STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, July 2025, pp 62–76.
Published online in International Academic Press (www.IAPress.org)

Semiparametric Biresponse Regression Modeling Mixed Spline Truncated, Fourier Series, and Kernel in Predicting Rainfall and Sunshine

Hartina Husain [1,*], Putri Indi Rahayu[2], Muhammad Rifki Nisardi [1], Muhammad Aslam Al-Fadhilah[1], Ahmad Husain[1]

[1] *Bacharuddin Jusuf Habibie Institute of Technology, Indonesia*
[2] *West Sulawesi University, Indonesia*

Abstract The biresponse semiparametric regression analysis combines parametric and nonparametric components to understand the relationship between two correlated response variables and predictor variables. In this approach, the nonparametric component are estimated using spline truncated, Fourier series, or kernel methods, each suitable for specific data patterns. This study aims to estimate the parameters of a mixed semiparametric regression model on climate data using the Weighted Least Square (WLS) method and to select optimal knot points, oscillation parameters, and bandwidth based on the smallest Generalized Cross Validation (GCV) value. The results show that the best model combines a spline truncated component with one-knot and a Fourier series component with one oscillation, yielding a minimum GCV of 7.401, an R^2 of 84.66%, and an MSE of 92.33. The findings suggest that the biresponse semiparametric regression model combining spline truncated, Fourier series, and kernel estimators are highly effective for modeling climate data with complex predictor patterns.

Keywords Biresponse semiparametric regression, Fourier Series , Generalized Cross Validation, Kernel, Spline Truncated

AMS 2010 subject classifications 62G08, 62J05, 62P12

DOI: 10.19139/soic-2310-5070-2166

1. Introduction

Indonesia, an archipelagic nation with rich biodiversity, faces significant challenges due to climate change. The steadily rising annual temperatures contribute to global warming, exacerbating the situation [1]. According to the 2022 report by the Meteorology, Climatology, and Geophysics Agency, over the past 40 years, Indonesia's temperatures have consistently increased, with an average air temperature anomaly of 1.2°C across 89 observation stations [2]. Paradoxically, Indonesia ranks 7th globally as the largest contributor to greenhouse gas emissions [3]. The impact of this crisis is increasingly evident in the heightened frequency and intensity of natural disasters [4]. Additionally, climate change threatens food security, degrades the environment, and escalates public health risks.

South Sulawesi and West Sulawesi are two Indonesian provinces significantly impacted by climate change. A rainfall anomaly occurred in 2023, characterized by a wetter January and a much drier June-July than previous years [5]. In January 2023, rainfall in South Sulawesi and West Sulawesi was notably higher than in January 2022. This increase in rainfall intensity poses a risk of hydrometeorological disasters, such as floods and landslides. For example, the flood in Makassar City in mid-January 2023 was recorded as the most severe in the last 20 years [6]. Conversely, a substantial reduction in rainfall could lead to severe drought, which would inevitably hinder and decrease productivity of crops requiring sufficient water supply [7].

*Correspondence to: Hartina Husain (Email: hartinahusain@ith.ac.id). Bacharuddin Jusuf Habibie Institute of Technology, Parepare, South Sulawesi Province, Indonesia (91125).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

In addition to rainfall, sunshine duration is a critical factor in the climate dynamics of South Sulawesi and West Sulawesi. Fluctuations in sunshine duration, whether through an increase or extreme decrease, can have significant ecological and environmental consequences for the region. Excessive increases in sunshine duration can raise air temperatures and dry out the soil, exacerbating existing drought conditions [8]. Conversely, extreme reductions in sunshine duration can disrupt plant growth patterns and photosynthetic activity, potentially reducing agricultural productivity and threatening food security in the region [9].

The complex and inconsistent dynamics of climate change present challenges in understanding trends and predicting future events. The uncertainty caused by extreme and irregular weather fluctuations further complicates the prediction process, hampering disaster mitigation efforts [1], [10]. Historical climate data indicate that some patterns can be modeled, while others remain unknown [1], [11]. Therefore, an adaptive predictive model is essential for accurately understanding inconsistent climate change patterns and forecasting future events.

Statistical modeling using semiparametric regression analysis offers an effective solution to address the complexity and uncertainty in climate data analysis [12]. This analysis combines the flexibility of nonparametric models with the linear structure of parametric models, allowing it to capture complex nonlinear patterns in climate data [13]. Its adaptability to unstable data patterns enhances its accuracy in predicting future data [14]. The nonparametric component of the data can be modeled using various estimators, such as spline truncated, Fourier series, or kernels [12]. This approach aligns with the characteristics of climate data, which often exhibit unknown patterns that are well-suited to semiparametric regression modeling.

Typically, research on semiparametric regression models employs a uniform estimation method for some or all predictor variables. However, data with multiple predictors often exhibit varying patterns for each predictor. Consequently, some researchers have developed mixed estimation methods to address this issue. Previous studies on semiparametric regression utilizing mixed estimators include Hesikumalasari et al. [15], which incorporated a combination of truncated spline and kernel estimators. Nisa explored a mix of truncated spline and Fourier series estimators [16]. Further research on biresponse data was conducted by Fariz et al. [17] using a Fourier series estimator and by Sauri et al. [18] employing a combined Fourier series and truncated spline estimator.

Previous research has primarily focused on the combination of two estimators in semiparametric regression. However, in other cases, it may be necessary to apply three or more estimators to different data patterns. This decision is guided by scatter plot analysis, comparing the characteristics of the truncated spline, Fourier series, and kernel estimators to determine the most appropriate modeling approach. Accordingly, this study advances the semiparametric regression model by incorporating mixed truncated spline, Fourier series, and kernel estimators, applied to data with two response variables (biresponse).

2. Literature Review

Semiparametric regression is a statistical modeling technique that combines both parametric and nonparametric components. In this approach, part of the model is specified with a predetermined functional form (parametric component), while the other part is left flexible to capture complex patterns in data without assuming a specific form (nonparametric component) [14].

Given the pairs of observational data (x, t, z, v, y_1, y_2), where x is the parametric component, t, z, v are the nonparametric components, and y_1, y_2 are response variables. The form of the biresponse semiparametric regression model that includes these variables can be expressed as follows [13]:

$$y_{ji} = f_j(x_i) + g_j(t_i) + h_j(z_i) + q_j(v_i) + \epsilon_{ji} \tag{1}$$

where i represents the number of data points, and j denotes the number of responses, with $i = 1, 2, \ldots, n$ and $j = 1, 2$. Meanwhile, ϵ_{ji} represents independent random errors that are normally distributed with a mean of zero and a variance of σ_{ji}^2.

Based on equation (1), it can be seen that the equation is a biresponse semiparametric function consisting of four components. The first function f is the parametric component, which is approximated using linear regression as follows:

$$f_j(x_i) = \beta_{0j} + \beta_j x_i \tag{2}$$

Meanwhile, the second component, g is a nonparametric function that is approximated using a linear spline truncated. This approach is used because the relationship between the response variable y_j and the predictor variables t is assumed to fluctuate within specific sub-intervals. The linear spline truncated regression curve with knots $K_1, K_2, \ldots, K_M$ $(K_1 \leq K_2 \leq \cdots \leq K_M)$ is defined by equation (3).

$$g_j(t_i) = \xi_j t_i + \sum_{k=1}^{M} \Phi_{jk}(t_i - K_{jk})_+ \tag{3}$$

where

$$(t_i - K_{jk})_+ = \begin{cases} (t_i - K_{jk}), & t_i \geq K_{jk} \\ 0, & t_i < K_{jk} \end{cases}$$

The function h is a nonparametric component that is approximated using a Fourier series function. This is based on the assumption that the relationship between the response variable y_j and the predictor variable z exhibits a repeating or seasonal pattern. The Fourier series regression curve with the oscillation parameter l can be defined by the following equation (4).

$$h_j(z_i) = b_j z_i + \frac{\alpha_{0j}}{2} + \sum_{l=1}^{L} \alpha_{jl} \cos(l z_i) \tag{4}$$

The function q is a nonparametric component that is approximated using a kernel function. The relationship pattern between the response variable y_j and the predictor variables v is assumed to be unknown. In this study, the Nadaraya-Watson kernel function is used, which is defined by the following equation (5).

$$q_j(v_i) = n^{-1} \sum_{i=1}^{n} \frac{K_{\phi_j}(v_j - v_i)}{n^{-1} \sum_{i=1}^{n} K_{\phi_j}(v_j - v_i)} y_{ji} \tag{5}$$

or

$$q_j(v_i) = n^{-1} \sum_{i=1}^{n} R_{\phi_{ji}}(v) y_{ji} \tag{6}$$

where

$$K_{\phi_j}(v_s - v_{si}) = \frac{1}{\phi_j} K\left(\frac{v - v_i}{\phi_j}\right)$$

and $q_j(v_i)$ is the Nadaraya-Watson kernel regression curve estimator, and ϕ represents the bandwidth parameter.

3. Research Method

The data used in this study are secondary data obtained from the Global Satellite Mapping of Precipitation (GSMAP) and The European Center for Medium-Range Weather Forecasts (ECMWF). The data pertains to the climate in the provinces of South Sulawesi and West Sulawesi collected from 78 data collection points from 2008 to 2013. This study uses two response variables and four predictor variables as shown in Table 1

Table 1. Research Variables

Variables	Description	Unit
y_1	Average of daily rainfall during January	mm/m^2
y_2	Average of daily sunshine duration during January	%
x_1	Average of daily wind speed during January	m⁄d
x_2	Average of daily relative humidity during January	%
x_3	Average of monthly maximum temperature before January	C
x_4	Average of monthly minimum temperature before January	C

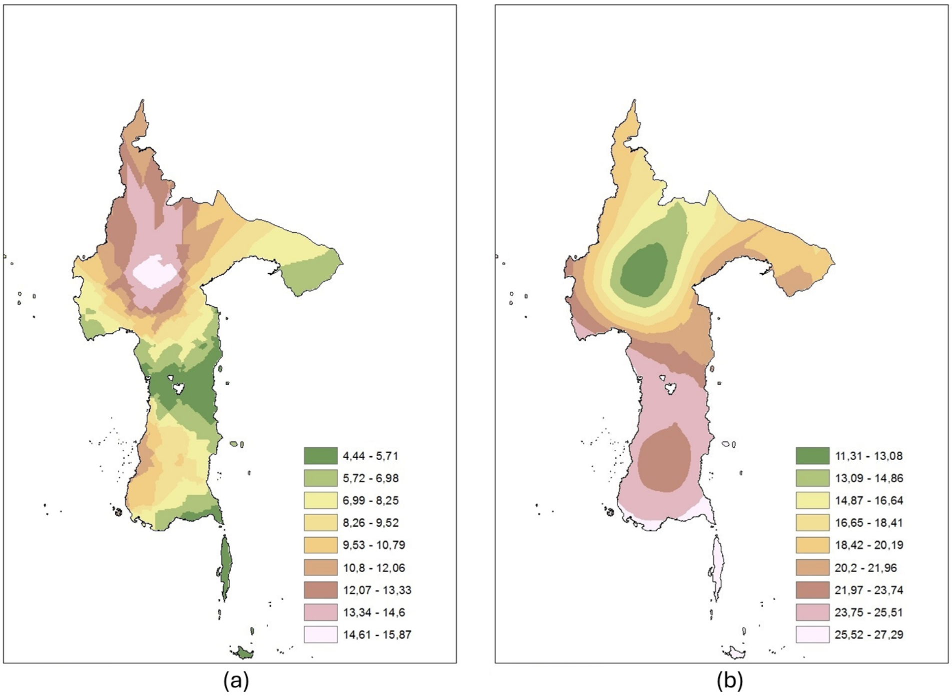

Figure 1. The response variables (a) Average of daily rainfall during January (y_1) (b) Average of daily sunshine duration during January (y_2).

To model the rainfall and sunshine duration data using semiparametric biresponse regression mixed spline truncated, Fourier series, and kernel, the following steps are taken :

1. Conduct a dependency test between the first and second responses.
2. Perform data exploration on the response variables (Rainfall and Sunshine Duration) and predictor variables (Wind Speed, Relative Humidity, Maximum Temperature, and Minimum Temperature) to determine the descriptive statistics of each research variable.
3. Create scatter plots to observe the relationship patterns between the response and predictor variables. For known relationship patterns, use parametric regression, and for unknown patterns, use nonparametric regression.
4. Determine which predictor variables should use spline truncated regression curves, Fourier series regression curves, and kernel regression curves based on the minimum Generalized Cross Validation (GCV) criterion.
5. Model the two responses, Rainfall and Sunshine Duration data, semiparametric biresponse regression mixed spline truncated, fourier series, and kernel.
6. Select the optimal knot points, oscillation parameters, and bandwidth based on the smallest GCV criterion.
7. Calculate the coefficient of determination for the obtained model.
8. Draw conclusions from the analysis results.

4. Result and Discussion

4.1. Testing Dependency of Response Variables

One of the assumptions that must be met for cases with more than one response is the dependence between the response variables. The Bartlett's Sphericity test is used to examine the dependency between the two response variables in this study, which are rainfall (y_1) and sunshine duration (y_2). The hypotheses used are:

Hypotheses:

- H_0: The response variables are independent.
- H_1: The response variables are dependent.

Based on the Bartlett's Sphericity test, $\chi^2 = 40.238 > \chi^2_{\text{table}} = 3.841$, so H_0 was rejected, indicating that the response variables are dependent. Additionally, the Pearson correlation between the two responses is -0.643. This means that there is a fairly strong inverse correlation between rainfall and sunshine duration. As the amount of rainfall at a location increases, sunshine duration at that location decreases, and vice versa.

4.2. Descriptive Analysis

The initial step before performing the mixed biresponse semiparametric regression modeling with truncated spline, Fourier series, and kernel is to conduct a descriptive analysis. This analysis aims to provide an overview of the variables used in the study, as shown in Table 2.

Table 2. Descriptive Analysis

Variables	Minimum	Maximum	Mean	Standard Deviation
y_1	0.541	25.194	8.292	5.959
y_2	8.266	27.296	22.869	5.020
x_1	0.944	6.156	2.692	1.474
x_2	0.651	0.961	0.798	0.072
x_3	20.242	32.122	27.900	2.643
x_4	14.193	28.012	23.298	4.190

Table 2 shows the distribution of variables from January 2008 to 2013 at 78 observation points in the Provinces of South Sulawesi and West Sulawesi. The average rainfall ranged from the lowest of 0.541 mm/m² observed in Luwu Regency, to the highest of 25.194 mm/m² observed around Takalar Regency. The maximum variability in sunshine duration was about 27.296 hours in central and southern South Sulawesi; the lowest value, however, was 8.266 hours in Tana Toraja and North Toraja. Wind speed varied from a maximum of 6.156 m/day in southern Selayar Regency to a minimum of 0.944 m/day in Luwu Regency. This description of rainfall and sunshine duration corresponds with the pattern shown in figure 1 . The highest relative humidity was 96.1% found in Luwu, Tana Toraja, and in the mountainous area of Gowa Regency, while the minimum was 65.1% in Mamuju Regency. The highest average maximum temperature was 32.122°C in Bone Regency, while the lowest, 20.242°C, was in Tana Toraja. The highest average minimum temperature was 28.012°C in Bulukumba and Selayar, and the lowest, 14.193°C, in Tana Toraja.

The scatter plots provided the preliminary relationships between the response and predictor variables, guiding whether to use a parametric or non-parametric approach — i.e., truncated spline, Fourier series, and kernel techniques — based on the theoretical data shape. Comparisons between scatter plots of rainfall and sunshine duration against their predictors of interest are shown below.

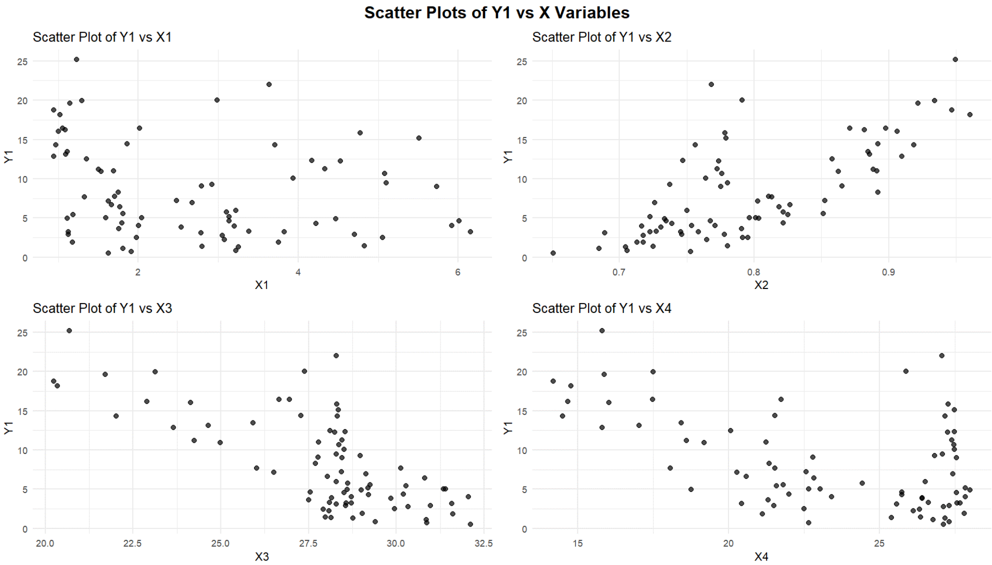

Figure 2. Scatter plots of Y_1 vs X variables

Figure 2 illustrates the scatter plot between rainfall (Y_1) and wind speed (X_1), which shows no clear pattern, suggesting a nonparametric kernel approach is suitable. The relationship between rainfall (Y_1) and relative humidity (X_2) shows a recurring trend as humidity increases with rainfall, indicating a Fourier series approach. The relationship between rainfall (Y_1) and maximum temperature (X_3) appears linear, making a linear parametric approach appropriate. Meanwhile, the relationship between rainfall (Y_1) and minimum temperature (X_4) changes behavior at certain intervals, suggesting a linear spline truncated approach.

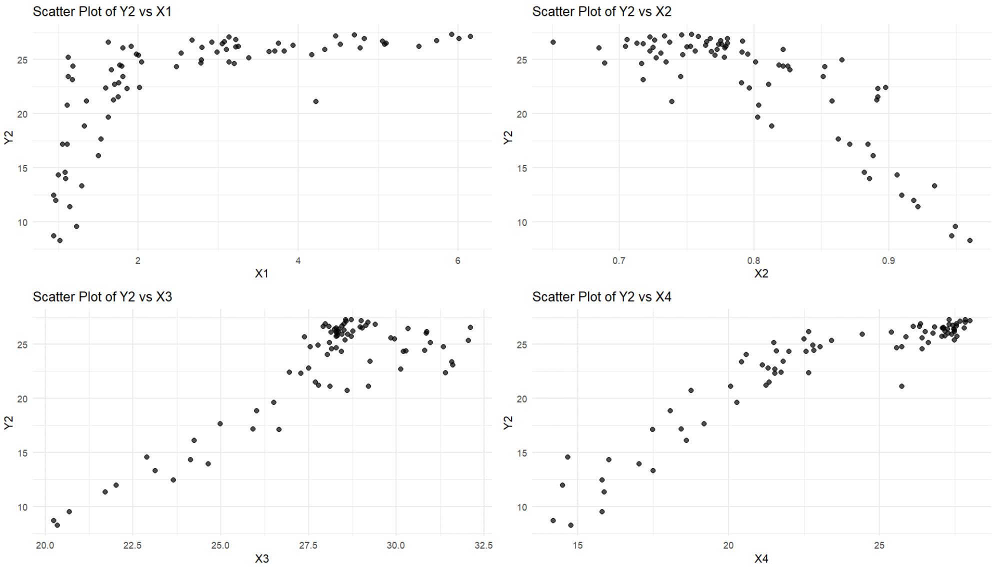

Figure 3. Scatter plots of Y_2 vs X variables

Figure 3 shows that the relationship between sunshine duration (Y_2) and wind speed (X_1) changes at certain intervals, with a sharp increase before interval 2 and a slower rate afterward. This suggests modeling with a nonparametric truncated spline. Similar patterns are observed for relative humidity (X_2) and minimum temperature

(X_4). In contrast, the relationship between sunshine duration (Y_2) and maximum temperature (X_3) is linear, suitable for a linear parametric approach.

Both rainfall and sunshine duration are linearly related to maximum temperature, indicating a linear parametric model for this predictor. However, other predictors, like wind speed, show different patterns with each response, suggesting different nonparametric approaches. For instance, rainfall and wind speed are modeled using a nonparametric kernel, while sunshine duration and wind speed use a truncated spline. The choice of nonparametric components — truncated spline, Fourier series, or kernel — is based on the smallest GCV value, considering only the nonparametric predictors X_1, X_2, and X_4, as X_3 follows a parametric approach.

Table 3. Comparison of GCV Values for Different Variables

No	Variables Spline Truncated	Variables Fourier Series	Variables Kernel	CGV
1	X_1	X_2	X_4	8.929
2	X_1	X_4	X_2	9.019
3	X_2	X_1	X_4	8.394
4	X_2	X_4	X_1	7.829
5	X_4	X_1	X_2	9.143
6	X_4	X_2	X_1	7.401*

Based on Table 3, the smallest GCV value is 7.401, with X_4 modeled using a truncated spline approach, X_2 using a Fourier series approach, and X_1 using a kernel approach. Variables modeled with parametric components are symbolized by x, while nonparametric components modeled with truncated spline, Fourier series, and kernel are symbolized by t, z, and v, respectively.

4.3. *Semiparametric Biresponse Regression Modeling with Mixed Estimator of Truncated Spline, Fourier Series, and Kernel*

The semiparametric biresponse regression modeling using mixed estimators of truncated spline, Fourier series, and kernel heavily relies on the selection of knot points, oscillation parameters, and optimal bandwidth. In accordance with the research limitations stated in Chapter 1, which restrict the study to one-knot point and two oscillation parameters ($l = 1, 2,$ and 3), the selection of knot points, oscillation parameters, and bandwidth is based on the smallest GCV value.

4.3.1. Model with One-knot Point for Spline truncated Component and One Oscillation for Fourier Series Component The following is the semiparametric mixed regression model of Spline truncated, Fourier series, and kernel using one-knot point and one oscillation.

$$\hat{y}_{ji} = \hat{\beta}_{0j} + \hat{\beta}_j x_i + \hat{\xi}_j t_i + \hat{\Phi}_{j1}(t_i - K_{j1})_+ + \hat{b}_j z_i + \frac{\hat{\alpha}_{0j}}{2} + \hat{\alpha}_{j1}\cos(z_i) + \frac{\sum_{i=1}^{n} \frac{1}{\varphi_j} K\left(\frac{v - v_i}{\varphi_j}\right)}{\sum_{i=1}^{n} \frac{1}{\varphi_j} K\left(\frac{v - v_i}{\varphi_j}\right)} y_{ji}$$

The GCV values produced using one-knot point and one oscillation are shown in Table 4.

Table 4. Comparison of GCV Values of Models with one-knot Point and One Oscillation

No.	Knot Point K_{11}	Knot Point K_{21}	Oscillation	Bandwidth φ_1	Bandwidth φ_2	GCV
1	22.987	19.644	1	3.018	2.730	7.401*
2	22.847	19.411	1	3.018	2.730	7.402
3	23.126	19.876	1	3.018	2.730	7.403
4	22.987	19.643	1	3.292	2.978	7.404
5	22.847	19.411	1	3.292	2.978	7.404
6	22.987	19.644	1	2.743	2.482	7.405
7	22.847	19.411	1	2.743	2.482	7.407
8	23.126	19.876	1	3.292	2.978	7.407
9	23.126	19.876	1	2.743	2.482	7.407
10	23.266	20.109	1	3.018	2.730	7.409

Based on Table 4, the smallest GCV value is 7.401 with the location of the knot point in the first response and the second response at 22.987 and 19.644, respectively. The bandwidth values in the first and second responses are 3.018 and 2.730, respectively. By using the knot point, oscillation parameters, and bandwidth, the parameter estimates for the parametric and nonparametric components are obtained as shown in Table 5.

Table 5. Model Parameter Estimation with one-knot Point and One Oscillation

Variable	Parameter	Estimate
x	$\hat{\beta}_{01}$	9.659×10^{-9}
	$\hat{\beta}_1$	-1.088
	$\hat{\beta}_{02}$	-8.038×10^{-8}
	$\hat{\beta}_2$	0.793
t	$\hat{\xi}_1$	0.631
	$\hat{\Phi}_{11}$	-0.690
	$\hat{\xi}_2$	-0.670
	$\hat{\Phi}_{21}$	81.164
z	$\hat{b}_1$	63.558
	$\hat{\alpha}_{01}$	-95.799
	$\hat{\alpha}_{11}$	8.972
	$\hat{b}_{21}$	553.136
	$\hat{\alpha}_{02}$	-2058.806
	$\hat{\alpha}_{21}$	844.715

Based on Table 4 and the estimation results in Table 5, the following semiparametric biresponse regression model of a mixture of Spline truncated, Fourier series, and kernel with one-knot point and one oscillation is obtained:

$$\hat{y}_{1i} = -47.899 - 1.088x_i + 0.631t_i - 0.690(t_i - 22.987)_+ + 63.558z_i$$
$$+ 8.972\cos(z_i) + \frac{\sum_{i=1}^{78} \frac{1}{3.018} K\left(\frac{v-v_i}{3.018}\right)}{\sum_{i=1}^{78} \frac{1}{3.018} K\left(\frac{v-v_i}{3.018}\right)} y_{1i}$$
$$\hat{y}_{2i} = -1029.403 + 0.793x_i - 0.670t_i + 81.164(t_i - 19.644)_+ + 553.136z_i$$
$$+ 844.715\cos(z_i) + \frac{\sum_{i=1}^{78} \frac{1}{2.730} K\left(\frac{v-v_i}{2.730}\right)}{\sum_{i=1}^{78} \frac{1}{2.730} K\left(\frac{v-v_i}{2.730}\right)} y_{2i}$$

4.3.2. Model with One-Point Knot Spline truncated Components and Two-Oscillation Fourier Series Components. The following is a semiparametric regression model of a mixture of Spline truncated, Fourier series, and kernel using one-knot point and two oscillations.

$$\hat{y}_{ji} = \hat{\beta}_{0j} + \hat{\beta}_j x_i + \hat{\xi}_j t_i + \hat{\Phi}_{j1}(t_i - K_{j1})_+ + \hat{b}_j z_i + \frac{\hat{\alpha}_{0j}}{2} + \hat{\alpha}_{j1}\cos(z_i) + \hat{\alpha}_{j2}\cos(2z_i) + \frac{\sum_{i=1}^{n} \frac{1}{\varphi_j} K\left(\frac{v-v_i}{\varphi_j}\right)}{\sum_{i=1}^{n} \frac{1}{\varphi_j} K\left(\frac{v-v_i}{\varphi_j}\right)} y_{ji}$$

The GCV values produced using one-knot point and two oscillations are shown in Table 6.

Table 6. Comparison of GCV Values of Models with one-knot Point and Two Oscillations

No.	Knot Point K_{11}	Knot Point K_{21}	Oscillation	Bandwidth φ_1	Bandwidth φ_2	GCV
1	23.824	21.038	2	2.743	2.482	7.629*
2	23.964	21.271	2	2.743	2.482	7.630
3	23.685	20.806	2	2.743	2.482	7.630
4	24.103	21.503	2	2.743	2.482	7.630
5	23.824	21.038	2	3.018	2.730	7.633
6	23.684	20.806	2	3.018	2.730	7.633
7	23.964	21.271	2	3.018	2.730	7.634
8	23.545	20.574	2	2.743	2.482	7.634
9	24.103	21.503	2	3.018	2.730	7.635
10	23.824	21.038	2	2.469	2.234	7.636

Based on Table 6, the smallest GCV value is 7.629 with the location of the knot point in the first response and the second response at 23.824 and 21.038, respectively. The bandwidth values in the first and second responses are 2.743 and 2.482, respectively. By using the knot point, oscillation parameters, and bandwidth, the parameter estimates for the parametric and nonparametric components are obtained as shown in Table 7.

Table 7. Model Parameter Estimation with one-knot Point and Two Oscillations

Variable	Parameter	Estimate
x	$\hat{\beta}_{01}$	7.213
	$\hat{\beta}_1$	-1.047
	$\hat{\beta}_{02}$	-27.785
	$\hat{\beta}_2$	0.698
t	$\hat{\xi}_1$	0.663
	$\hat{\Phi}_{11}$	0.644
	$\hat{\xi}_2$	-8.665
	$\hat{\Phi}_{21}$	0.794
z	$\hat{b}_1$	548.091
	$\hat{\alpha}_{01}$	-916.420
	$\hat{\alpha}_{11}$	38.144
	$\hat{\alpha}_{12}$	232.547
	$\hat{b}_2$	-904.788
	$\hat{\alpha}_{02}$	1092.777
	$\hat{\alpha}_{21}$	233.025
	$\hat{\alpha}_{22}$	-540.312

Based on Table 6 and the estimation results in Table 7, the following semiparametric biresponse regression model of a mixture of Spline truncated, Fourier series, and kernels with one-knot point and two oscillations is obtained:

$$\hat{y}_{1i} = -450.997 - 1.047x_i + 0.663t_i + 0.644(t_i - 23.824)_+ + 548.091z_i + 38.144\cos(z_i) + 232.547\cos(2z_i) + \frac{\sum_{i=1}^{78} \frac{1}{2.743} K\left(\frac{v-v_i}{2.743}\right)}{\sum_{i=1}^{78} \frac{1}{2.743} K\left(\frac{v-v_i}{2.743}\right)} y_{1i}$$

$$\hat{y}_{2i} = 518.604 + 0.698x_i - 8.665t_i + 0.794(t_i - 21.038)_+ - 904.788z_i + 233.025\cos(z_i) - 540.312\cos(2z_i) + \frac{\sum_{i=1}^{78} \frac{1}{2.482} K\left(\frac{v-v_i}{2.482}\right)}{\sum_{i=1}^{78} \frac{1}{2.482} K\left(\frac{v-v_i}{2.482}\right)} y_{2i}$$

4.3.3. Model with One-Point Knot Spline truncated Components and Three-Oscillation Fourier Series Components
The following is a semiparametric regression model of a mixture of Spline truncated, Fourier series, and kernels using one-knot point and three oscillations.

$$\hat{y}_{ji} = \hat{\beta}_{0j} + \hat{\beta}_j x_i + \hat{\xi}_j t_i + \hat{\Phi}_{j1}(t_i - K_{j1})_+ + \hat{b}_j z_i + \frac{\hat{\alpha}_{0j}}{2} + \hat{\alpha}_{j1}\cos(z_i) + \hat{\alpha}_{j2}\cos(2z_i) + \hat{\alpha}_{j3}\cos(3z_i) + \frac{\sum_{i=1}^{n} \frac{1}{\hat{\varphi}_j} K\left(\frac{v-v_i}{\hat{\varphi}_j}\right)}{\sum_{i=1}^{n} \frac{1}{\hat{\varphi}_j} K\left(\frac{v-v_i}{\hat{\varphi}_j}\right)} \hat{y}_{ji}$$

The GCV values produced using one-knot point and three oscillations are shown in Table 8.

Table 8. Comparison of GCV Values of Models with one-knot Point and Three Oscillations

No.	Knot Point K_{11}	Knot Point K_{21}	Oscillation	Bandwidth φ_1	Bandwidth φ_2	GCV
1	27.453	27.082	3	0.823	0.745	7.806*
2	19.637	14.065	3	1.097	0.993	7.882
3	19.497	13.833	3	1.097	0.993	7.885
4	19.358	13.600	3	1.097	0.993	7.889
5	19.218	13.368	3	1.097	0.993	7.896
6	19.079	13.135	3	1.097	0.993	7.904
7	18.939	12.903	3	1.097	0.993	7.913
8	27.453	27.082	3	1.097	0.993	7.914
9	18.799	12.671	3	1.097	0.993	7.924
10	18.660	12.438	3	1.097	0.993	7.934

Based on Table 8, the smallest GCV value is 7.806 with the location of the knot point in the first response and the second response respectively 27.453 and 27.082. The bandwidth value in the first and second responses is also obtained as 0.823 and 0.745 respectively. By using the knot point, oscillation parameters, and bandwidth, the parameter estimates for the parametric and nonparametric components are obtained as in Table 9.

Table 9. Model Parameter Estimation with one-knot Point and Three Oscillations

Variable	Parameter	Estimate
x	$\hat{\beta}_{01}$	0.665
	$\hat{\beta}_1$	-1.046
	$\hat{\beta}_{02}$	-27.977
	$\hat{\beta}_2$	0.698
t	$\hat{\xi}_1$	0.300
	$\hat{\Phi}_{11}$	0.634
	$\hat{\xi}_2$	-115.01
	$\hat{\Phi}_{21}$	-0.057
z	$\hat{b}_1$	1155.339
	$\hat{\alpha}_{01}$	-611.313
	$\hat{\alpha}_{11}$	-966.563
	$\hat{\alpha}_{12}$	1047.841
	$\hat{\alpha}_{13}$	-136.610
	$\hat{b}_2$	-777.150
	$\hat{\alpha}_{02}$	389.865
	$\hat{\alpha}_{21}$	633.263
	$\hat{\alpha}_{22}$	-664.728
	$\hat{\alpha}_{23}$	44.376

Based on Table 8 and the estimation results in Table 9, a semiparametric biresponse regression model of a mixture of Spline truncated, Fourier series, and kernel with one-knot point and three oscillations is obtained as follows:

$$\begin{aligned}\hat{y}_{1i} &= -304.992 - 1.046x_i + 0.300t_i + 0.634(t_i - 27.453)_+ + 1155.339z_i \\ &\quad - 966.563\cos(z_i) + 1047.841\cos(2z_i) - 136.610\cos(3z_i) \\ &\quad + \frac{\sum_{i=1}^{78} \frac{1}{0.823} K\left(\frac{v-v_i}{0.823}\right)}{\sum_{i=1}^{78} \frac{1}{0.823} K\left(\frac{v-v_i}{0.823}\right)} y_{1i} \\ \hat{y}_{2i} &= 166.954 + 0.698x_i - 115.01t_i - 0.057(t_i - 27.082)_+ \\ &\quad - 777.150z_i + 633.263\cos(z_i) - 664.728\cos(2z_i) + 44.376\cos(3z_i) \\ &\quad + \frac{\sum_{i=1}^{78} \frac{1}{0.745} K\left(\frac{v-v_i}{0.745}\right)}{\sum_{i=1}^{78} \frac{1}{0.745} K\left(\frac{v-v_i}{0.745}\right)} y_{2i}\end{aligned}$$

Next, a comparison of the GCV values for the Model with one-knot Spline Truncated Component and Fourier Series Component for 1, 2, and 3 oscillations can be seen in Table 10.

Table 10. Comparison of GCV Values of Models with one-knot Point and 1, 2, 3 Oscillations

No	Oscillation Parameter	GCV
1	1 oscillation	7.401*
2	2 oscillation	7.629
3	3 oscillation	7.806

Based on Table 10, the smallest GCV value is 7.401 in the model with one-knot spline truncated component and one oscillation Fourier series component. From the model obtained, a plot is obtained between the y value and the $\hat{y}$ for each response, namely rainfall and sunshine duration in South Sulawesi and West Sulawesi Province. The plot of actual rainfall data (y_1) and the plot of predicted data ($\hat{y}_1$) are shown in Figure 4. and The plot of actual solar radiation duration data (y_2) and the plot of predicted data data ($\hat{y}_2$) are shown in Figure 5.

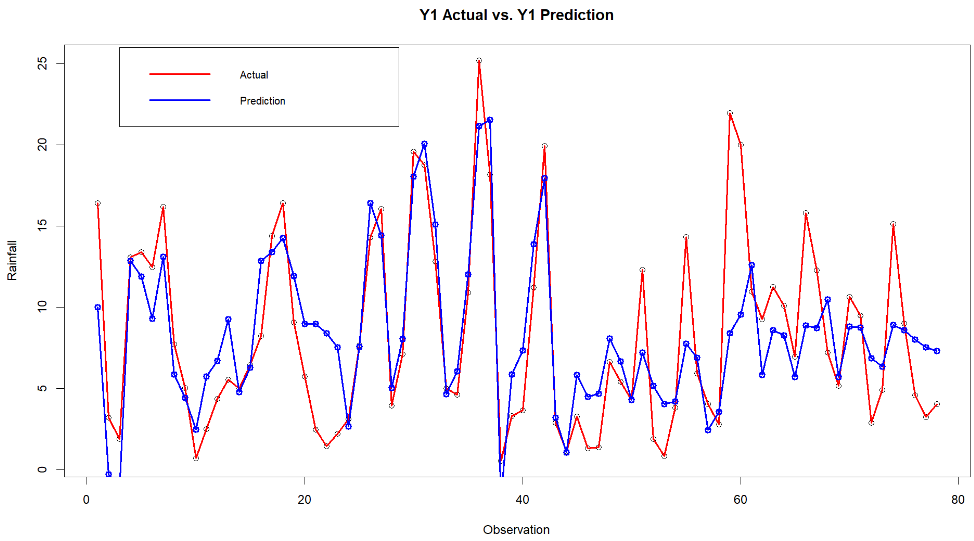

Figure 4. Plot of data y_1 and $\hat{y}_1$.

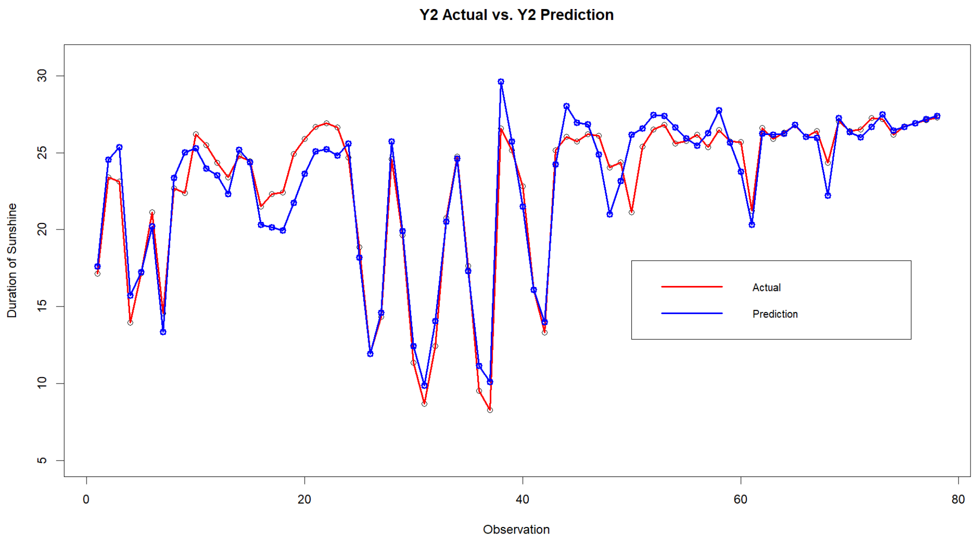

Figure 5. Plot of data y_2 and $\hat{y}_2$.

Figures 4 and 5 provide a visual comparison between the predicted and actual values for both rainfall and sunshine duration, revealing that the predictions closely approximate the actual data. This alignment is quantitatively supported by a coefficient of determination (R^2) of 84.66%, indicating that a significant proportion of the variance in the response variables is explained by the model. Additionally, the model's mean square error (MSE) of 92.33 further underscores its accuracy, as this relatively low error value suggests that the differences between predicted and actual values are minimal. The combination of a high R^2 and a low MSE reflects the robustness of the semiparametric model employed, affirming its effectiveness in capturing the underlying patterns in the climate data. Consequently, these findings illustrate that the model is well-suited for predicting both rainfall and sunshine duration, making it a reliable tool for understanding and forecasting climate-related variables in the studied regions.

Semiparametric biresponse mixed Spline Truncated, Fourier series, and kernels can be compared with models that only use single estimator. The model was also tested on the same dataset using single estimator: Spline truncated, Fourier series, and kernels. Table 11 shows a comparison of GCV values between models with mixed estimators and each single estimator.

Table 11. Comparison of GCV Values of Mixed Model and Single Estimator

No	Model	GCV
1	Mixed Estimator	7.401*
2	Spline Truncated	15.457
3	Fourier Series	36,159

From Table 11, it can be seen that the GCV value for the model using a combination of three estimators is lower than the models that only use one estimator. This shows that the mixed model of three estimators is superior in modeling the average data of Rainfall and Sunshine Duration in South Sulawesi and West Sulawesi Provinces.

5. Conclusion

The best model resulting from the selection of knot points, oscillation parameters, and optimum bandwidth with the smallest GCV value is a model with one-knot truncated spline component and one oscillation Fourier series component:

$$\hat{y}_{1i} = -47.899 - 1.088x_i + 0.631t_i - 0.690(t_i - 22.987)_+ + 63.558z_i + 8.972\cos(z_i) + \frac{\sum_{i=1}^{78} \frac{1}{3.018} K\left(\frac{v-v_i}{3.018}\right)}{\sum_{i=1}^{78} \frac{1}{3.018} K\left(\frac{v-v_i}{3.018}\right)} y_{1i}$$

$$\hat{y}_{2i} = -1029.403 + 0.793x_i - 0.670t_i + 81.164(t_i - 19.644)_+ + 553.136z_i + 844.715\cos(z_i) + \frac{\sum_{i=1}^{78} \frac{1}{2.730} K\left(\frac{v-v_i}{2.730}\right)}{\sum_{i=1}^{78} \frac{1}{2.730} K\left(\frac{v-v_i}{2.730}\right)} y_{2i}$$

The regression model above has a GCV value of 7.401, a coefficient of determination (R^2) of 84.66%, and a mean squared error (MSE) of 92.33. Therefore, the semiparametric biresponse mixed spline truncated regression model, Fourier series, and kernel are able to model and explain the relationship between rainfall and sunshine duration in South Sulawesi and West Sulawesi Province with variables of wind speed, relative humidity, maximum temperature, and minimum temperature of 84.66%..

Acknowledgment

This research was specifically funded by the Government of Indonesia, with primary support from the Directorate General of Higher Education, Research, and Technology, as well as the Ministry of Education, Culture, Research, and Technology. We extend our gratitude to all parties who have provided support and input during the preparation of this research. All contributions have been invaluable in enhancing the quality of this study.

References

[1] S. Aripiyanto, D. Khairani, and A. Hartono, "Co2 emission forecasting in indonesia until 2030: Evaluation of ets smoothing prediction models and their implications for global climate change mitigation," *Journal of Applied Data Sciences*, vol. 5, no. 1, pp. 46–55, 2024. DOI: 10.47738/jads.v5i1.154.

[2] BMKG, "Informasi iklim bmkg untuk kenaikan suhu dan perkembangan iklim," Tech. Rep., 2022. [Online]. Available: https://dataonline.bmkg.go.id/.

[3] D. Hafidz, A. Wardhana, and M. R. Prawira, "The analysis of indonesia's climate change policies in response to the 2021 intergovernmental panel on climate change (ipcc) assessment report / ar6 group 1 (2021-2023)," *Unknown Journal*, vol. 1, pp. 42–53, 2023.

[4] T. Neya, G. Yanon, J. Soubeiga, R. Kiribou, O. Neya, and J. Magistro, "Climate change impact assessment and disaster risk financing strategies in mali: A comprehensive analysis of drought and flood events," *International Journal of Environment and Climate Change*, vol. 14, no. 3, pp. 126–138, 2024. DOI: 10.9734/ijecc/2024/v14i34025.

[5] Indonesian Agency for Meteorological Climatological and Geophysics, "Indonesia air quality annual report 2023," Tech. Rep., 2023, pp. 1–28. [Online]. Available: https://iklim.bmkg.go.id/bmkgadmin/storage/buletin/BMKG_20Climate_20Outlook_202023.pdf.

[6] N. Ainun. "Banjir makassar meluas, warga mengungsi bertambah jadi 482 orang." Retrieved March 29, 2024. (2023), [Online]. Available: https://www.detik.com/sulsel/berita/d-7147863/banjir-makassar-meluas-warga-mengungsi-bertambah-jadi-482-orang#:~:text=Berdasarkan%20data%20Badan%20Penanggulangan%20Bencana,dan%2017%20jiwa%20dari%20Biringkanaya.

[7] A. Yamin and J. Wiratmo, *Klasifikasi wilayah rawan kekeringan pada lahan sawah non irigasi di sulawesi selatan*, 2020. [Online]. Available: https://www.researchgate.net/publication/345387748_KLASIFIKASI_WILAYAH_RAWAN_KEKERINGAN_PADA_LAHAN_SAWAH_NON_IRIGASI_DI_SULAWESI_SELATAN.

[8] C. Zhang, Y. Zhang, J. Pu, Z. Liu, Z. Wang, and L. Wang, "An hourly solar radiation prediction model using extreme gradient boosting algorithm with the effect of fog-haze," *Energy and Built Environment*, Apr. 2023. DOI: 10.1016/j.enbenv.2023.08.001.

[9] M. Ehteram and H. Shabanian, "Unveiling the salstm-m5t model and its python implementation for precise solar radiation prediction," *Energy Reports*, vol. 10, pp. 3402–3417, 2023. DOI: 10.1016/j.egyr.2023.10.029.

[10] E. P. Prasetya and E. C. Djamal, "Rainfall forecasting for the natural disasters preparation using recurrent neural networks," in *Proceedings of the International Conference on Electrical Engineering and Informatics*, 2019, pp. 52–57. DOI: 10.1109/ICEEI47359.2019.8988838.

[11] M. Ramli, I. N. Budiantara, and V. Ratnasari, "A method for parameter hypothesis testing in nonparametric regression with fourier series approach," *MethodsX*, vol. 11, p. 102 468, 2023. DOI: 10.1016/j.mex.2023.102468.

[12] L. Hidayati, N. Chamidah, and I. N. Budiantara, "Spline truncated estimator in multiresponse semiparametric regression model for computer based national exam in west nusa tenggara," *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, p. 052 029, 2019. DOI: 10.1088/1757-899X/546/5/052029.

[13] H. Husain, I. N. Budiantara, and I. Zain, "Mixed estimator of spline truncated, fourier series, and kernel in biresponse semiparametric regression model," *IOP Conference Series: Earth and Environmental Science*, vol. 880, no. 1, p. 012 046, 2021. DOI: 10.1088/1755-1315/880/1/012046.

[14] B. Lestari, N. Chamidah, I. N. Budiantara, and D. Aydin, "Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator," *Journal of King Saud University - Science*, vol. 35, no. 5, p. 102 664, 2023. DOI: 10.1016/j.jksus.2023.102664.

[15] Hesikumalasari, I. N. Budiantara, and V. Ratnasari, "Pemodelan regresi semiparametrik menggunakan estimator campuran spline truncated dan kernel," *Institut Teknologi Sepuluh Nopember*, pp. 1–8, 2016.

[16] K. Nisa, "Model regresi semiparametrik campuran spline truncated dan deret fourier (studi kasus: Angka harapan hidup provinsi jawa timur)," *Unknown Journal*, pp. 1–120, 2017.

[17] M. Fariz, F. Mardianto, P. Supervisor, N. Budiantara, and M. Si, "Semiparametric regression model of biresponse using fourier series," *Unknown Journal*, 2015.

[18] M. S. Sauri, M. Hadijati, and N. Fitriyani, "Spline and kernel mixed nonparametric regression for malnourished children model in west nusa tenggara," *Jurnal Varian*, vol. 4, no. 2, pp. 99–108, 2021. DOI: 10.30812/varian.v4i2.1003.