Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:69
https://doi.org/10.28919/cmbn/9853
ISSN: 2052-2541

MODELING OF SHORTWAVE RADIATION IN SEMARANG USING SEMIPARAMETRIC FOURIER SERIES REGRESSION

TIANI WAHYU UTAMI[1,*], ELSA NUDYAWATI[1], IHSAN FATHONI AMRI[2], ALWAN FADLUROHMAN[2]

[1]Department of Statistics, Faculty of Science and Agriculture Technology, Universitas Muhammadiyah Semarang, Central Java 50273, Indonesia

[2]Department of Data Science, Faculty of Science and Agriculture Technology, Universitas Muhammadiyah Semarang, Central Java 50273, Indonesia

Abstract: This study presents semiparametric Fourier series regression model to analyze shortwave radiation (SWR), photosynthetically active radiation (PAR), and time in Semarang. The semiparametric approach combines the flexibility of nonparametric regression with the structure of parametric models, allowing for better representation of complex seasonal patterns in solar radiation data. The nonparametric component was modeled using a cosine-based Fourier series, which effectively captures the periodic nature of the observations. Data simulation was conducted by generating several training and testing data splits (60:40, 70:30, 80:20, and 90:10) to identify the optimal proportion for model estimation. Based on the Generalized Cross Validation (GCV) criterion, the 90:10 split with one Fourier coefficient produced the most efficient model. Model evaluation using actual SWR and PAR data yielded an in-sample coefficient of determination (R^2) of 0.9938, indicating that the model explains nearly all data variability, and an out-sample Mean Absolute Percentage Error (MAPE) of 0.9003%, reflecting very high predictive accuracy. These results

*Corresponding author
E-mail address: tianiutami@unimus.ac.id
Received March 03, 2026

demonstrate that the semiparametric Fourier series regression can accurately predict shortwave radiation based on PAR and time, providing valuable insights for agricultural planning and the development of solar radiation–based renewable energy in Semarang.

Keywords: semiparametric regression; Fourier series; shortwave radiation; photosynthetically active radiation.

2020 AMS Subject Classification: 62G08.

1. INTRODUCTION

Indonesia, as a tropical country located around the equator, receives high solar radiation intensity throughout the year. This radiation plays a crucial role as the primary energy source regulating climate, hydrological cycles, and vegetation growth [1]. One of the main components of solar radiation is Shortwave Radiation (SWR), with wavelengths ranging from 300–3000 nm, which produces heat when it reaches the Earth's surface. The intensity of SWR fluctuates due to environmental and atmospheric factors such as cloud cover, humidity, and aerosol particles [2]. These fluctuations complicate the analysis of the relationship between SWR and other components, one of which is Photosynthetically Active Radiation (PAR). PAR represents light with wavelengths between 400–700 nm, used by plants for photosynthesis to convert solar energy into chemical energy stored as glucose. As part of SWR, PAR has a strong relationship with the total solar radiation received in a region. Howell et al. [3] found a strong linear relationship between PAR and SWR, with a coefficient of determination (R^2) greater than 99%, indicating that variations in PAR can almost entirely be explained by changes in SWR.

As one of Indonesia's metropolitan cities, Semarang exhibits a tropical climate characterized by distinct wet and dry seasons that influence solar radiation intensity [4]. Solar radiation is a major factor in the climate system, where weather phenomena arise due to variations in the distribution and intensity of solar energy [5]. Such variations potentially alter the relationship pattern between PAR and SWR. Understanding this relationship is crucial for enhancing renewable energy efficiency, optimizing agricultural productivity, and supporting sustainable environmental management. The relationship between SWR and PAR in Semarang is analyzed using regression methods, which assess the functional relationship between response and predictor variables [6]. Regression analysis can be categorized into two main types: parametric and nonparametric. Parametric regression assumes specific functional form such as linear or quadratic, while nonparametric regression offers greater flexibility since it does not rely on distributional assumptions [7]. Several smoothing techniques are commonly used in nonparametric regression, including local linear estimators, Fourier series, local polynomial, spline, and kernel estimators.

In some cases, datasets involve more than one predictor variable, with some functional relationships known and others unknown—forming what is known as a semiparametric regression model. This model integrates both parametric and nonparametric regression approaches, combining their strengths in capturing complex relationships between response and predictor variables. Numerous studies have explored semiparametric regression models using various estimation techniques, such as kernel estimators [8], [9], local linear estimators [10], [11], local polynomial estimators [12], [13], spline estimators [6], [14], and Fourier series estimators [15], [16]. Among these, the Fourier series estimator is particularly advantageous due to its ability to model sinusoidal and cosinusoidal data patterns, whether trending or not [17]. This method is suitable for cyclic or periodic data, where predictor variables exhibit repeating variations [18], [19]. Furthermore, semiparametric regression based on Fourier series can effectively capture seasonal fluctuations with high accuracy without requiring strict distributional assumptions. Fourier series are highly flexible since any periodic function can be expanded into a sum of harmonic functions composed of sine and cosine components representing periodic behavior [20].

In this study, the fluctuating nature of Shortwave Radiation (SWR) data serves as the main rationale for employing the Fourier series estimator. This model is considered the most appropriate for characterizing seasonal patterns and periodic changes in solar radiation intensity. The trigonometric sine and cosine functions in the Fourier estimator allow flexible adaptation to both cyclical patterns and short-term fluctuations in Semarang's radiation data. The study focuses on analyzing SWR variations throughout 2024, particularly the significant decrease observed in the early months of the year. Such a decline may have substantial environmental implications, especially for photosynthesis, which relies heavily on sunlight. A decrease in SWR intensity could slow plant growth, reduce crop yields, and affect surface temperature, weather patterns, and the overall energy balance in the Semarang area. Therefore, applying semiparametric regression with Fourier series estimator is expected to provide a deeper understanding of shortwave radiation dynamics and their environmental implications in tropical regions.

This research uses Shortwave Radiation (SWR) data from January to June 2024. The SWR variable over time is treated as a nonparametric component, while SWR and Photosynthetically Active Radiation (PAR) are considered parametric components. Based on the assumption that both parametric and nonparametric components exist, a semiparametric regression approach is applied. The method was chosen because radiation data exhibit strong fluctuations due to variations in solar intensity and weather changes throughout the year, making the Fourier series estimator highly suitable for representing such periodic patterns. The semiparametric regression approach with the Fourier estimator enables a more comprehensive analysis of complex environmental data such as solar radiation in Semarang. The objectives of this study are to analyze the relationship between PAR and SWR in Semarang City, identify their interaction patterns, assess fluctuations, and construct a statistical model for practical applications. Through this approach, the study aims to provide clearer insights into solar radiation dynamics in tropical regions. The results are expected to contribute significantly to understanding solar radiation patterns in Semarang and to serve as a reference for similar research in regions with comparable climatic characteristics. With an accurate and relevant model, this study may also support data-driven decision-making in agriculture, renewable energy, and sustainable environmental management.

2. PRELIMINARIES

Previous studies have conducted estimations of semiparametric regression models employing Fourier series with a Partial Least Squares (PLS) approach [21]. It is assumed that the relationship between y_i, x_i and t_i follows a semiparametric regression model, with paired data (y_i, x_i, t_i), and can be expressed as follows [15]:

$$y = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \ldots + \beta_p u_p + g(t) + \varepsilon \tag{1}$$

Where y_i is the response variable for the i-th observation, $x_1, x_2, \ldots, x_p$ are predictor variables representing the parametric component, $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ are the corresponding parametric parameters, $g(t_i)$ is the nonparametric regression function, and ε_i is a random error with assumed to be normally distributed with a mean of zero and variance σ^2.

Explicitly, the semiparametric regression model in Equation (1) can be written in matrix form as:

$$Y = X\beta + g(t) + \varepsilon \tag{2}$$

Where Y is an $n \times 1$ response vector, X is an $n \times (p+1)$ predictor matrix for the parametric component, t contains the predictors for the nonparametric component. β is a $(p+1) \times 1$ vector of regression parameters. $g(t)$ represents an unknown-shape regression function, and ε is a random error vector normally distributed with mean zero and variance σ^2.

The regression function $g(t_{\mathrm{i}})$ exhibits a pattern that tends to repeat over time, allowing the regression function to be approximated using a Fourier series estimator, which is expressed as follows [22]:

$$g(t_{\mathrm{i}}) = \frac{1}{2}\alpha_0 + \gamma t_{\mathrm{i}} + \sum_{k=1}^{K} \alpha_k \cos kt_{\mathrm{i}}, for\ i = 1, 2, \ldots, n \tag{3}$$

Where $a_0, a_1, a_2, \ldots, a_K$ are the parameters in the nonparametric Fourier series regression model. The term t_i represents the time index for the i-th observation, and cos kt_i is the cosine function at the K-th harmonic frequency, with k denoting the Fourier coefficient.

The parameters in the semiparametric regression function can be estimated using the Ordinary Least Squares (OLS) optimization method. The estimation of the parameter θ in semiparametric regression is obtained by minimizing the squared errors, formulated as follows:

$$\psi(\theta) = \varepsilon^T\varepsilon = (Y^* - \mathrm{C}\theta)^T((Y^* - \mathrm{C}\theta) = Y^{*\mathrm{T}}Y^* - \theta^{\mathrm{T}}\mathrm{C}^{\mathrm{T}}Y^* - Y^{*\mathrm{T}}\mathrm{C}\theta + \theta^{\mathrm{T}}\mathrm{C}^{\mathrm{T}}\mathrm{C}\theta \tag{4}$$

Where is, $Y^* = Y - X\beta$ and based on equation (4), the function for the OLS method within the semiparametric model using the Fourier series estimator produces the following parameter estimate values: $\hat{\theta} = (\mathrm{C}^{\mathrm{T}}\,\mathrm{C})^{-1}\mathrm{C}^{\mathrm{T}}Y^*$, $\hat{g}(t) = D(Y - X\beta)$ and $\hat{Y} = X\hat{\beta} + \hat{g}(t)$

After obtaining the general form of the OLS estimator, the components of the model namely the response vector, the parametric predictor matrix, the regression parameters, the nonparametric function, and the error vector are expressed as follows:

$$Y=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix};\ X=\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p}\\ 1 & x_{21} & x_{22} & \dots & x_{2p}\\ \vdots & \vdots & \dots & \ddots & \vdots\\ 1 & x_{n1} & x_{n2} & \dots & x_{np}\end{bmatrix};\beta=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_p\end{bmatrix};g(t)=\begin{bmatrix} g(t_{11},t_{12},\cdots,t_{1q})\\ g(t_{21},t_{22},\cdots,t_{2q})\\ \vdots\\ g(t_{n1},t_{n2},\cdots,t_{nq})\end{bmatrix};\ \varepsilon=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix} \quad (5)$$

Next, the matrix C is constructed to represent the nonparametric component, which consists of the time function and the Fourier series transformations:

$$\mathrm{C}=\begin{bmatrix} 1 & t_{11} & \cos t_{12} & \cos 2t_{11} & \cdots & \cos\mathrm{K}\, t_{11} & \cdots & t_{1q} & \cos t_{1q} & \cos 2t_{1q} & \dots & \cos\mathrm{K}\, t_{1q}\\ 1 & t_{21} & \cos t_{21} & \cos 2t_{21} & \cdots & \cos\mathrm{K}\, t_{21} & \cdots & t_{2q} & \cos t_{2q} & \cos 2t_{2q} & \cdots & \cos\mathrm{K}\, t_{2q}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & t_{n1} & \cos t_{n1} & \cos 2t_{n1} & \cdots & \cos\mathrm{K}\, t_{n1} & \cdots & t_{nq} & \cos t_{nq} & \cos 2t_{nq} & \cdots & \cos\mathrm{K}\, t_{nq}\end{bmatrix} \quad (6)$$

After constructing the matrix C, the nonparametric parameter vector θ is defined as a combination of the time coefficients and the Fourier coefficients:

$$\theta=[\emptyset \quad \gamma_1 \quad \alpha_{11} \quad \cdots \quad \alpha_{1K} \quad \cdots \quad \gamma_q \quad \alpha_{q1} \quad \cdots \quad \alpha_{qK}]^T \quad (7)$$

$$\emptyset=\frac{1}{2}\sum_{j=1}^{q}\alpha_{0j} \quad (8)$$

Finally, the smoothing matrix D are calculated to obtain the final estimator of the nonparametric function $g(t)$:

$$D=\ \mathrm{C}(\mathrm{C}^{\mathbf{T}}\ \mathrm{C})^{-\mathbf{1}}\mathrm{C}^{\mathbf{T}} \quad (9)$$

To obtain the most accurate estimates, selecting an optimal bandwidth and its corresponding kernel function is crucial. This selection can be made using the Generalized Cross-Validation (GCV) criterion, defined as follows:

$$GCV==\frac{n^{-1}\left\|(I-A(K))y\right\|^2}{\left(n^{-1}trace\big(I-A(K)\big)\right)^2} \quad (10)$$

The coefficient of determination (R^2) is an indicator used to measure the percentage of variation in the response variable (y) that is explained by the predictor variable (x) [23]. The value of the coefficient of determination (R^2) is obtained as follows [24]:

$$R^2=\frac{\sum_{i=1}^{n}(\hat{y}_\mathrm{i}-\bar{y})^2}{\sum_{i=1}^{n}(y_\mathrm{i}-\bar{y})^2} \quad (11)$$

The error rate used to determine the best estimator is evaluated using the following measure:

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{|y_t-\hat{y}_t|}{y_t}\times 100\% \quad (12)$$

Where n is the size of the sample, $\bar{y}$ is mean of the response variable $\hat{y}_t$ is the value predicted by the model for time point t and y is the value observed at time point t.

In the initial stage of the study, a simulation-based analysis was conducted to validate the method and determine the optimal data-splitting proportion before applying the approach to the actual dataset. The simulated data were generated following the structure of a semiparametric regression model consisting of parametric and nonparametric components, and were then evaluated under several data-splitting scenarios, namely 60:40, 70:30, 80:20, and 90:10. Each scenario was assessed using the Mean Absolute Percentage Error (MAPE) on the out-sample data to measure prediction accuracy. Based on the evaluation results, the 90:10 proportion produced the best performance, yielding the lowest out-sample MAPE; therefore, this proportion was adopted as the basis for data splitting in the analysis of the actual dataset.

After determining the optimal data-splitting proportion through the simulation analysis, the same methodological framework was applied to the actual dataset. The stages of the actual data analysis are as follows:

1. Input the actual research data, which include Shortwave Radiation (SWR), Photosynthetically Active Radiation (PAR), and the observation time.
2. Split the data into in-sample and out-sample sets using the optimal proportion obtained from the simulation analysis, namely 90:10.
3. Use the in-sample data to calculate correlations, construct scatterplots, and perform linearity tests. The results are used to proceed with the development of the semiparametric model. If the model assumptions are not satisfied, the data input process is repeated until all assumptions are met.
4. Perform the analysis by determining the optimal Fourier coefficient (K) and selecting the best coefficient using the minimum Generalized Cross-Validation (GCV) criterion.
5. Construct the semiparametric Fourier series model using the optimal Fourier coefficient (K) obtained in Step 4.
6. Evaluate the goodness of fit of the resulting model using the in-sample R^2 value.
7. Validate the model using the out-sample data and assess its performance using the Mean Absolute Percentage Error (MAPE).

8. Generate a plot comparing the forecasting results with the actual in-sample data to assess the model's performance.
9. Compute the forecasting accuracy.
10. Perform forecasting for the number of out-sample observations and generate a plot comparing the forecasting results with the actual out-sample data.

The flow diagram shown in Figure 1 illustrates the process used to analyze and forecast Shortwave Radiation (SWR) based on Photosynthetically Active Radiation (PAR) and time. The process begins with data collection and input, followed by the creation of scatter plots to visualize the relationships between SWR, PAR, and time. A semiparametric regression model using the Fourier series approach is then selected, optimized using the Generalized Cross-Validation (GCV) criterion, and evaluated using the Mean Absolute Percentage Error (MAPE). The best-performing model is subsequently applied for forecasting SWR values, and the results are analyzed to identify the underlying patterns and variations in SWR based on PAR and time (Figure 1).

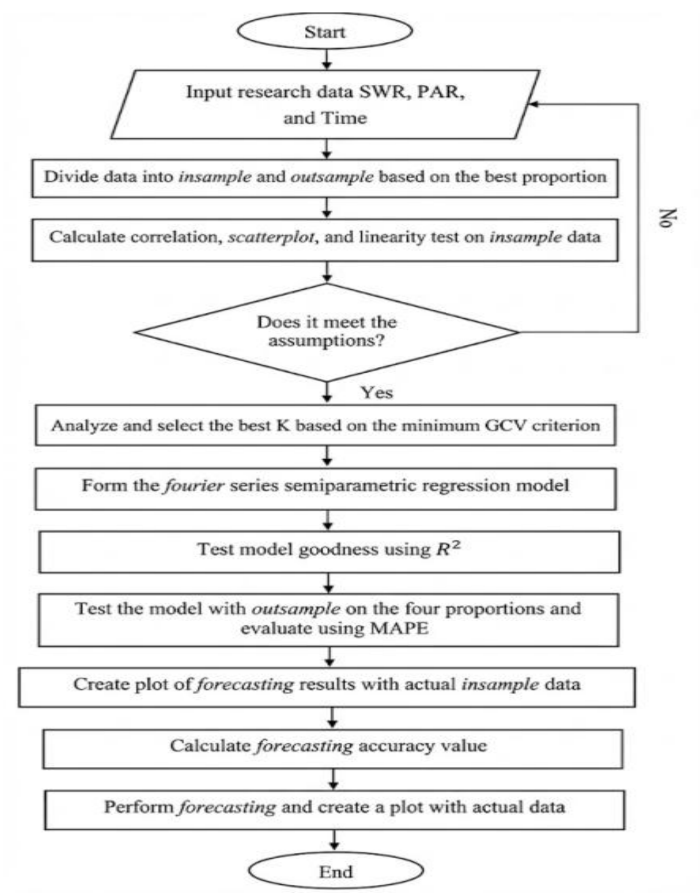

Figure 1. Research Flow Diagram of the Semiparametric Fourier Series Regression Model

3. Main Results

Semiparametric regression modeling using the Fourier series estimator is carried out through several stages until an appropriate model is obtained. In this study, the process began with developing an algorithm and program for simulation purposes. The initial analysis was conducted using simulated data to evaluate and determine the best data-splitting proportion for semiparametric regression modeling. After the optimal proportion was identified through simulation, the next step was to apply the model to the Shortwave Radiation (SWR) data with Photosynthetically Active Radiation (PAR) and Time as predictor variables, based on 182 observations. Before applying the model to the actual data, scatterplots of the relationship between SWR and PAR as well as between SWR and Time were presented. This visualization aims to observe the initial patterns in the data, where the relationship between SWR and PAR provides a preliminary understanding of the characteristics of the dataset.

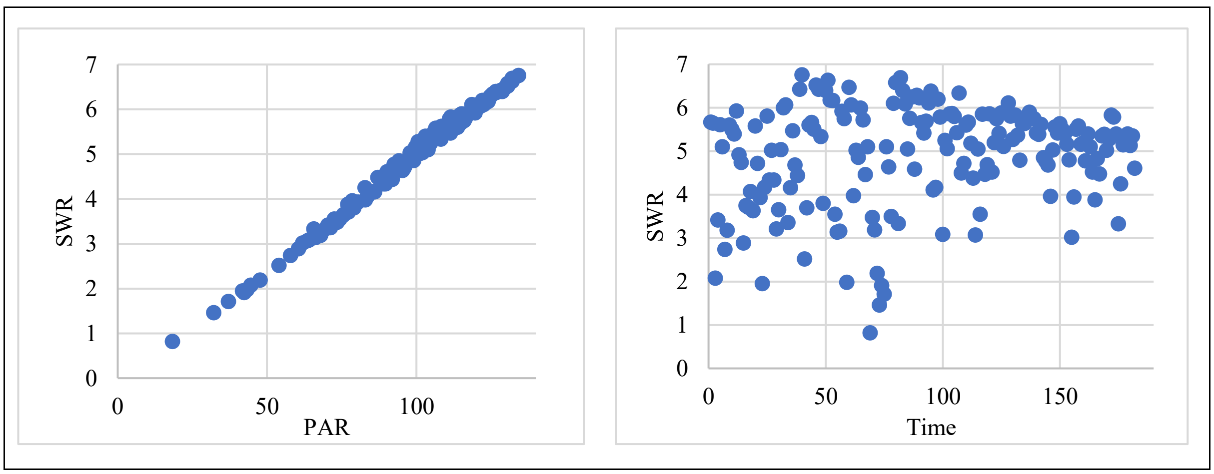

Figure 2. Scatterplots of SWR versus PAR (left) and SWR versus Time (right)

Based on Figure 2, it can be concluded that the PAR variable forms a linear pattern with SWR, while the Time variable does not show a specific pattern and tends to fluctuate, thus requiring a special approach. Therefore, semiparametric regression modeling using the Fourier series estimator was selected because it is capable of handling a combination of linear and nonlinear variables simultaneously.

The following is a summary of the simulation results, namely the MAPE values for various out-sample data proportions, which are presented in Table 1 below:

Table 1. Correlation Matrix

Data Proportion	*Fourier* Coefficient	MAPE Value
60:40	1	2.554%
70:30	1	2,462%
80:20	1	2,506%
90:10	**1**	**1,997%**

All out-sample MAPE values reported in Table 1 are below the 10% threshold, indicating that the model has a very high level of forecasting accuracy. The 90:10 proportion yields the lowest out-sample MAPE value of 1.997%, which means that the model under this proportion is the most accurate for forecasting. These results clearly show that the semiparametric regression model with a Fourier coefficient of 1 is capable of producing excellent forecasts across all data proportions.

Based on these results, the model with the 90:10 proportion was used as a reference for application to the actual data, namely shortwave radiation (SWR), photosynthetically active radiation (PAR), and time. This application aims to evaluate how well the semiparametric Fourier series regression model is able to capture the patterns and variations in the actual data.

Before constructing the semiparametric Fourier series regression model on the original data, the first step that needs to be taken is to examine the correlation between the response variable (y) and the predictor variable (x). This examination aims to ensure that a linear relationship exists between the two variables. The following figure shows the correlation plot between variables y and x:

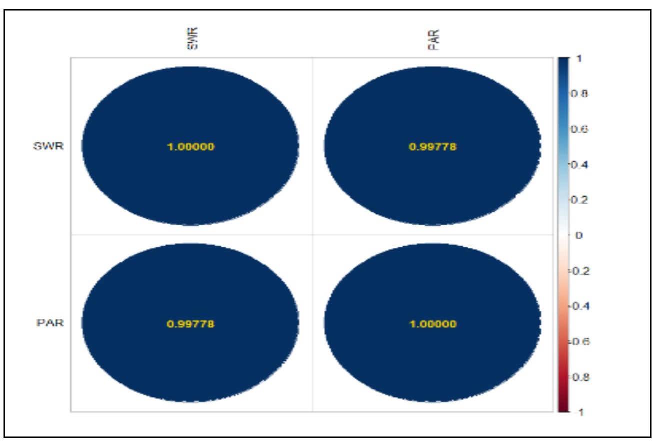

Figure 3. Correlation Plot between SWR and PAR Variables

Figure 3 presents a visualization of the correlation between the SWR and PAR variables, which illustrates the strength and direction of the relationship between the two variables. The size and intensity of the blue color in the circles indicate that the relationship between these variables is very strong and positive. To further support the visualization above, a table containing the Pearson correlation coefficient is included.

The table is provided to complement the visual display with more precise numerical information. To examine the correlation values between the variables in more detail, Table 2 presents the Pearson correlation coefficients for the SWR and PAR variables:

Table 2. Pearson Correlation Coefficient Values between SWR and PAR Variables

	SWR	**PAR**
SWR	1,000	0,997
PAR	0,997	1,000

Table 2 shows that the Pearson correlation coefficient between the SWR and PAR variables is 0.997. This value is very close to 1, indicating a very strong positive linear relationship between the two variables. This means that any increase in photosynthetically active radiation (PAR) is followed by an increase in shortwave radiation (SWR). This very high correlation reinforces the assumption that these variables are highly interrelated.

Since the SWR and PAR variables exhibit a very strong relationship based on the Pearson correlation value, the next step is to visualize this relationship using scatterplot. This visualization aims to provide an illustration of the pattern of the relationship between the variables. To further strengthen the analysis of the relationship between variables, scatterplots of SWR versus PAR and SWR versus Time are presented:

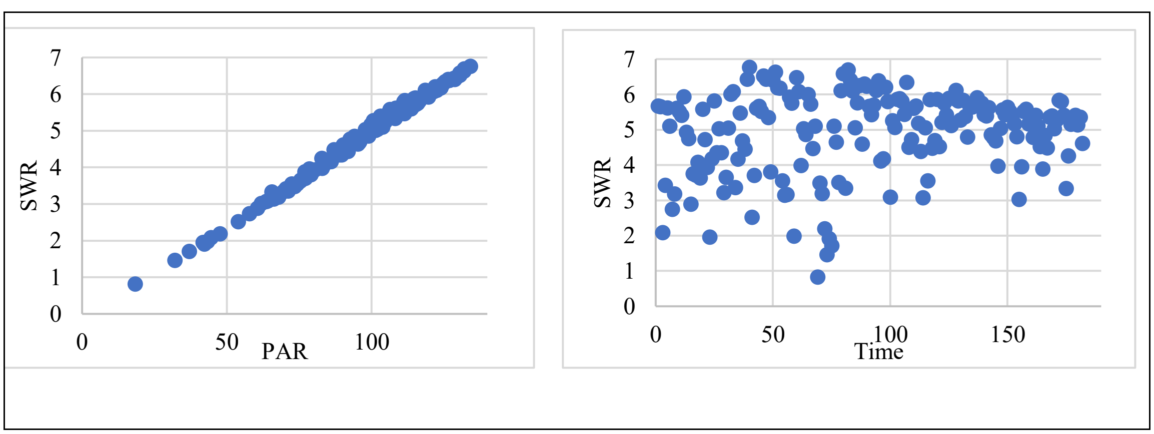

Figure 4. Scatterplot of SWR vs PAR (left) and SWR vs Time (right) for 90:10 Proportion Data.

Figure 4 shows the scatterplot between SWR and PAR (left) and between SWR and Time (right).

On the left, the data points form a pattern that is almost aligned with a straight line, indicating a linear relationship between SWR and PAR. This pattern suggests that increases in PAR are consistently followed by increases in SWR, making the parametric regression approach suitable for modeling the relationship between SWR and PAR.

Meanwhile, on the right side of Figure 4, the SWR values fluctuate over time with a regular up-and-down pattern. This suggests the presence of a seasonal or periodic component in the SWR data, making the semiparametric regression approach with a Fourier series estimator appropriate for capturing such variation in the model. To ensure that the scatterplot assumption above closely resembles a linear pattern, a linear regression approach was used to test the linearity assumption between the variables. The significance test for linearity shows a p-value of 2×10^{-16}. Since the p-value < 0.05, H_0 is rejected, indicating a significant linear relationship between the response variable (y) and the predictor variable (x).

The next step is to determine the optimal Fourier coefficient in the semiparametric regression model, which is carried out by considering the Generalized Cross Validation (GCV) value as the evaluation criterion. The minimum GCV value is used as the basis for determining the optimal Fourier coefficient to be applied in the semiparametric regression model. The following figure presents the relationship between the number of Fourier coefficients and the resulting GCV values.

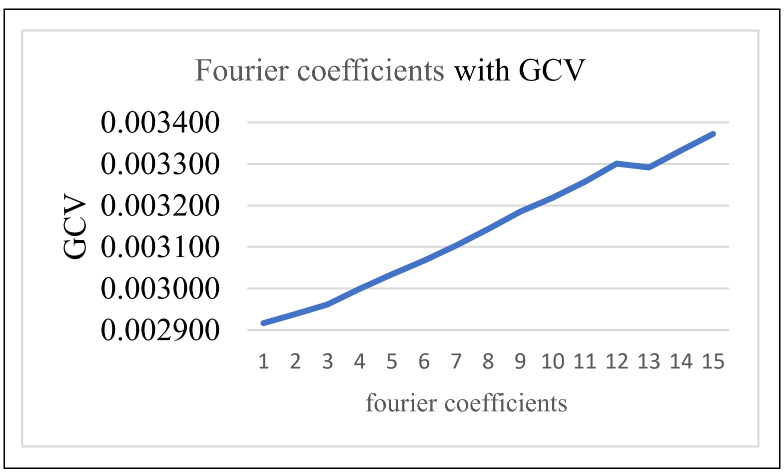

Figure 5. Plot of Fourier coefficients with GCV values

The GCV values corresponding to the plot shown in Figure 5 are presented in detail in Table 3 for the 90:10 data-splitting proportion. This presentation aims to facilitate the identification of the minimum GCV value for the 90:10 proportion, thereby determining the most optimal number of fourier coefficients.

Table 3. GCV Values for Each Fourier Coefficient

Fourier Coefficient	GCV
1	**0,00291**
2	0,00293
3	0,00296
4	0,00300
5	0,00303
6	0,00306
7	0,00310
8	0,00314
9	0,00318
10	0,00321

Table 3 shows that the GCV values increase as the number of Fourier coefficients increases. Based on Table 3, it is evident that a Fourier coefficient of 1 produces the lowest GCV value, and therefore it was selected as the coefficient to construct the best semiparametric regression model. The following is the semiparametric Fourier series regression model using a Fourier coefficient of 1:

$$
\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \frac{1}{2}\hat{\alpha}_0 + \hat{\gamma} t_{1i} + \sum_{k=1}^{K} \hat{\alpha}_k \cos k t_{1i}, i = 1, 2, \ldots, n \\
\hat{y}_i &= -3.381777 \times 10^{-20} + 0.052499 x_{1i} - 0.363654 + 0.001301 t_{1i} - 0.001515 \cos t_{1i} \\
\hat{y}_i &= -0.363654 + 0.052499 x_{1i} + 0.001301 t_{1i} - 0.001515 \cos t_{1i}
\end{aligned}
\tag{13}
$$

The semiparametric Fourier series regression model in Equation (13) was constructed to forecast SWR values based on PAR and time. Based on this model, the intercept value is -0.363654, representing the estimated SWR value when all predictor variables (PAR, time, and the periodic component) are zero. The PAR variable acts as the linear component with the largest influence on SWR, as indicated by its coefficient of 0.052499. This means that every one-unit increase in PAR increases the SWR value by 0.052499 units, assuming other variables remain constant. This aligns with the theory that higher PAR results in higher SWR received.

Meanwhile, time is included in the model as a nonlinear component with a coefficient of 0.001301, indicating that as time increases, there is a tendency for SWR to rise by 0.001301 units per unit of time. This reflects a long-term increasing trend in SWR. The periodic component in the form of cos(t) has a coefficient of -0.001515, meaning that when cos(t) is positive, its effect tends to decrease SWR, and when cos(t) is negative, its effect tends to increase SWR. This component captures seasonal variations that cannot be explained by the linear trend alone. This reinforces that the semiparametric Fourier series model is capable of capturing fluctuations in SWR over time.

Overall, the model demonstrates that SWR values are strongly influenced by PAR and seasonally affected by time, making it suitable for representing solar and photosynthetically active radiation conditions. The next step is to evaluate the performance of the constructed model by calculating the in-sample R^2 and out-sample MAPE for the 90:10 proportion. This evaluation aims to measure how well the model represents the influence among variables and the accuracy of the resulting forecasts. Manual calculations are carried out to provide a more precise assessment of the model's performance in explaining the data and generating predictions.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 0.998$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right)x100\% = 0.9003\%$$

The manual calculations conducted show the in-sample R^2 and out-sample MAPE as model evaluation metrics. To make the results easier to understand, the in-sample R^2 and out-sample MAPE values can be seen in the table below:

Table 4. In-Sample R^2 and Out-Sample MAPE Values

Kriteria	Nilai
R^2	99.8%
MAPE	0.9003%

Based on Table 4, the in-sample R^2 value is 0.998 (99.8%), indicating that the model has an excellent ability to explain the relationship between SWR, PAR, and time. This value is close to 100%, meaning that almost all the variation in the SWR data is explained by the model. This indicates that the combination of linear and periodic variables effectively captures the characteristics of the data, resulting in very small forecasting errors.

Furthermore, the out-sample MAPE value of 0.9003% shows that the average forecasting error of the model compared to the actual data is very low. With an out-sample MAPE below 10%, the model is considered to have a very high level of accuracy. Therefore, the constructed semiparametric regression model can represent the SWR data pattern well and is effective for both analysis and forecasting purposes.

As a visual representation of the model's performance, a plot of the in-sample data comparing actual y values and forecasted ŷ values is presented. The following figure shows the comparison between y and ŷ in the in-sample data:

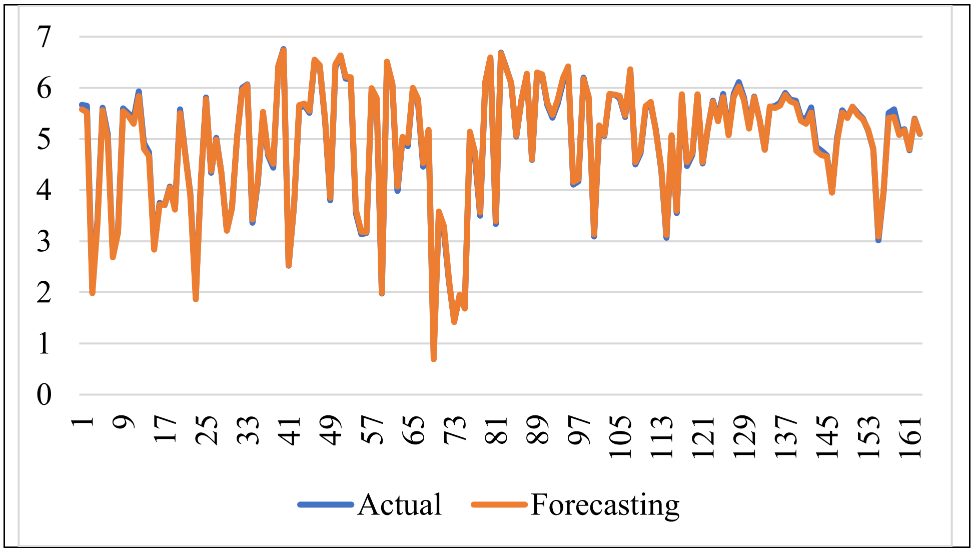

Figure 6. Plot of Actual and In-Sample Forecasted Data for the Original Data

Figure 6 presents a plot of the actual data and the in-sample forecasted results. The plot shows that the forecasted data pattern closely follows the actual data, as seen from the red line almost always tracking the blue line, despite slight differences in some time periods. From the resulting plot, it can be concluded that the obtained model is suitable for forecasting Shortwave Radiation (SWR) data. The model's ability to match the pattern of the actual data indicates a high potential for accurate forecasting. To assess the quality of the model's forecasting results, the next step is to calculate the forecasting accuracy based on the previously obtained out-sample MAPE value.

Evaluating the forecasting accuracy is an important step in assessing the quality of the constructed regression model. This accuracy can be measured using the out-sample Mean Absolute Percentage Error (MAPE), which indicates the average percentage error between the actual values and the forecasted values. The out-sample MAPE obtained from the semiparametric Fourier series regression model is 0.9003%, indicating a very low forecasting error.

The low out-sample MAPE indicates that the model's performance is very good, as it is below the 10% threshold. Based on the out-sample MAPE value, the forecasting accuracy of the model is 99.0997%, meaning that the forecasted results are very close to the actual values. These results reinforce that the semiparametric approach with linear and seasonal components based on the Fourier series is an effective method for accurately representing environmental data dynamics.

Model accuracy evaluation is not only conducted through the numerical indicator of out-sample MAPE but is also supported by a comparison table between actual data and forecasted results. This comparison serves to evaluate the extent of differences between actual and forecasted values. Through the table of differences between actual and forecasted values, these discrepancies can be more clearly identified. Table 5 presents a comparison between the actual data and forecasts from the semiparametric regression model for time points 164 to 182.

Table 5. Actual and Forecasts from the Semiparametric Regression Model

Time	Actual	*Forecasting*
164	4.520	4.579
165	3.880	3.925
166	4.830	4.861
167	4.480	4.468
168	5.350	5.325
169	5.390	5.427
170	5.020	5.047
171	5.280	5.181
172	5.830	5.771
173	5.790	5.751
174	5.400	5.311
175	3.330	3.288
176	4.250	4.209
177	5.150	5.182
178	5.320	5.365
179	5.400	5.462
180	5.130	5.186
181	5.350	5.378
182	4.610	4.593

The data presentation is not only in tabular form but is also visualized in a plot in Figure 7. This visualization aims to provide an illustration of the alignment between actual values and forecasts, making it easier to observe the movement pattern between actual and forecasted values. The following figure shows a plot comparing actual and forecasted values for time points 164 to 182:

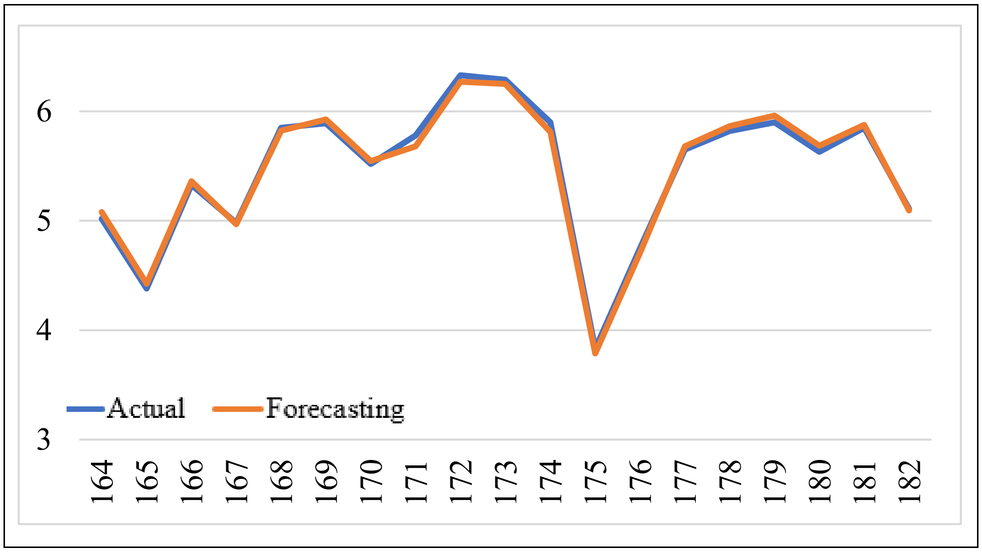

Figure 7. Plot of Actual and Out-Sample Forecasted Data

Figure 7 shows that the two lines in the plot are very close and follow almost the same movement pattern, both in upward and downward trends. This indicates that the semiparametric regression model used has excellent forecasting capability, as it can accurately capture fluctuations in the actual data. Thus, this visualization provides evidence that the model is able to represent data value fluctuations well and demonstrates consistent forecasting performance.

4. CONCLUSIONS

Based on the analysis using a semiparametric regression model with a first-order Fourier series estimator and a 90:10 data split, the optimal model obtained is $\hat{y}_i = -0{,}363654 + 0{,}052499x_{1i} + 0{,}001301t_{1i} - 0{,}001515\cos t_{1i}$. The model indicates that PAR has a positive influence on SWR values, while time contributes a seasonal pattern captured through the cosine function. The model evaluation shows an in-sample coefficient of determination (R^2) value of 0.998 (99.8%) and an out-sample MAPE of 0.9003%, demonstrating that the model has an excellent ability to explain nearly all variations in the data, with a prediction accuracy of 99.0997%. These results highlight the strong performance of the semiparametric regression with Fourier series estimator in forecasting SWR values. Furthermore, the comparison between actual and forecasted data reveals that the model successfully follows the fluctuation patterns consistently, confirming its reliability as a forecasting approach for SWR.

ACKNOWLEDGMENT

The author extends gratitude to the Institute for Research and Community Service (LPPM) of Universitas Muhammadiyah Semarang—supported by the PDP Internal Research Grant (Decree No. 034/UNIMUS.L/PG/PDP/PJ.INT/2025)—for providing the necessary support and facilities to conduct this research. Appreciation is also expressed to all parties who provided technical and non-technical support, enabling the successful execution of this study; it is hoped that the findings will prove beneficial to other researchers.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] S.K. Jha, A.C. Patra, G.P. Verma, I.S. Iyer, D.K. Aswal, Natural Radiation and Environment, in: Handbook on Radiation Environment, Volume 1, Springer, 2024: pp. 27-72. https://doi.org/10.1007/978-981-97-2795-7_2.

[2] H. Rusmaryadi, Sukarmansyah, T.P.O. Sianipar, H. Setiadi, Pengaruh Cermin Reflektor terhadap Daya dan Kenaikkan Temperatur Sel Surya, Turbulen: J. Tek. Mesin 1 (2018), 85-94.

[3] T.A. Howell, D.W. Meek, J.L. Hatfield, Relationship of Photosynthetically Active Radiation to Shortwave Radiation in the San Joaquin Valley, Agric. Meteorol. 28 (1983), 157-175. https://doi.org/10.1016/0002-1571(83)90005-5.

[4] O. Asime, C.D. Lawson, The Effects of Rainfall and Relative Humidity on Solar Radiation Intensity during Dry Season in Port Harcourt Metropolis—Quantitative Approach, Res. Dev. Sci. Technol. 6 (2022), 98-111. https://doi.org/10.9734/bpi/rdst/v6/15307D.

[5] Y. Sianturi, Pengukuran dan Analisa Data Radiasi Matahari di Stasiun Klimatologi Muaro Jambi, Megasains 12 (2021), 40-47. https://doi.org/10.46824/megasains.v12i1.45.

[6] N. Nurdiani, N. Herrhyantom, D. Dasari, Regresi Nonparametrik Birespon Spline, J. Eurekamatika 5 (2017), 106-121.

[7] A. Suciani, D. Ruhiat, S.D. Rahayu, Komparasi Hasil Analisis Beda Rata-rata Menggunakan Metode Statistik Parametrik dan Nonparametrik, J. Ris. Mat. Sains Teknol. 2 (2022), 92-103.

[8] T.W. Utami, I.M. Nur, E. Suryaningsih, Pemodelan Produksi Padi Menggunakan Regresi Semiparametrik Kernel, J. Stat. 6 (2018), 160-165.

[9] D.A. Nanda, Suparti, A. Hoyyi, Analisis Pengaruh Jumlah Uang Beredar dan Nilai Tukar Rupiah terhadap Indeks Harga Saham Gabungan Menggunakan Pemodelan Regresi Semiparametrik Kernel, J. Gaussian 5 (2016), 373-382.

[10] N. Chamidah, B. Zaman, L. Muniroh, B. Lestari, Designing Local Standard Growth Charts of Children in East Java Province Using a Local Linear Estimator, Int. J. Innov. Creat. Change 13 (2020), 45-67.

[11] T.W. Utami, N. Chamidah, T. Saifudin, Hematocrit Modeling in DHF Patients Based on Longitudinal Data Using Local Linear Estimator in Semiparametric Regression, AIP Conf. Proc. 3302 (2025), 020006. https://doi.org/10.1063/5.0262183.

[12] V. Fibriyani, N. Chamidah, T. Saifudin, Estimating Time Series Semiparametric Regression Model Using Local Polynomial Estimator for Predicting Inflation Rate in Indonesia, J. King Saud Univ. Sci. 36 (2024), 103549. https://doi.org/10.1016/j.jksus.2024.103549.

[13] A. Prahutama, Suparti, Pemodelan Harga Cabai di Kota Semarang terhadap Harga Inflasi Menggunakan Regresi Semiparametrik Polinomial Lokal, J. Stat. 5 (2017), 1-7.

[14] T.W. Utami, A. Prahutama, Regresi Semiparametrik Spline Truncated dengan Software R, in: Seminar Nasional Pendidikan, Sains dan Teknologi, Universitas Muhammadiyah Semarang, pp. 17-23, 2017.

[15] L.R. Khairunnisa, A. Prahutama, R. Santoso, Pemodelan Regresi Semiparametrik dengan Pendekatan Deret Fourier (Studi Kasus: Pengaruh Indeks Dow Jones dan BI Rate terhadap Indeks Harga Saham Gabungan), J. Gaussian 9 (2020), 50-63.

[16] L.J. Asrini, I.N. Budiantara, Fourier Series Semiparametric Regression Models (Case Study: The Production of Lowland Rice Irrigation in Central Java), ARPN J. Eng. Appl. Sci. 9 (2014), 1501-1506.

[17] N.Y. Adrianingsih, A.T.R. Dani, A. Ainurrochmah, Pemodelan dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Nusa Tenggara Timur, in: Prosiding Seminar Edusainstech, FMIPA UNIMUS, pp. 400-407, 2020.

[18] L.P. Inasari, M. Irwan, W. Abidin, W. Alwi, Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Kemiskinan di Sulawesi Selatan, J. Mat. Stat. Apl. 12 (2024), 79-87. https://doi.org/10.24252/msa.v12i2.51107.

[19] T.W. Utami, I.M. Nur, Aplikasi Regresi Nonparametrik Deret Fourier pada Data High Water Level (HWL) Kota Semarang, J. Stat. 5 (2017), 57-61.

[20] S. Sahidah, K. Kuzairi, M.F.F. Mardianto, Estimator Deret Fourier Dalam Regresi Nonparametrik dengan Penalti untuk Perencanaan Penjualan Produk Musiman, Zeta Math J. 7 (2022), 69-78. https://doi.org/10.31102/zeta.2022.7.2.69-78.

[21] I.F. Amri, N. Chamidah, T. Saifudin, B. Lestari, Determining Relative Humidity in Surakarta City Using Semiparametric Regression Based on Fourier Series Penalized Least Square, Commun. Math. Biol. Neurosci. 2025 (2025), 61. https://doi.org/10.28919/cmbn/9184.

[22] A.T.R. Dani, A.F. Dewi, L. Ni'matuzzahroh, Studi Simulasi dan Aplikasi: Estimator Deret Fourier pada Pemodelan Regresi Nonparametrik, in: Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya, pp. 279-288, 2022.

[23] A. Latief, D. Rosalina, D. Apiska, Analisis Hubungan Antar Manusia terhadap Kinerja Karyawan, J. Educ. Humanit. Soc. Sci. 1 (2019), 127-131. https://doi.org/10.34007/jehss.v1i3.34.

[24] M.A.D. Octavanny, Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Papua Dengan Regresi Semiparametrik Truncated Spline, in: Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya, pp. 29-41, 2023.