error variances of 0.3; 0.5; 0.8.
4. Establish multicollinearity with low, medium, and high categories. Multicollinearity is a condition where there is a linear relationship or high correlation between predictor variables in a regression model. The presence of multicollinearity can be identified if the Variance Inflation Factor (VIF) value exceeds 10.

$$VIF_h = \frac{1}{1-R_h^2} \quad (4)$$

R_h is the coefficient of determination of a single predictor variable in a regression model as the response variable (X_h) and the predictor variables in the regression model as other predictor variables $(X_i, i \neq h)$. The criteria for VIF values are as follows [14].

Table 1. Category of Multicollinearity

Criteria	Multicollinearity Level Categories
$VIF < 5$	Low
$5 < VIF < 10$	Middle
$VIF > 10$	High

5. Forming simulation data by combining predictor variables and response variables based on scenarios with each scenario performed 100 times.
6. The generation scenario was designed to determine the impact of different semiparametric components on the regression model and different multigroup conditions. The designed scenarios consisted of possible models that were obtained and designed in such a way as to fulfill the conditions in accordance with the design of the scenarios described in Table 2.

Table 2. Data Simulation Scenario

Scenario	Relationship Between Variable					
	Group 1			Group 2		
	$X_1 \to Y_1$ $X_1 \to Y_2$ $X_1 \to Y_3$	$X_2 \to Y_1$ $X_2 \to Y_2$ $X_2 \to Y_3$	$X_3 \to Y_1$ $X_3 \to Y_2$ $X_3 \to Y_3$	$X_1 \to Y_1$ $X_1 \to Y_2$ $X_1 \to Y_3$	$X_2 \to Y_1$ $X_2 \to Y_2$ $X_2 \to Y_3$	$X_3 \to Y_1$ $X_3 \to Y_2$ $X_3 \to Y_3$
1	L	L1K	L1K	L	L1K	L1K
2	L	L	L2K	L	L	L2K

Scenario	Relationship Between Variable					
	Group 1			Group 2		
	$X_1 \to Y_1$ $X_1 \to Y_2$ $X_1 \to Y_3$	$X_2 \to Y_1$ $X_2 \to Y_2$ $X_2 \to Y_3$	$X_3 \to Y_1$ $X_3 \to Y_2$ $X_3 \to Y_3$	$X_1 \to Y_1$ $X_1 \to Y_2$ $X_1 \to Y_3$	$X_2 \to Y_1$ $X_2 \to Y_2$ $X_2 \to Y_3$	$X_3 \to Y_1$ $X_3 \to Y_2$ $X_3 \to Y_3$
3	L	Q1K	Q1K	L	Q1K	Q1K
4	L	L	Q2K	L	L	Q2K
5	L	L1K	Q2K	L	L2K	Q1K

L: Linear; L1K: Linear 1 Knot; L2K : Linear 2 Knot; Q1K: Quadratic 1 Knot, Q2K: Quadratic 2 Knots

RESULTS AND DISCUSSION

Ramsey RESET Modification

The Ramsey RESET (Regression Equation Specification Error Test) modification is a diagnostic tool designed to assess whether a linear regression model suffers from misspecification, such as omitted variables or incorrect functional forms. The steps of testing the modified Ramsey RESET are as follows:

- Regress the estimated value of the response variable as follows.

  $$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \qquad (5)$$

  Calculate the coefficient of determination of the regression with equation (6)

  $$R_1^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{y})^2} \qquad (6)$$

  Where
  Y_i : Response variable at the i-th observation
  $\hat{Y}_i^*$: Estimated response variable with the i-th additional predictor variable
  $\bar{y}$: Average value of response variable
  R_2^2 : Coefficient of determination of linear regression
- Form a regression equation by adding exogenous variables, namely so that the estimated value of the response variable is obtained as follows.

  Linear Order Truncated Spline with 1 Knot

  $$\hat{Y}_i = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2^* (X_{1i} - k_1)_+ \qquad (7)$$

  Linear Order Truncated Spline with 2 Knots

  $$\hat{Y}_i = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2^* (X_{1i} - k_1)_+ + \hat{\beta}_3^* (X_{1i} - k_2)_+ \qquad (8)$$

  Calculate the coefficient of determination of the regression with equation (9).

  $$R_2^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i^*)^2}{\sum_{i=1}^{n}(Y_i - \bar{y})^2}$$

  Where:
  Y_i : Response variable at the i-th observation
  $\hat{Y}_i^*$: Estimated response variable with the i-th additional predictor variable
  $\bar{y}$: Average value of response variable
  R_2^2 : Coefficient of determination of linear regression with additional predictor variables
- Test the linearity between predictor variables on the response with the following hypothesis

Linear Order Truncated Spline with 1 Knot

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

Linear Order Truncated Spline with 2 Knots

$H_0 : \beta_2 = \beta_3 = 0$

$H_1 :$ at least one $\beta_j \neq 0; j = 2,3$

With the value of the F test statistic using formula (10).

$$F_{statistics} = \frac{(R_2^2 - R_1^2)/m}{(1 - R_2^2)/(\text{n} - p - 1 - m)} \sim F_{m,n-p-1-m} \quad (10)$$

j = 1,2,…,m; m : additional exogenous variables
p : initial exogenous variable
i = 1,2,3,…n; n : number of observations

- Determining the critical value approximates the F distribution at α = 5%.

The test criterion is if the test statistic $F_{hitung} \geq F_{(\alpha,m,n-p-1-m)}$ or p-value < α then it's H_0 rejected, meaning the model is not linear.

The classic Ramsey RESET linearity test aims to detect the relationship between predictor variables and response variables. The Ramsey RESET test was applied to the temporally stratified data (No Group, Day Group, and Night Group) presented in Table 3.

Table 3. Ramsey RESET Classic Linearity Test Results

Num	Relationship between Variables	***p*-value**		
		Without Group	**Group 1 (Day)**	**Group 2 (Night)**
1	*Oxya sp.* pests. (X_1) → Natural Enemies of *Sympetrum* sp. (Y_1)	0,856	0,953	0,953
2	*Nilaparvata* sp pests. (X_2) → Natural Enemies of *Sympetrum* sp. (Y_1)	0,963	0,926	0,951
3	*Chilo* sp. pests. (X_3) → Natural Enemies of *Pardosa* sp. (Y_2)	0,942	0,990	0,252
4	*Oxya* sp. pests. (X_1) → Natural Enemies of *Pardosa* sp. (Y_2)	**<0,001***	0,987	0,272
5	*Nilaparvata* sp pests. (X_2) → Natural Enemies of *Pardosa* sp. (Y_2)	**0,001***	**0,010***	0,968
6	*Chilo* sp. pests. (X_3) → Natural Enemies of *Coccinella* sp. (Y_3)	0,647	0,565	0,693
7	*Oxya* sp. pests. (X_1) → Natural Enemies of *Coccinella* sp. (Y_3)	0,268	0,469	**<0,001***
8	*Nilaparvata* sp. pests. (X_2) → Natural Enemies of *Coccinella* sp. (Y_3)	0,440	0,985	0,115
9	*Chilo* sp. pests. (X_3) → Natural Enemies of *Coccinella* sp. (Y_3)	0,235	0,800	**<0,001***

Table 3 shows that the non-linear relationship in the category without groups is found in the relationship between *Oxya* sp. against natural enemies *Pardosa* sp. and the relationship between *Nilaparvata* sp. against natural enemies *Pardosa* sp. While in the day group category, the non-linear relationship is found in the relationship between *Nilaparvata* sp. against *Pardosa* sp. In the night group category, the non-linear relationship is found in the relationship between *Oxya* sp. pests against *Coccinella* sp. and the relationship between *Chilo* sp. pests against natural enemies *Coccinella* sp. These four relationships will be further identified regarding the non-linear relationship of the most suitable truncated spline approach by means of a linearity test with modified Ramsey RESET. The results of the Ramsey RESET modification from the application of equations (5) to (10).

Table 4. Linearity Test Results Modified Ramsey RESET Linear Order

Relationship Between Variables	Linear 1 Knot		Linear 2 Knots	
	Group 1 (Day)	Group 2 (Night)	Group 1 (Day)	Group 2 (Night)
Oxya sp pests (X_1) → Natural Enemies of *Pardosa* sp. (Y_2)	0,512	**0,019***	0,389	**0,061***
Nilaparvata sp pests (X_2) → Natural Enemies of *Pardosa* sp. (Y_2)	**0,001***	0,154	**0,006***	0,969
Oxya sp pests (X_1) → Natural Enemies of *Coccinella* sp. (Y_3)	**0,087***	0,866	0,302	0,905
Chilo sp pests (X_3) → Natural Enemies of *Coccinella* sp. (Y_3)	0,106	**<0,001***	0,161	**<0,001***

* means significant

Overall, the p-value of the 2-knot linear approach is higher than the 1-knot linear approach, which indicates that the 2-knot linear model is less sensitive in capturing the nonlinearity of the relationship between variables in the study. The linearity test results show that there is a linear and non-linear relationship between variables. The relationship between variables that are linear will be solved with a parametric approach. In contrast, the relationship between variables that are not linear will be solved by a nonparametric approach using a linear order truncated spline with one knot point. The combination of parametric and nonparametric approaches is called semiparametric so the next step is a multi-response multi-group truncated spline semiparametric regression analysis.

Function Estimation of Semiparametric Truncated Spline Multi-Response Multi-Group Regression Model

Multi-response multi-group semiparametric regression is an equation model that facilitates multi-response semiparametric regression on certain categories through a *dummy* approach. One form of multi-response regression involves three predictor variables (x_1, x_2, x_3) and three response variables (y_1, y_2, y_3) with some parametric and nonparametric relationships. The relationship between the variables follows the modified Ramsey RESET in Table 2 and Table 3. The multirespon multigroup *truncated spline* semiparametric regression equation model can be formulated according to this pattern as equation (11).

$$\begin{aligned}
\hat{y}_{1i} &= \beta_{01} + \beta_{11}x_{1i} + \beta_{21}x_{1i}D_i + \beta_{31}x_{2i} + \beta_{41}x_{2i}D_i + \beta_{51}x_{3i} + \beta_{61}x_{3i}D_i \\
\hat{y}_{2i} &= \beta_{02} + \beta_{12}x_{1i} + \beta_{22}x_{1i}D_i + \alpha_{12}(x_{1i} - k_1)_+D_i + \beta_{32}x_{2i} + \beta_{42}x_{2i}D_i + \\
&\quad \alpha_{22}(x_{2i} - k_2)_+(1 - D_i) + \beta_{52}x_{3i} + \beta_{62}x_{3i}D_i \\
\hat{y}_{3i} &= \beta_{03} + \beta_{13}x_{1i} + \beta_{23}x_{1i}D_i + \alpha_{13}(x_{1i} - k_3)_+(1 - D_i) + \beta_{33}x_{2i} + \beta_{43}x_{2i}D_i + \\
&\quad \beta_{53}x_{3i} + \beta_{63}x_{3i}D_i + \alpha_{23}(x_{3i} - k_4)_+D_i
\end{aligned} \tag{11}$$

The matrix form of equation (11) is:

$$\underset{\sim}{y} = \mathbf{X}\underset{\sim}{\beta} + f(\mathbf{X}) + \underset{\sim}{\varepsilon}$$

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3n} \end{bmatrix}_{3n \times 1} ; \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \vdots \\ \varepsilon_{3n} \end{bmatrix}_{3n \times 1}$$

$$f(\mathbf{X}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \alpha_{12}(x_{11} - k_1)_+ D_i + \alpha_{22}(x_{21} - k_2)_+ (1 - D_i) \\ \alpha_{12}(x_{12} - k_1)_+ D_i + \alpha_{22}(x_{22} - k_2)_+ (1 - D_i) \\ \vdots \\ \alpha_{12}(x_{1n} - k_1)_+ D_i + \alpha_{22}(x_{2n} - k_2)_+ (1 - D_i) \\ \alpha_{13}(x_{11} - k_3)_+ (1 - D_i) + \alpha_{23}(x_{31} - k_4)_+ D_i \\ \alpha_{13}(x_{12} - k_3)_+ (1 - D_i) + \alpha_{23}(x_{32} - k_4)_+ D_i \\ \vdots \\ \alpha_{13}(x_{1n} - k_3)_+ (1 - D_i) + \alpha_{23}(x_{3n} - k_4)_+ D_i \end{bmatrix}_{3n \times 1}$$

$$\mathbf{X}\underset{\sim}{\beta} = \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{n}\times 7} & \mathbf{0}_{\boldsymbol{n}\times 7} & \mathbf{0}_{\boldsymbol{n}\times 7} \\ \mathbf{0}_{\boldsymbol{n}\times 7} & \boldsymbol{X}_{\boldsymbol{n}\times 7} & \mathbf{0}_{\boldsymbol{n}\times 7} \\ \mathbf{0}_{\boldsymbol{n}\times 7} & \mathbf{0}_{\boldsymbol{n}\times 7} & \boldsymbol{X}_{\boldsymbol{n}\times 7} \end{bmatrix}_{3n \times 21} \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \\ \beta_{61} \\ \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{52} \\ \beta_{62} \\ \beta_{03} \\ \beta_{13} \\ \beta_{23} \\ \beta_{33} \\ \beta_{43} \\ \beta_{53} \\ \beta_{63} \end{bmatrix}_{21 \times 1}$$

Where:

$$\boldsymbol{X}_{\boldsymbol{n}\times\boldsymbol{7}} = \begin{bmatrix} 1 & x_{11} & x_{11}D_i & x_{21} & x_{21}D_i & x_{31} & x_{31}D_i \\ 1 & x_{12} & x_{12}D_i & x_{22} & x_{22}D_i & x_{32} & x_{32}D_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{1n}D_i & x_{2n} & x_{2n}D_i & x_{3n} & x_{33}D_i \end{bmatrix}_{n\times 7}$$

Notes:

y_{ij} : *i-th* value of *j-th* response variable

x_{ij} : i-th value of *j-th* predictor *variable*

β_{0j} : Intercept on the *jth* response variable

β_{ij} : Regression coefficient on the *j-th* response variable

ε_{ij} : The i-th error *on the j-th* response variable

Equation (11) is written in the form of multirespon regression for each group with D_i value is bellow.

Model equation for the day group with the value of D is 0.

$$\begin{aligned} \hat{y}_{11i} &= \beta_{01} + \beta_{11}x_{1i} + \beta_{31}x_{2i} + \beta_{51}x_{3i} \\ \hat{y}_{21i} &= \beta_{02} + \beta_{12}x_{1i} + \beta_{32}x_{2i} + \alpha_{22}(x_{2i} - k_2)_+ + \beta_{52}x_{3i} \\ \hat{y}_{31i} &= \beta_{03} + \beta_{13}x_{1i} + \alpha_{13}(x_{1i} - k_3)_+ + \beta_{33}x_{2i} + \beta_{53}x_{3i} \end{aligned} \quad (12)$$

Model equation for the night group with D value of 1.

$$\begin{aligned} \hat{y}_{12i} &= \beta_{01} + (\beta_{11} + \beta_{21})x_{1i} + (\beta_{31} + \beta_{41})x_{2i} + (\beta_{51} + \beta_{61})x_{3i} \\ \hat{y}_{22i} &= \beta_{02} + (\beta_{12} + \beta_{22})x_{1i} + \alpha_{12}(x_{1i} - k_1)_+ + (\beta_{32} + \beta_{42})x_{2i} + (\beta_{52} + \beta_{62})x_{32i} \\ \hat{y}_{32i} &= \beta_{03} + (\beta_{13} + \beta_{23})x_{1i} + (\beta_{33} + \beta_{43})x_{2i} + (\beta_{53} + \beta_{63})x_{3i} + \alpha_{23}(x_{3i} - k_4)_+ \end{aligned} \quad (13)$$

Where the *truncated* function is as follows.

$$(x_{1i} - k_1)_+ = \begin{cases} (x_{1i} - k_1), & x_{1i} \geq k_1 \\ 0 & , x_{1i} < k_1 \end{cases}$$

$$(x_{2i} - k_2)_+ = \begin{cases} (x_{2i} - k_2), & x_{2i} \geq k_2 \\ 0 & , x_{2i} < k_2 \end{cases}$$

$$(x_{1i} - k_3)_+ = \begin{cases} (x_{1i} - k_3), & x_{1i} \geq k_3 \\ 0 & , x_{1i} < k_3 \end{cases}$$

$$(x_{3i} - k_4)_+ = \begin{cases} (x_{3i} - k_4), & x_{3i} \geq k_4 \\ 0 & , x_{3i} < k_4 \end{cases}$$

The parameter estimation used for multi-response multi-group semiparametric regression is *Weighted Least Square* (WLS). This is because in multi-response semiparametric regression, the errors are not identical resulting in unequal error variances for each i, denoted by $var(\varepsilon_i) = \sigma_i{}^2$ so that the OLS method cannot be done because the OLS method has the assumption that the errors are identical (homogeneity in error variance). In order to fulfill the identical assumption, the transformation is done by multiplying ϵ_i with $\boldsymbol{W}$.

The diagonal matrix whose elements consist of the component vectors $\boldsymbol{W}$ is called the weighting matrix where $\boldsymbol{W}$ **is** equal to $\hat{\boldsymbol{\Sigma}}^{-1}$. The WLS method estimates the parameters by minimizing equation (14).

$$\underset{\sim}{\varepsilon}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\underset{\sim}{\varepsilon} = (\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}) \quad (14)$$

$$= (\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta})^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} - \hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\underset{\sim}{\beta})$$
$$= \underset{\sim}{y}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} - \underset{\sim}{y}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\underset{\sim}{\beta} - \underset{\sim}{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} + \underset{\sim}{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\underset{\sim}{\beta}$$

The minimum value of $\underset{\sim}{\varepsilon}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{\varepsilon}$ obtained when $\frac{\partial(\underset{\sim}{\varepsilon}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{\varepsilon})}{\partial\underset{\sim}{\beta}} = \underset{\sim}{0}$ resulting in equation (15).

$$\hat{\underset{\sim}{\beta}} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} \quad (15)$$

After that, equation (15) is formed into $\hat{\underset{\sim}{f}} = \mathbf{X}\hat{\underset{\sim}{\beta}}$ hence the form of the *spline* function estimator as equation (16).

$$\hat{\underset{\sim}{f}} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} \quad (16)$$

Knot Point Determination with GCV

Linearity test results with Ramsey modification in Table 2, there are 4 out of 18 relationships between predictor variables and response variables that have a non-linear relationship with linear order and one knot point. Completion with *truncated spline* requires knot points in the modeling so that the selection of the best knot points in nonparametric analysis is done using GCV. The best knot is obtained from the knot with the smallest GCV value. Table 5 is a table containing knots for each non-linear relationship between predictor variables and their response variables along with GCV values.

Table 5. Optimal Knot Points for Each Relationship

Group	Predictor Variable	Response Variabke	Knots Point	GCV	R^2
1	*Nilaparvata* sp. pests (X_2)	Natural Enemies of *Pardosa* sp. (Y_2)	2	2,254	89,174%
1	*Oxya* sp pests. (X_1)	Natural Enemies of *Coccinella* sp. (Y_3)	2		
2	*Oxya* sp pests (X_1)	Natural Enemies of *Pardosa* sp. (Y_2)	1		
2	*Chilo* sp pests (X_3)	Natural Enemies of *Coccinella* sp. (Y_3)	1		

Table 5 shows each of the best knot points that will be used to model the *truncated spline* on each relationship between predictor variables and response variables. The results of the multi-response multi-group *truncated spline* semiparametric regression model when linear with 1 knot point for three predictor variables and three response variables are as follows.

$$\hat{y}_{1i} = 2{,}516 + 0{,}085x_{1i} - 0{,}549x_{1i}D_i - 0{,}166x_{2i} - 1{,}012x_{2i}D_i + 0{,}015x_{3i} - 0{,}054x_{3i}D_i$$
$$\hat{y}_{2i} = 0{,}494 - 0{,}021x_{1i} + 0{,}367x_{1i}D_i - 0{,}573(x_{1i} - 1)_+D_i + 0{,}304x_{2i} - 0{,}366x_{2i}D_i + 0{,}001(x_{2i} - 2)_+(1 - D_i) - 0{,}003x_{3i} + 0{,}063x_{3i}D_i$$
$$\hat{y}_{3i} = 0{,}363 - 0{,}0004x_{1i} - 0{,}066x_{1i}D_i + 0{,}178(x_{1i} - 2)_+(1 - D_i) - 0{,}049x_{2i} + 0{,}339x_{2i}D_i + 0{,}052x_{3i} - 0{,}511x_{3i}D_i + 0{,}552(x_{3i} - 1)_+D_i$$

Model equation in the day group with D value 0.

$$\hat{y}_{11i} = 2{,}516 + 0{,}085x_{1i} - 0{,}166x_{2i} + 0{,}015x_{3i}$$
$$\hat{y}_{21i} = 0{,}494 - 0{,}021x_{1i} + 0{,}304x_{2i} + 0{,}001(x_{2i} - 2)_+ - 0{,}003x_{3i}$$
$$\hat{y}_{31i} = 0{,}363 - 0{,}0004x_{1i} + 0{,}178(x_{1i} - 2)_+ - 0{,}049x_{2i} + 0{,}052x_{3i}$$

Model equation for the night group with D value of 1.

$$\hat{y}_{12i} = 2{,}516 - 0{,}464x_{1i} - 1{,}178x_{2i} - 0{,}039x_{3i}$$
$$\hat{y}_{22i} = 0{,}494 + 0{,}351x_{1i} - 0{,}573(x_{1i} - 1)_+ - 0{,}062x_{2i} + 0{,}060x_{3i}$$

$\hat{y}_{32i} = 0{,}363 - 0{,}0664x_{1i} + 0{,}29x_{2i} - 0{,}459x_{3i} + 0{,}552(x_{3i} - 1)_+$

The coefficient of determination is 89.174%. This means that the model is able to explain the diversity of the response variable by 89.174% and the rest is explained by other unknown factors in the model. Furthermore, partial regression coefficients were conducted to determine the contribution of each predictor variable to the response variable. The largest contribution was the arrival of *Nilaparvata* sp. pests to the arrival of *Sympetrum* sp. natural enemies by 33.5%. Meanwhile, the smallest contribution was the arrival of *Oxya* sp. pests against the arrival of natural enemies *Pardosa* sp. by 1.5%. The percentage contribution of each variable varies. Highlights that the effect of pests on natural enemies is different, where certain pests such as the arrival of *Nilaparvata* sp. play an important role in showing the arrival of natural enemies *Sympetrum* sp. In addition, the arrival of *Chilo* sp. contributes greatly to the arrival of natural enemies *Pardosa* sp. The arrival of *Chilo* sp. contributes greatly to natural enemies *Coccinella* sp.

Test Results with Bootstrap Resampling

The results of the hypothesis test using *bootstrap resampling* conducted 1000 times are presented in Table 6 (the * sign indicates significant at the 5% real level).

Table 6. Significance Test Results with Bootstrap Resampling

Relation	Group	Coefficient	Test Statistics	P Value	Description
$x_1 \to y_1$	1	0,085	3,954	<0,001*	Sig
$x_2 \to y_1$	1	-0,166	-7,687	<0,001*	Sig
$x_3 \to y_1$	1	0,015	0,690	0,313	Not Sig
$x_1 \to y_2$	1	-0,021	-0,988	0,243	Not Sig
$x_2 \to y_2$	1	0,304	14,064	<0,001*	Sig
$x_2 \to y_2$	1	-0,001	-0,001	0,398	Not Sig
$x_3 \to y_2$	1	-0,003	-0,183	0,391	Not Sig
$x_1 \to y_3$	1	-0,001	-0,019	0,397	Not Sig
$x_1 \to y_3$	1	0,178	8,236	<0,001*	Sig
$x_2 \to y_3$	1	-0,049	-2,268	0,031*	Sig
$x_3 \to y_3$	1	0,052	2,388	0,024*	Sig
$x_1 \to y_1$	2	-0,046	-21,442	<0,001*	Sig
$x_2 \to y_1$	2	-1,178	-54,526	<0,001*	Sig
$x_3 \to y_1$	2	-0,039	-1,831	0,075	Sig
$x_1 \to y_2$	2	0,345	15,990	<0,001*	Sig
$x_1 \to y_2$	2	-0,573	-26,526	<0,001*	Sig
$x_2 \to y_2$	2	-0,063	-2,897	0,006*	Sig
$x_3 \to y_2$	2	-0,067	-3,111	0,003*	Sig
$x_1 \to y_3$	2	-0,066	-3,062	0,003*	Sig
$x_2 \to y_3$	2	0,291	13,444	<0,001*	Sig
$x_3 \to y_3$	2	-0,460	-21,283	<0,001*	Sig
$x_3 \to y_3$	2	0,552	25,544	<0,001*	Sig

Overall, during the day and night, the arrival of the natural enemy *Sympetrum sp.* is influenced by the presence of pests such as *Oxya sp.* and *Nilaparvata sp.*. *Sympetrum sp.* belongs to the dragonfly genus, which are active predators during the day. Dragonflies possess excellent vision for detecting prey movement, making the presence of pests like *Oxya sp.* (grasshoppers) and *Nilaparvata sp.* (planthoppers), often seen in open areas during the day, trigger the predator's activity. Dragonflies also have a preference for open and watery habitats, where pests often lay eggs or gather for feeding, making it easier for *Sympetrum sp.* to detect their prey. At night, although the activity of pests such as *Nilaparvata sp.* persists, the activity level of dragonflies may be lower. However, some species of *Sympetrum* remain active during dusk and nighttime, utilizing their sensitive motion detection abilities to target pest insects, though less frequently than during the day.

During the day, the arrival of the natural enemy *Pardosa sp.* is influenced by the presence of *Nilaparvata sp.*, while at night, the arrival of *Pardosa sp.* is influenced by *Oxya sp.*, *Nilaparvata sp.*, and *Chilo sp. Pardosa sp.*, a type of wolf spider, exhibits nocturnal behavior or activity during low-light conditions. Wolf spiders rely on movement and vibrations from pests as their primary prey detection methods. Planthoppers (*Nilaparvata sp.*) are generally active and visible on plants during the day, allowing spiders to detect and wait around such areas. However, during the day, the influence of *Oxya sp.* and *Chilo sp.* may be insignificant as the spiders tend to hide and become active only at dusk [15]. At night, the presence of *Oxya sp.* is more noticeable to *Pardosa sp.* as these spiders have good night vision and quick reactions to pursue slow-moving prey like grasshoppers. However, *Chilo sp.*, which tends to hide within stems or plant cavities, is less likely to be detected by *Pardosa sp.* during either day or night.

During both day and night, the arrival of the natural enemy *Coccinella sp.* is influenced by *Oxya sp.*, *Nilaparvata sp.*, and *Chilo sp*. However, during the day, the presence of *Oxya sp.* only has an effect when there are more than two grasshopper pests. Meanwhile, at night, the arrival of *Coccinella sp.* is influenced by *Nilaparvata sp.* and *Chilo sp. Coccinella sp.*, or ladybugs, are predatory insects that are more effective during the day as they rely on vision to locate prey such as aphids and planthoppers. The influence of *Oxya sp.* and *Chilo sp.* during the day can be explained by the increased activity of these pests on the surface of plants, which attracts the attention of *Coccinella*. At night, although ladybugs are not fully active, some species continue to hunt, particularly targeting pests that do not hide, such as planthoppers [16]. Meanwhile, *Oxya sp.*, which typically hides at night, becomes less visible to *Coccinella sp.*, resulting in an insignificant influence.

Overall, this study shows that natural predators have varying effectiveness depending on the type of pest and environmental conditions. These results support the importance of biological pest control strategies in agriculture, where the use of natural enemies can help manage pest populations without relying on chemical pesticides that may harm the ecosystem.

Simulation Study Results

Simulated data scenarios are model shapes that are set up to illustrate whether the best multigroup semiparametric multirespon regression model is in the form of a linear, quadratic, or flexible relationship as listed in Table 1.

Table 7. Simulation Results

Error Variance	VIF	Scenario	R^2	r^2_{rata}
0,1	Low	1	0,9612	0,2609
		2	0,9414	0,2499
		3	0,9329	70,2502
		4	0,9377	0,2461
		5	0,9431	0,2506
	Middle	1	0,9512	0,2537
		2	0,9148	0,2379
		3	0,9274	0,2455
		4	0,9251	0,2471
		5	0,8707	0,2269
	High	1	0,9532	0,2653

Error Variance	VIF	Scenario	R^2	r^2_{rata}
		2	0,9214	0,2522
		3	0,9322	0,2580
		4	0,9345	0,2584
		5	0,9456	0,2502
0,3	Low	1	0,9517	0,2578
		2	0,9231	0,2406
		3	0,9378	0,2460
		4	0,9381	0,2482
		5	0,9372	0,2509
	Middle	1	0,9345	0,2495
		2	0,9285	0,2421
		3	0,9317	0,2463
		4	0,9231	0,2423
		5	0,8907	0,2264
	High	1	0,9575	0,2594
		2	0,9307	0,2492
		3	0,9213	0,2498
		4	0,9016	0,2454
		5	0,9290	0,2344
0,5	Low	1	0,9588	0,2623
		2	0,9397	0,2472
		3	0,9306	0,2435
		4	0,9195	0,2411
		5	0,9484	0,2528
	Middle	1	0,9414	0,2552
		2	0,9316	0,2442
		3	0,9290	0,2421
		4	0,9178	0,2398
		5	0,8743	0,2259
	High	1	0,9488	0,2496
		2	0,9173	0,2363
		3	0,8887	0,2246
		4	0,8955	0,2292
		5	0,8995	0,2188

The results in Table 7 show the performance of the model under the five simulation scenarios, which includes the overall coefficient of determination as well as the partial coefficient of determination for some combinations of variables. The coefficient of determination in each scenario indicates how well the model generally explains the variability of the data. Scenario 1 has the highest coefficient of determination, at 0.9612, which means that the model is able to explain about 96.12% of the data variability, while scenario 5 has the lowest value with an R^2 of 0.8743. This shows that the model's ability to explain data variability decreases slightly as the scenarios change. The average value of the partial determination coefficient in each scenario shows the average performance of the model in explaining the relationship between predictor variables and response variables.

The error bars in this simulation range from 0,1 to 0,5. The EV value indicates the amount of error variability in the model, the higher of which will affect the accuracy of the model in capturing the relationship pattern between the dependent and independent variables. Based on table 4.14, the larger the error variance value, the model tends to experience a decrease in the coefficient of determination, which indicates a decrease in the model's ability to explain data variability.

In the aspect of multicollinearity (VIF), there are three categories that indicate the level of multicollinearity between predictors in the model, namely low, medium, and high. At a low VIF level, multicollinearity between predictors is not so influential, then the predictors in the model are relatively more independent of each other. There is a multicollinearity effect at a medium VIF level, but still within insignificant limits. Meanwhile, at a high VIF level, multicollinearity between predictors is quite strong, so there is a large relationship between predictors in the model.

Looking at the coefficient of determination, which shows how well the model explains the variability of the data, it can be seen that at an error variance of 0.1, the coefficient of determination is quite high, even in scenarios with high VIF. This indicates that at low error variance, the model is able to explain the variability of the data well despite multicollinearity. However, at an error variance of 0.3, the coefficient of determination begins to decrease, especially in scenarios with high VIF, indicating that as the error variance increases, the model's ability to explain data variability begins to weaken. At an error variance of 0.5, the decrease in the coefficient of determination is even more significant, especially at high VIF levels, where the coefficient of determination is in the range of 0.87 to 0.90, indicating that the model experiences a significant decline in performance when the error variance and multicollinearity levels are both high.

$r^2_{average}$ is the partial determination coefficient, indicating the contribution of the predictor variable in partially explaining the variance of the response variable. At an error variance of 0.1 and a low VIF, the value of the partial determination coefficient is around 0.25, indicating a moderate contribution from the predictors, suggesting that most of the variability can still be explained. However, at moderate VIF levels, the value of the partial determination coefficient decreases slightly, especially in scenario 5 where it reaches a value of 0.2269, indicating that multicollinearity is having an impact, albeit a modest one. At high VIF at all levels of error variance, the coefficient of partial determination fluctuates more. For example, at an error variance of 0.5 and a high VIF, the partial determination coefficient decreases to a minimum value of 0.2188, indicating that the partial contribution of each predictor weakens as multicollinearity increases.

Overall, Table 7 shows that the performance of the model is affected by an increase in the error variance and VIF. At low to moderate error variances and low to moderate VIF levels, the model is more stable and able to capture the data pattern well. However, as the error variance and VIF increase to higher levels, the model experiences a significant decline in performance, both in terms of the lower coefficient of determination and in terms of the partial coefficient of determination, which shows a weakening partial contribution. This confirms that multi-response regression models are more optimal when the error variance and multicollinearity levels are kept low to moderate.

CONCLUSIONS

The model using truncated splines demonstrated a high coefficient of determination (89.17%), indicating good accuracy in predicting the arrival of natural enemies based on rice pests. The development of a multi-response multi-group semiparametric regression model using a truncated spline approach at linear and quadratic orders with varying knot levels allows for more flexible estimation in modeling the relationship between pests and their natural enemies in rice plants. By applying a dummy variable approach to create two groups, day and night, the function estimation process can be simplified. Based on rice pest abundance data, pest arrivals had a greater influence on the arrival of natural enemies at night than during the day. The significance tests revealed that the arrival of *Nilaparvata sp.* and *Oxya sp.* affected the arrival of *Sympetrum sp.* during both day and night. The arrival of *Nilaparvata sp.* also influenced the arrival of *Pardosa sp.* during the day. Meanwhile, all three pests, namely *Oxya sp., Nilaparvata sp.,* and *Chilo sp.,* influenced the arrival of *Coccinella sp.* during both day and night. Simulation study results indicated that the best model was found in Scenario 1, which involved a semiparametric model with a nonparametric component of linear order, one knot point, and low multicollinearity levels.

REFERENCES

[1] Fernandes, A.A.R. & Solimun. *Analisis Regresi dalam Pendekatan Fleksibel*. Universitas Brawijaya Press. 2022.

[2] Fernandes, A. A. R., Rahmawati, I., Muflikhah, L., Ubaidillah, F., & Al Rohimi, A. Research on Structural Flexibility and Acceptance Model (SFAM) Reconstruction Based on Disruption Innovation in the Social Humanities and Education Sector. *WSEAS Transactions on Mathematics*, 20, 657-675. 2021. DOI: 10.37394/23206.2021.20.70

[3] Dani, A. T. R., Adrianingsih, N. Y., Ainurrochmah, A., & Sriningsih, R. Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern. *Jurnal Litbang Edusaintech*, 2(1), 37-43. 2021. https://doi.org/10.51402/jle.v2i1.30

[4] Fernandes, A. A. R., Darmanto, Astuti, A. B., Solimun, Amaliana, L., Nurjannah, Yanti, I., Arisoesilaningsih, E., & Isaskar, R. Smoothing Spline Nonparametric Path: Application for Green Product and Green Marketing Strategy towards Green Product Purchasing Intention. *IOP Conference Series: Earth and Environmental Scienc*e, 239, 012018. 2019. DOI: 10.1088/1755-1315/239/1/012018

[5] Schumaker, L. *Spline Functions: Basic Theory*. Cambridge University Press. 2007.

[6] Dani, A. T. R., & Ni'matuzzahroh, L. Penerapan Keluarga Model Spline Truncated Polinomial pada Regresi Nonparametrik. *Inferensi*, 5(1), 37-44. 2022. http://dx.doi.org/10.12962/j27213862.v5i1.12537

[7] Härdle, W., & Liang, H. Partially Linear Models. *In Handbook of Computational Statistics* (pp. 707-737). Springer. 2007.

[8] Sari, W. (2021). Uji Efektivitas Beberapa Bahan Perangkap Organik terhadap Populasi Walang Sangit (Leptocorisa Oratorius) Pada Tanaman Padi. *Jurnal Agroscience*, *11*(1), 66-75.

[9] Saragih, I. R., Chalil, D., & Ayu, S. F.. Analisis Risiko Produksi Padi dalam Pengembangan Asuransi Usahatani Padi (AUTP) di Desa Panca Arga, Kecamatan Rawang Panca Arga, Kabupaten Asahan. *Jurnal AGRISEP: Kajian Masalah Sosial Ekonomi Pertanian dan Agribisnis*, 187-196. 2018.

[10] Khalil, A. A., Budiantara, I., & Zain, I. Comparison of linear and quadratic bi-response semiparametric regression models using spline truncated. *Journal of Physics*, 1511(1), p. 012046). 2020. DOI: 10.1088/1742-6596/1511/1/012046

[11] Nurcahayani, H., Budiantara, I. N., & Zain, I. The curve estimation of combined truncated spline and fourier series estimators for multiresponse nonparametric regression. *Mathematics*, 9(10), 1141. 2021. https://doi.org/10.3390/math9101141

[12] Di Bucchianico, A. Coefficient of determination (R2). *Encyclopedia of statistics in quality and reliability*, 1. 2008.

[13] Quinino, R. C., Reis, E. A., & Bessegato, L. F. Using the coefficient of determination R2 to test the significance of multiple linear regression. *Teaching Statistics*, 35(2), 84-88. 2013.

[14] Daoud, J. I. Multicollinearity and Regression Analysis. *Journal of Physics*, 949, 012009. 2017. https://doi.org/10.1088/1742-6596/949/1/012009

[15] Trabalon, M. Effects of wolf spiders' captive environment on their locomotor and exploratory behaviours. *Insects*, 13(2), 135. 2022. https://doi.org/10.3390/insects13020135

[16] Chellappan, M., & Ranjith, M. T. Insect Predators. *Insect Predators in Pest Management* (pp. 1-54). CRC Press. 2023.