Data and Metadata. 2025; 4:890
doi: 10.56294/dm2025890

ORIGINAL

Forecasting Temperature of Earth Surface in Sragen Regency Using Semiparametric Regression Based on Penalized Fourier Series Estimator

Pronóstico de la temperatura de la superficie terrestre en la regencia de Sragen mediante regresión semiparamétrica basada en el estimador de series de Fourier penalizadas

Ihsan Fathoni Amri[1,2], Nur Chamidah[3,4], Toha Saifudin[3,4], Budi Lestari[5], Dursun Aydin[6,7]

[1]Airlangga University, Doctoral Study Program of Mathematics and Natural Sciences, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[2]Universitas Muhammadiyah Semarang, Department of Data Science, Faculty of Science and Agriculture Technology. Semarang 50273, Indonesia.
[3]Airlangga University, Department of Mathematics, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[4]Airlangga University, Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[5]University of Jember, Department of Mathematics, Faculty of Mathematics and Natural Sciences. Jember 68121, Indonesia.
[6]Muğla Sıtkı Koçman University, Department of Statistics, Faculty of Science. Muğla 48000, Turkey.
[7]University of Wisconsin, Research Scholar at Department of Mathematics. Oshkosh Algoma Blvd, Oshkosh, WI 54901, USA.

Cite as: Amri IF, Chamidah N, Saifudin T, Lestari B, Aydin D. Forecasting Temperature of Earth Surface in Sragen Regency Using Semiparametric Regression Based on Penalized Fourier Series Estimator. Data and Metadata. 2025; 4:890. https://doi.org/10.56294/dm2025890

Submitted: 21-07-2024 **Revised:** 21-11-2024 **Accepted:** 30-04-2025 **Published:** 01-05-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Nur Chamidah

ABSTRACT

Sragen regency that is located in Central Java Province of Indonesia, is one of the areas that feels the direct impact of the high earth surface temperature. The various sectors in Sragen regency, including agriculture, health, and the environment are affected by the high temperature of the earth's surface. The Sragen regency is geographically dominated by agricultural areas, which are very vulnerable to extreme earth surface temperatures. This has a direct effect on agricultural productivity and the availability of water for irrigation. This study examines the use of a semiparametric regression model with a Penalized Least Squares (PLS)-based Fourier Series estimator to analyze the relationship between earth surface temperature and relative humidity in Sragen regency. The combining parametric and nonparametric components, the model effectively addresses complex climate data patterns. A dataset of 100 observations was analyzed under three training data scenarios N = 70, N = 80, and N = 90, yielding optimal Fourier coefficients of 1, 1, 1 and lambda values of 0,035, 0,028, and 0,02. The resulting minimum Generalized Cross Validation (GCV) values of 0,3534871, 0,3711413, and 0,3918924. This model successfully made good predictions for testing data sizes of 30, 20, and 10, with MAPE values of 1,606545, 1,518221, and 1,018482. These results underscore the model's ability to capture the inverse relationship between earth surface temperature and relative humidity. The study highlights the Fourier-based semiparametric approach's effectiveness in dynamic scenarios and recommends applying it to other climate variables or regions to further evaluate its adaptability and robustness.

Keywords: Semiparametric Regression; Penalized Fourier Series Estimator; Earth Surface Temperature; Relative Humidity; Climate Change.

RESUMEN

La regencia de Sragen, ubicada en la provincia de Java Central, Indonesia, es una de las zonas que sufre el impacto directo de las altas temperaturas superficiales. Diversos sectores de la regencia, como la agricultura, la salud y el medio ambiente, se ven afectados por las altas temperaturas. La regencia de

Sragen está geográficamente dominada por zonas agrícolas, muy vulnerables a las temperaturas extremas de la superficie terrestre. Esto tiene un efecto directo en la productividad agrícola y la disponibilidad de agua para riego. Este estudio examina el uso de un modelo de regresión semiparamétrica con un estimador de series de Fourier basado en Mínimos Cuadrados Penalizados (MCP) para analizar la relación entre la temperatura de la superficie de la Tierra y la humedad relativa en la regencia de Sragen, Indonesia. Al combinar componentes paramétricos y no paramétricos, el modelo aborda de manera efectiva patrones complejos de datos climáticos. Se analizó un conjunto de datos de 100 observaciones en tres escenarios de datos de entrenamiento N = 70, N = 80 y N = 90, lo que arrojó coeficientes de Fourier óptimos de 1, 1, 1 y valores lambda de 0,035, 0,028 y 0,02. Los valores mínimos resultantes de Validación Cruzada Generalizada (VCG) de 0,3534871, 0,3711413 y 0,3918924. Este modelo realizó predicciones satisfactorias para tamaños de datos de prueba de 30, 20 y 10, con valores MAPE de 1,606545, 1,518221 y 1,018482. Estos resultados subrayan la capacidad del modelo para capturar la relación inversa entre la temperatura de la superficie terrestre y la humedad relativa. El estudio destaca la eficacia del enfoque semiparamétrico basado en Fourier en escenarios dinámicos y recomienda aplicarlo a otras variables climáticas o regiones para evaluar más a fondo su adaptabilidad y robustez.

Palabras clave: Regresión Semiparamétrica; Estimador de Series de Fourier Penalizado; Temperatura de la Superficie de la Tierra; Humedad Relativa; Cambio Climático.

INTRODUCTION

Drastic climate change is a phenomenon that we are currently experiencing in various countries in the world, not apart from Indonesia. Climate change in Indonesia has a huge impact because it causes several adverse natural phenomena such as natural phenomena such as floods, landslides, fires, earthquakes, volcanoes erupting and many others.[1] Basically, climate change is caused by the greenhouse effect by an increase in the concentration of gases in the air, one of which is carbon dioxide (CO2).[2] An increase in the concentration of CO_2 gas causes the earth surface temperature to increase drastically.[3] This is because the concentration level of CO_2 gas contained in the air plays a role in the percentage level of air humidity, if the percentage level of humidity in the air is smaller, the temperature of the earth's surface will rise higher.[4]

Facing the challenge of climate change requires a comprehensive and integrated action from various sectors in society. A comprehensive and integrated action in addressing the impacts of climate change depends not only on how to reduce the level of gas emissions due to the greenhouse effect, but also on adaptation to the changes that occur.[5] Adaptations that can be made, such as increasing the resilience of agricultural systems by using plant varieties resistant to high earth surface temperatures, efficient and effective water management so that droughts do not occur, and infrastructure improvements to prepare for natural disaster risk management that may occurs.[6] In addition, sustainable environmental management, such as reforestation and land management in the right way, is highly recommended as a countermeasure, because it can reduce its negative impacts.[7] As we know that tackling before an event occurs is more effective than tackling after an event occurs.

The high temperature of the earth's surface is now a very troubling complaint for the community, because it has an impact on many sectors in Indonesia. Due to the high temperature of the earth's surface, several existing sectors such as the agriculture, health, environment, and so on sectors are affected.[8,9] The high temperature of the earth's surface in the agricultural sector causes irregular rainfall patterns so that the level of rainfall is very small. This has an impact on the agricultural sector by disrupting the growing season and potentially reducing crop yields.[10] As a result, local food security has a significant decrease. Furthermore, the rise in the earth surface temperature also affects the health sector because extreme heat causes skin problems. High earth surface temperatures can also cause dehydration and other diseases due to high temperature.[11] In addition, climate change can affect the spread of diseases such as dengue fever. In terms of the environmental sector, the high temperature of the earth's surface causes prolonged drought which has an impact on fires in several Indonesian districts.[12] The impact of these fires usually occurs in places such as agriculture, plantations, settlements, and forestry. The high temperature of the earth's surface is one of the obstacles that must be handled appropriately, and one of the factors that affect the temperature of the earth's surface is the percentage of air humidity level. The relationship between the earth surface temperature and the percentage of air humidity level is an inverse relationship because the lower the percentage of air humidity level, the higher the earth surface temperature, and vice versa.[13] On the other hand, the earth surface temperature data obtained is in the form of time sequence data so that the time variable is also chosen as one of the variables in this study.

Drought and flooding in Indonesia are natural disasters that often occur due to climate change. This disaster can cause various other problems to arise. Drought is one of them, it can cause several issues such as existing

plants are not as productive as usual or even wilt or die, and the land can no longer be planted due to the extreme drought. This often causes huge economic losses due to crop failure or crop failure due to lack of water resources. The El Nino La Nina phenomenon that often occurs in Indonesia is one of the causes of changes in rainfall in Sragen Regency which has an impact on the occurrence of drought disasters. Indonesia's National Disaster Management Agency said that 7 sub-districts in Sragen Regency experienced drought while other areas experienced floods and landslides.[14] The Drought Disaster Risk Index of Sragen Regency in 2020 was also high with a score of 24,0. Sragen Regency, which is known as a rice producing regency and also a rice supplier for Central Java, is capable of producing up to ±440 000 tons of rice. Rice farming of course requires large amounts of water consumption beyond the daily water needs of the population. It is feared that this large need for water will not be met when the dry season arrives. Sragen Regency is an area with an area of 941,55 km2, consisting of 20 districts, 12 sub-districts, and 196 villages. The area of Sragen Regency is physiologically divided into 42,62 % rice fields (40 129 Ha) and 57,38 % dry land (54 026 Ha). The average height of the plains in Sragen district is 109 m above sea level with a population of 976 951 people in 2020.[15]

Sragen regency, located in Central Java Province, Indonesia, is one of the areas that feels the direct impact of the high earth surface temperature.[16] Various sectors in Sragen regency, including agriculture, health, and the environment are affected by the high temperature of the earth's surface. Sragen is geographically dominated by agricultural areas, which are very vulnerable to extreme earth surface temperatures. This has a direct effect on agricultural productivity and the availability of water for irrigation.[17] Not only that, the people of Sragen face an increased risk of increasing the incidence of weather-related diseases, due to the high temperature of the earth's surface. One of them is dengue fever, which tends to increase during and after the rainy period due to the breeding of the Aedes aegypti mosquito.[18] The impact of high earth surface temperatures is also very visible in the environmental sector in Sragen. One of the impacts on the environmental sector is the occurrence of fires in the Sragen area due to the high temperature of the earth's surface, so that many plants die and facilitate the occurrence of fires.[19]

The above description is the main reference in taking the case to be researched, because it really needs to be researched immediately so that the problems that occur can get the right solution. Based on this urgency, an in-depth analysis is needed to model the functional relationship between the earth surface temperature (as the Y response variable) and the percentage level of air humidity (as the X predictor variable) using statistical modeling techniques. There are three main types of regression models, namely parametric regression model,[20] nonparametric regression model,[21,22,23,24,25,26,27,28,29,30] and semiparametric regression model[31,32,33,34,35,36,37] that can be used to determine the functional relationship between response variables and predictor variables. The parametric regression model assumes that the functional relationship between the response variable and the predictor follows a certain pattern.[20] In contrast, nonparametric regression models do not assume a specific form of curve pattern on the functional relationship between response variables and predictor variables.[21,22,23,24,25] In the nonparametric regression model, it has high flexibility because the functional relationship of the regression curve is considered smooth and can be estimated using certain smoothing methods based on data patterns.[26,27,28,29,30] Furthermore, there is also a semiparametric regression model which is a combination of parametric and nonparametric models, which has high flexibility in analyzing the functional relationship between response and predictor variables that is presented by a regression curve.[31,32,33,34,35,36,37] In the semiparametric regression model, there are parametric and nonparametric components where some of the smoothing techniques are used to estimate the nonparametric component, i.e., kernel,[38,39] spline,[39,40] local linear,[41] local polynomial,[42] and Fourier series.[43,44,45,46,47,48]

One of the nonparametric regression approaches used to overcome fluctuating data patterns is to rise high or fall far in the range of values such as the shape of the sine and cosine curve patterns, namely the Fourier series estimator.[47] In temperature data where there is a fluctuation around a high value and at a certain time then decreases to a much lower value, it is very suitable for use in the Fourier series estimator. The repeating data pattern is very consistent with the Fourier series estimator, in the form of repetition of different independent or dependent variables.[48] In the optimization of the Fourier series estimator, there are several optimization methods that can be used, namely Least Squares (LS),[49] Weighted Least Squares (WLS),[49] and Penalized Least Squares (PLS).[50] The PLS is an optimization method that provides smoothing components to the LS method with optimization criteria that combine data matching with the smoothness of the curve. The PLS estimation is carried out to balance data adjustments and avoid excessive roughness.[50] Basically, the PLS is very good to use because in general, Generalized Cross Validation (GCV) cannot select really good parameters because the overfitting effect is ignored in the resulting model.[51] Based on this, penalized in an estimation method is an optimization that can be used to overcome overfitting.[51]

The application of Fourier series estimators on semiparametric regression using the PLS optimization has been applied in various cases and shows significant potential in overcoming the challenges faced by data that have fluctuating patterns.[43,44,47] Previous research on estimating semiparametric regression model using fourier series based on PLS has been carried out by Fan et al.[50] for estimating weakly dependent data,

namely a meteorological data set. The PLS optimization is very effective in balancing the need to follow data accuracy by avoiding overfitting.[50] The Fourier series estimators, with their ability to model data fluctuations as trigonometric functions, are well suited for data that show repetitive patterns such as the earth surface temperature. This capability is particularly relevant for the study of climate change, where parameters such as temperature and humidity have clear seasonal patterns.[52]

Further research that integrates the Fourier series approach to semiparametric regression using the PLS optimization is expected to provide a more flexible method for understanding and forecasting changes in the earth surface temperature. This is not only important for academic purposes but also for policymaking, where data-driven decisions and accurate forecasting are urgently needed to plan and implement adaptation and mitigation strategies.

METHOD

Fourier series are trigonometric polynomials that have flexibility, so they can adapt effectively to the local nature of the data. Fourier series are good for describing curves that show sine and cosine. Semiparametric regression model estimation using Fourier series is one of the semiparametric regression model estimations that has good statistical interpretation and visual interpretation among other semiparametric regression model estimators. The advantage of model estimation using Fourier series is that it is able to handle periodic data characters, namely following a repeating pattern at a certain interval trend such as cyclical climate data patterns, and has good statistical and visual interpretation.[53]

Next, suppose that we have a paired data ($y,t,u_1,u_2,\ldots,u_p$) follows the following regression model:[50,51]

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

Based on Equation 1 i with observation then variable as ($y_i,t_i,u_{1i},u_{2i},\ldots,u_{pi}$) and the regression model follow:

$$y_i = \beta_0 + \beta_1 u_{1i} + \beta_2 u_{2i} + \ldots + \beta_p u_{pi} + g(t_i) + \varepsilon_i, i = 1,2,\ldots,n. \qquad (2)$$

Equation 2 can be written to matrix:

$$\boldsymbol{y} = \mathbf{U}'\boldsymbol{\beta} + \boldsymbol{g}(\boldsymbol{t}) + \boldsymbol{\varepsilon} \qquad (3)$$

Where y is vector respon variable with i observation, U is matric predictor variable for parametric component, ß is vector parameter for parametric component, g(t) is vector nonparametric regression function and $\varepsilon \sim N_n(0,\sigma^2 I)$.

Parametric component g(t) can be approach by fourier series estimator. The fourier series estimator have high fleksibelity, then really good to use in volatile data. The function for Fourier series can be written as follows:[49,51]

$$\hat{g}(t_i) = \hat{\alpha}_0 + \sum_{j=1}^{J}\left[c_j(cos(2\pi j t_i)) + d_j(sin(2\pi j t_i))\right]. \qquad (4)$$

The PLS is a good optimization method to avoid overfitting effect and the PLS has been used to solved estimation semiparametric model using Fourier series estimator. The function for PLS method on semiparametric using Fourier series estimator can be written as follows:[50]

$$\min_{\beta \epsilon R^{r+1}, g \epsilon C(0,1)} \left[n^{-1}(\boldsymbol{y}^* - \boldsymbol{g}(\boldsymbol{t}))'(\boldsymbol{y}^* - \boldsymbol{g}(\boldsymbol{t})) + \lambda \int_0^1 \left(\boldsymbol{g}^{(2)}(\boldsymbol{t})\right)^2 dt \right] \qquad (5)$$

Based on equation 5, the function for the PLS method in a semiparametric model using the Fourier series estimator yields the following estimates:

$$\hat{\boldsymbol{y}}(\boldsymbol{u},\boldsymbol{t}) = \boldsymbol{U}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{g}}(\boldsymbol{t}); \hat{\boldsymbol{\beta}} = [\boldsymbol{U}'V\boldsymbol{U}]^{-1}\boldsymbol{U}'V\boldsymbol{y} \text{ ; and } \hat{\boldsymbol{g}}(\boldsymbol{t}) = \boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H}\boldsymbol{U}[\boldsymbol{U}'V\boldsymbol{U}]^{-1}\boldsymbol{U}'\boldsymbol{V})\boldsymbol{y}$$

Where:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \mathbf{U}' = \begin{bmatrix} 1 & u_{11} & u_{21} & \cdots & u_{p1} \\ 1 & u_{12} & u_{22} & \cdots & u_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{1n} & u_{2n} & \cdots & u_{pn} \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}; \quad \boldsymbol{H} = \boldsymbol{F}(n^{-1}\boldsymbol{F}'\boldsymbol{F} + \lambda^*\boldsymbol{D})^{-1}n^{-1}\boldsymbol{F}' \quad ; \quad \text{and} \quad \boldsymbol{V} = (\boldsymbol{I} - \boldsymbol{H})'(\boldsymbol{I} - \boldsymbol{H}).$$

To get the best estimation, one of the most important things is to choose an optimal bandwidth with associated Kernel function. This can be performed using Generalized Cross-Validation criterion with the following formula:

$$GCV = \frac{n^{-1}\|(\boldsymbol{I}-\boldsymbol{UK}-\boldsymbol{L})\boldsymbol{y}\|^2}{\left(n^{-1}trace(\boldsymbol{I}-\boldsymbol{UK}-\boldsymbol{L})\right)^2} \qquad (6)$$

Where K=[U' VU]$^{-1}$ U' V, L=H(I-HU[U' VU]$^{-1}$ U'V), and ||.|| is norm of a vector.

The error rate measurement to compare the best estimator is based on MAPE value proposed by Moreno et al.[52] as follows:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t} \times 100\ \% \qquad (7)$$

Where T is the size of the sample, y^t is the value predicted by the model for time point t and y is the value observed at time point t. The MAPE values criteria are shown in table 1.[52]

Table 1. MAPE Value Criteria

MAPE	Definition
< 10	Highly Accurate
10 - 20	Accurate
20 - 50	Reasonable
>50	Inaccurate

Source: Moreno et al.[52]

The process of conducting statistical analysis with a semiparametric method utilizing a The penalized Fourier series can be outlined as follows:

1. Gather data on earth surface temperature and relative humidity as the primary variables for the analysis.
2. Enter earth surface temperature data at time t and relative humidity data at time t into the analysis framework for processing.
3. Develop scatter plots to visualize the relationship between earth surface temperature at time t and relative humidity at time t, as well as between earth surface temperature at time t and time, to observe the interaction patterns among these variables.
4. Determine an appropriate semiparametric regression model to describe the relationships between earth surface temperature, relative humidity, and time.
5. Implement the selected semiparametric regression model to study the associations between earth surface temperature at time t, relative humidity at time t, and time.
6. Calibrate the model by employing the Generalized Cross-Validation (GCV) method to identify optimal parameter values.
7. Evaluate the model's predictive accuracy using the Mean Absolute Percentage Error (MAPE) metric.
8. Select the model with the highest accuracy based on the MAPE evaluation results.
9. Leverage the chosen model to forecast earth surface temperature and verify the predictions using datasets.
10. Interpret the prediction outcomes and document the patterns observed in the relationship between earth surface temperature and relative humidity as identified by the model.

The process for analyzing and predicting earth surface temperature and relative humidity data is summarized in the flow diagram (figure 1). The process begins with data collection and entry, followed by scatter plot

generation to examine relationships among earth surface temperature, relative humidity, and time. Next, a suitable semiparametric regression model is selected, optimized using GCV, and assessed with MAPE. The model with the best performance is used for predictions, and the results are analyzed to uncover patterns in earth surface temperature and relative humidity data (figure 1).

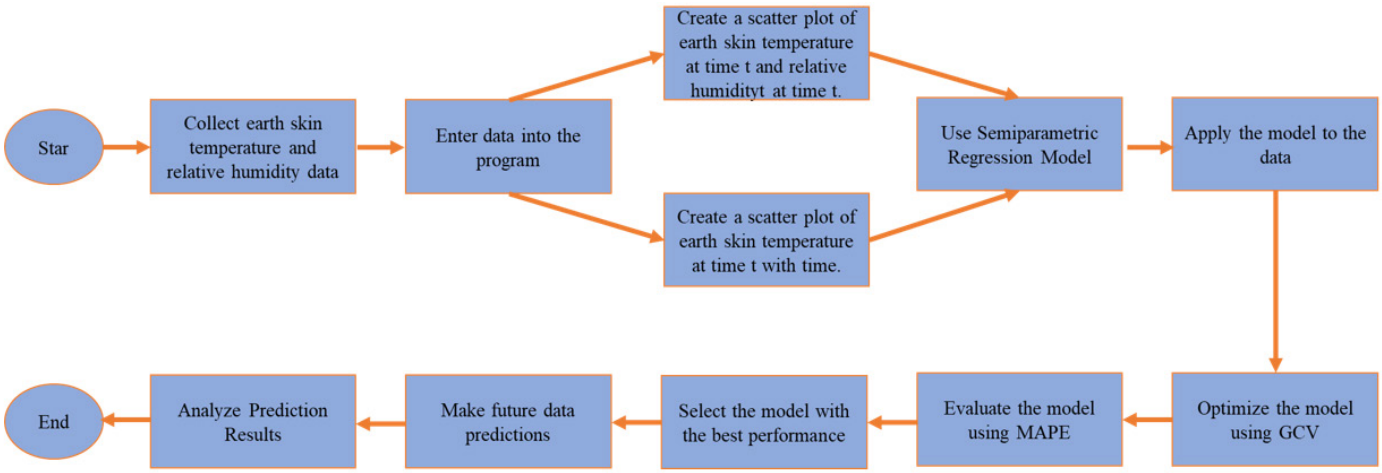


Figure 1. Flowchart of the study on predicting earth surface temperature variations in Sragen

RESULTS AND DISCUSSION

The data used in this study is 100 data, which is divided into training data and data testing. For the distribution of training data, it was divided consecutively, namely 70, 80, and 90 data. As for the distribution of testing data, it is divided consecutively, namely 30, 20, and 10 data. The first thing in this study is to check the relationship or correlation between earth surface temperature and relative humidity on the three training data. This check was carried out using statistical analysis of correlation which aims to determine the strength and direction of the relationship between the two earth surface temperature variables symbolized by EST and relative humidity symbolized by RH. To measure correlation, the Pearson correlation coefficient is used, which gives a value between -1 to 1. Positive values indicate a positive correlation, while negative values indicate negative correlations. The following figure (figure 2) shows the correlation plots between EST and RH for 70 training data, 80 training data, and 90 training data.

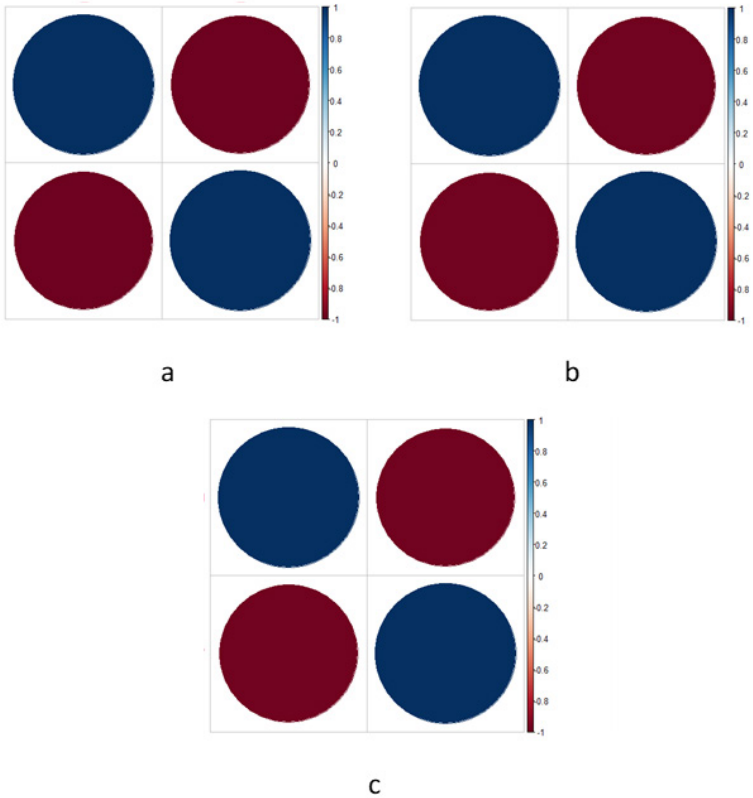


Figure 2. (a) Correlation plot of 70 training data. (b) Correlation plot of 80 training data. (c) Correlation plot of 90 training data

https://doi.org/10.56294/dm2025890

Based on figure 2, the following are the results of the correlation values in the form of a table based on the Correlation matrix between EST and RH:

Table 2. Correlation matrix for each N

	N =70		N = 80		N = 90	
	EST	**RH**	**EST**	**RH**	**EST**	**RH**
EST	1,0000000	-0,9700351	1,0000000	-0,9656242	1,0000000	-0,9640268
RH	-0,9700351	1,0000000	-0,9656242	1,0000000	-0,9640268	1,0000000

In table 2, with the correlation values shown in the matrix -0,9700351, -0,9656242, and -0,9640268, which show a correlation between EST and RH. A negative correlation value means that the relationship between temperature and relative humidity is inversely proportional, that is, when the earth surface temperature rises, the relative humidity value will decrease, and vice versa.

The next stage is to form a time series plot on each training data, namely earth surface temperature with relative humidity to check the distribution of data. Figure 3 shows scatter plots of the earth surface temperature (EST) and the relative humidity (RH).

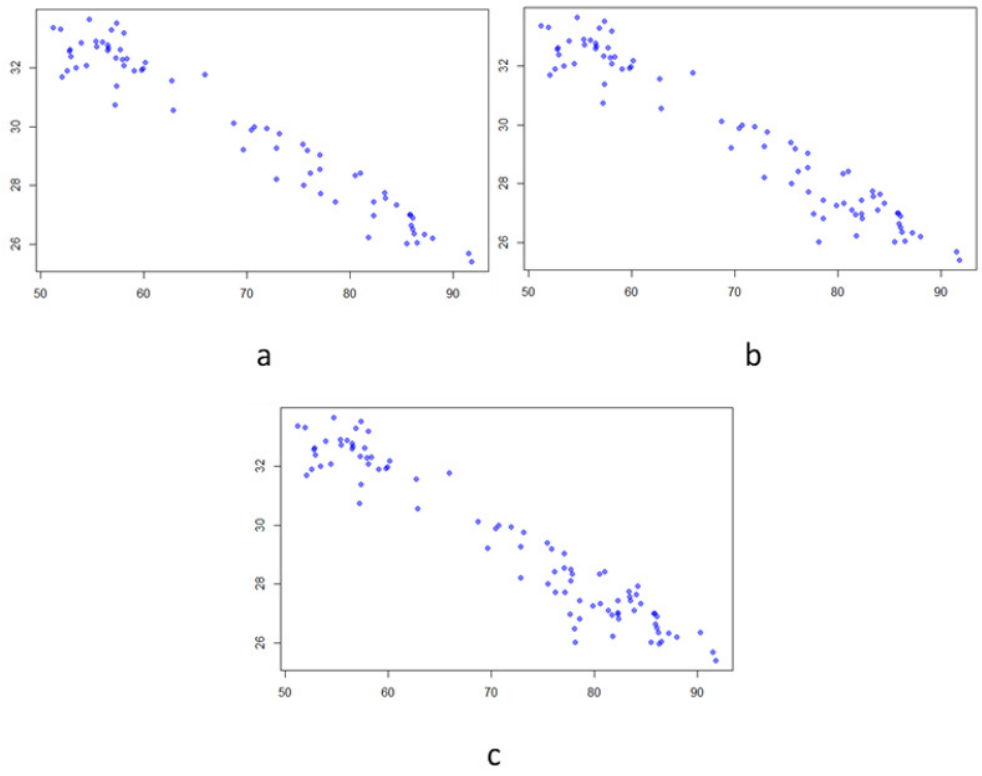


Figure 3. (a) Scatter plot of 70 training data. (b) Scatter plot of 80 training data. (c) Scatter plot of 90 training data

The results of the scatter plots in figure 3 show that the three training data have a pattern that follows a linear assumption with a descending line which means showing the parametric negative relationship between the three training data between EST and RH. To strengthen the assumption, a linearity test was carried out by following a linear regression model and the following results were obtained:

Table 3. Significance test of linear model parameters

	N = 70	N = 80	N = 90
	Pr(>|t|)	**Pr(>|t|)**	**Pr(>|t|)**
(Intercept)	<2e-16 ***	<2e-16 ***	<2e-16 ***
RH2M	<2e-16 ***	<2e-16 ***	<2e-16 ***

https://doi.org/10.56294/dm2025890

It can be seen in the table above for each training data competition that the coefficient value is less than α=0,05 the meaning of the significant linear model coefficient. It can be concluded that the variables between EST and RH have a linear relationship. Next, we create a scatter plot between EST and time, and the scatter plot is shown in figure 4.

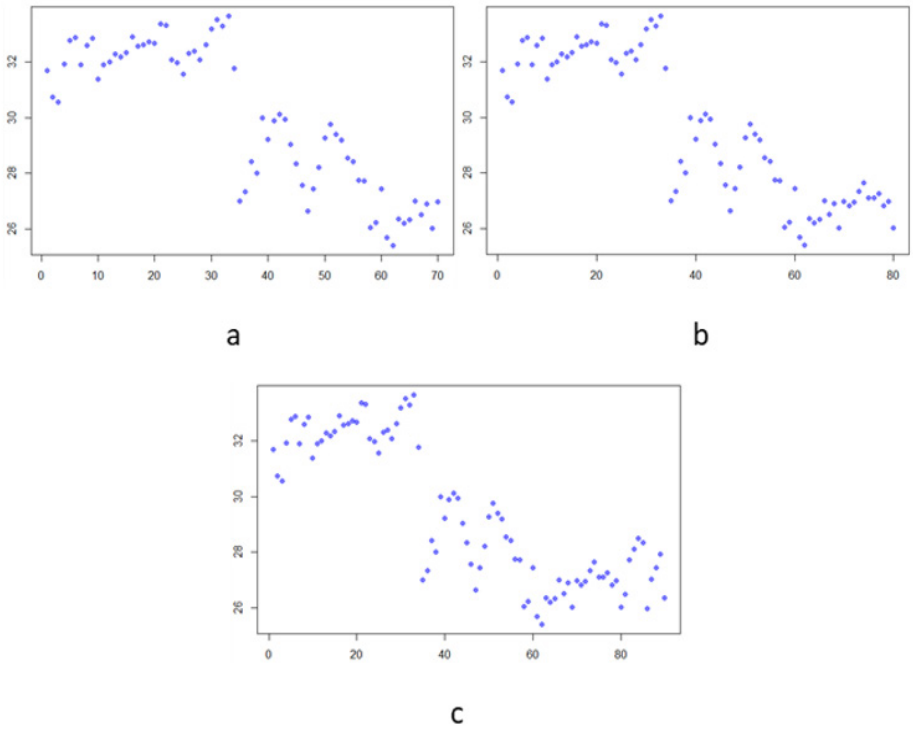
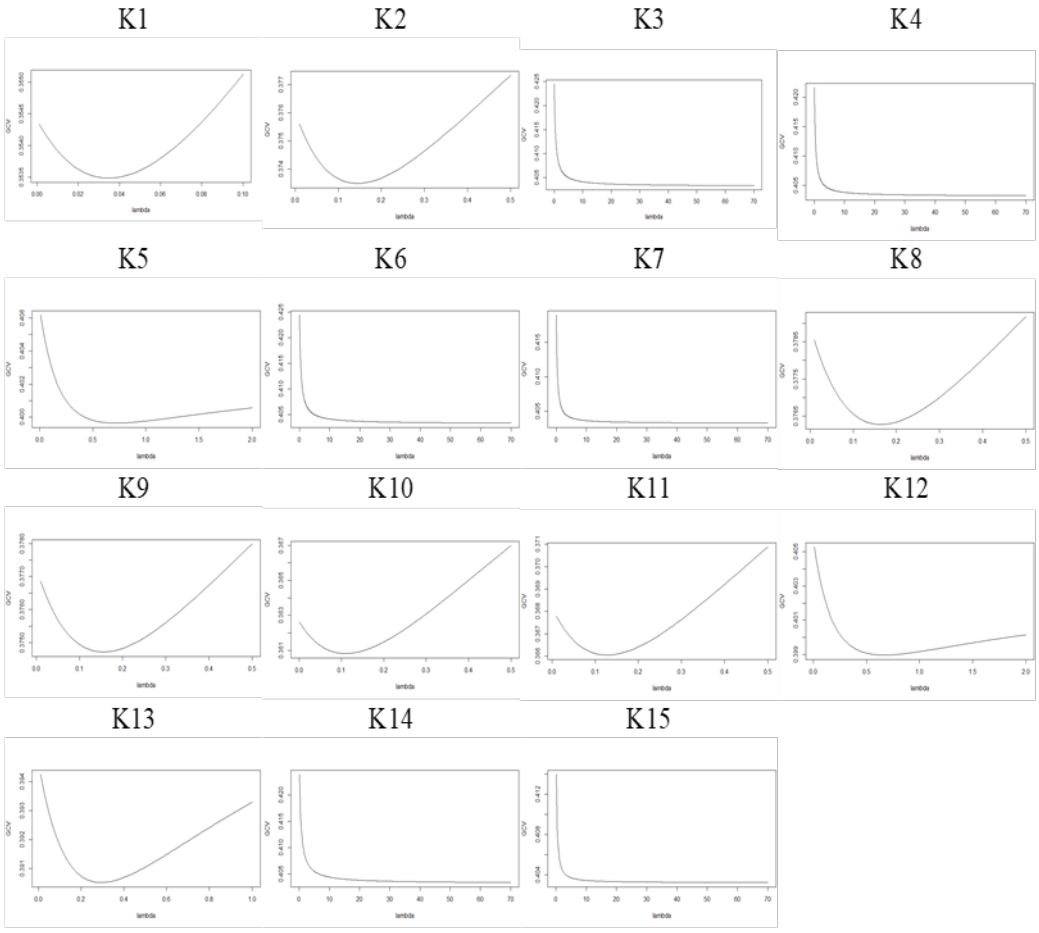
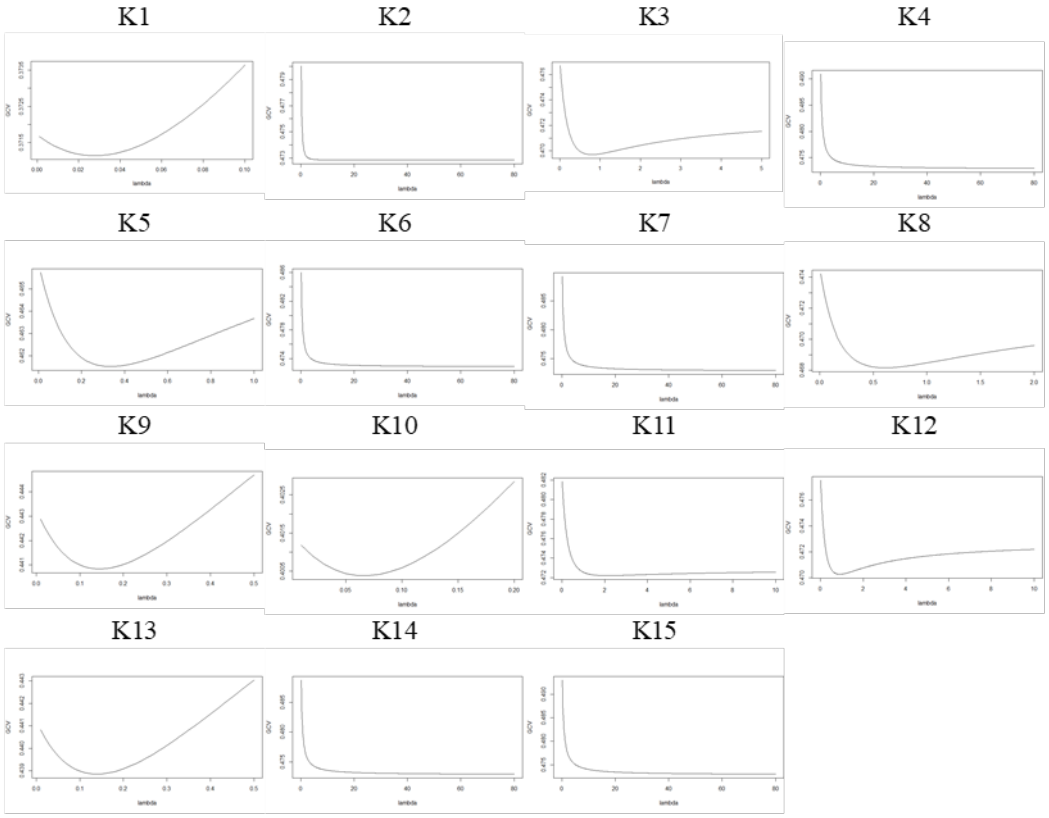


Figure 4. (a) Scatter plot of 70 training data. (b) Scatter plot of 80 training data. (c) Scatter plot of 90 training data

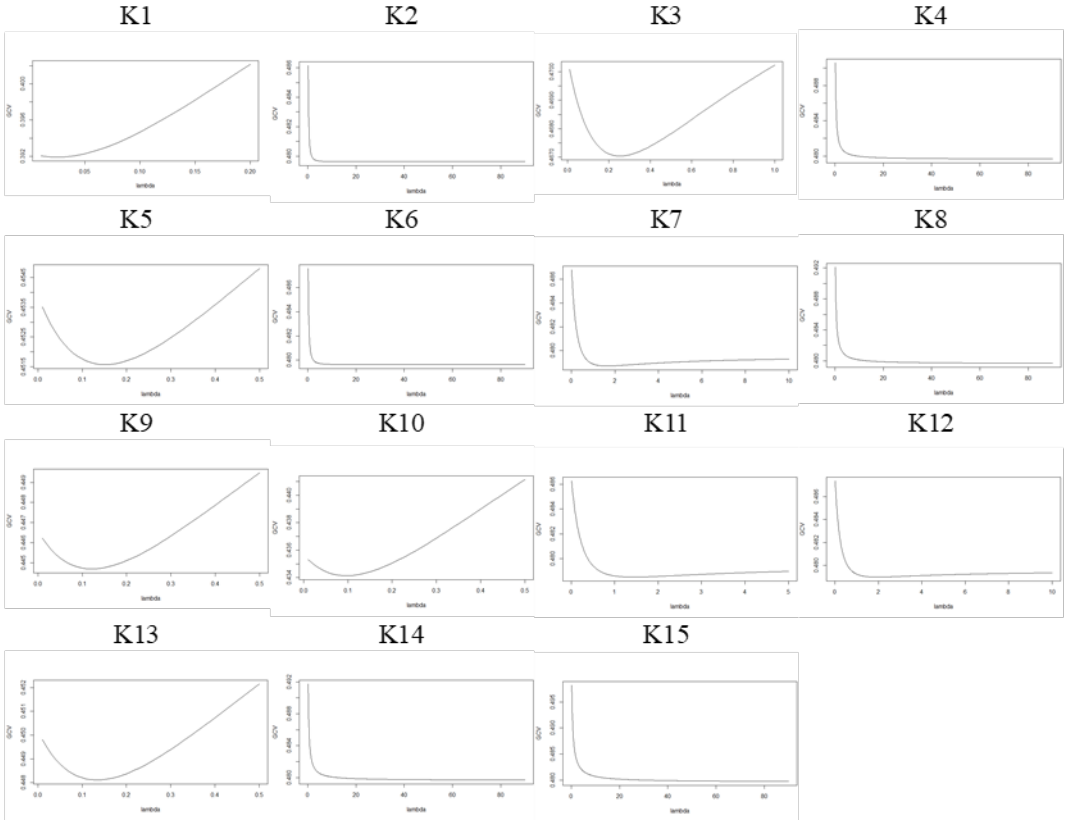


(a)

https://doi.org/10.56294/dm2025890



(b)



(c)

Figure 5. (a) GCV plot for N=70. (b) GCV plot for N=80. (c) GCV plot for N=90

Based on figure 4, the scatter plot between EST and time does not seem to form a specific pattern. This indicates that a nonparamteric regression approach can be used. Therefore, knowing that the functional

https://doi.org/10.56294/dm2025890

relationship between EST and RH is linear while the functional relationship between EST and time does not form a specific pattern, the semi-parametric regression approach is used in this case. This approach involves finding the minimum Generalized Cross Validation (GCV) value using a Fourier series estimator. In this study, the limit for the Fourier series coefficient is 15. The GCV plots for each Fourier coefficient of all three training data are presented in figure 5.

Based on figure 5 it can be seen that each Fourier coefficient has its minimum GCV value and the corresponding lambda, respectively. The values of each GCV and lambda on each of the Fourier coefficients are given in table 4.

Table 4. GCV and Lambda values from K = 1 to K = 10 for each N

N = 70			N = 80			N = 90		
K	**GCV**	**Lambda**	**K**	**GCV**	**Lambda**	**K**	**GCV**	**Lambda**
1	0,3534871	0,035	1	0,3711413	0,028	1	0,3918924	0,02
2	0,3734924	0,14	2	0,4728418	22,8	2	0,4796277	17,9
3	0,4033425	69	3	0,4697023	0,79	3	0,467037	0,25
4	0,4032988	69	4	0,4729677	79	4	0,4796707	89
5	0,3996436	0,72	5	0,4615144	0,34	5	0,4515693	0,15
6	0,4033412	69	6	0,4729007	79	6	0,4796329	62,9
7	0,403263	69	7	0,4729464	79	7	0,4787159	1,63
8	0,3762748	0,16	8	0,4681612	0,61	8	0,479689	89
9	0,3747158	0,16	9	0,4408315	0,15	9	0,4446873	0,12
10	0,3608138	0,109	10	0,4003833	0,07	10	0,4341343	0,1
11	0,3660474	0,13	11	0,4722082	2,13	11	0,4785247	1,44
12	0,398974	0,67	12	0,470282	0,9	12	0,4789971	1,96
13	0,3905162	0,29	13	0,4388478	0,14	13	0,4481078	0,13
14	0,4033847	69	14	0,4729398	79	14	0,4796829	89
15	0,403237	69	15	0,4729951	79	15	0,4797591	89

Furthermore, based on the figure 5 and table 4, the GCV value for each Fourier coefficient in the training data has a minimum GCV value at the first fourier coefficient for N=70, then the fourier coefficient one for N=80 and the fourier coefficient one for N=90. It can be concluded that the best semiparametric model to be where is the sixth Fourier coefficient with a minimum GCV value of N=70 from 0,3534871 and lambda 0,035. Then lie down in the first quadruier coefficient with a minimum GCV value of N=80 from 0,3711413 and lambda 0,028. The first Fourier coefficient with a minimum GCV value of N=90 is 0,3918924 and lambda is 0,02. The best semiparametric model based on the characteristics of the minimum GCV using a Fourier series estimator can be written as follows:

For N=70, we have:

$$y_i = -2{,}583678e-19 + (-0{,}1863643)\,u_{1i} + 42{,}82742 + (-0{,}0007323958)\cos(2\pi t_i) + 0{,}00178221\sin(2\pi t_i)$$

For $N = 80$, we have:

$$y_i = -2{,}609251e-17 + (-0{,}183842)\,u_{1i} + 42{,}55465 + -0{,}0006943916\cos(2\pi t_i) + 0{,}002908398\sin(2\pi t_i)$$

For $N = 90$, we have:

$$y_i = 6{,}291741e-19 + (-0{,}16612)\,u_{1i} + 41{,}26539 + 0{,}0002542628\ \cos(2\pi t_i) + 0{,}003947357\ \sin(2\pi t_i)$$

After obtaining the best semiparametric regression model on each training data, the next step in this study is to measure the accuracy and performance of the semiparametric regression model using a Fourier series estimator. This is done using the Mean Absolute Percentage Error (MAPE) for each of the best semiparametric regression models obtained. MAPE measures forecasting error in the form of percentages, with smaller values indicating higher accuracy. Plot of the actual data and predictions of each N along with the Fourier Coefficient are presented in figure 6.

https://doi.org/10.56294/dm2025890

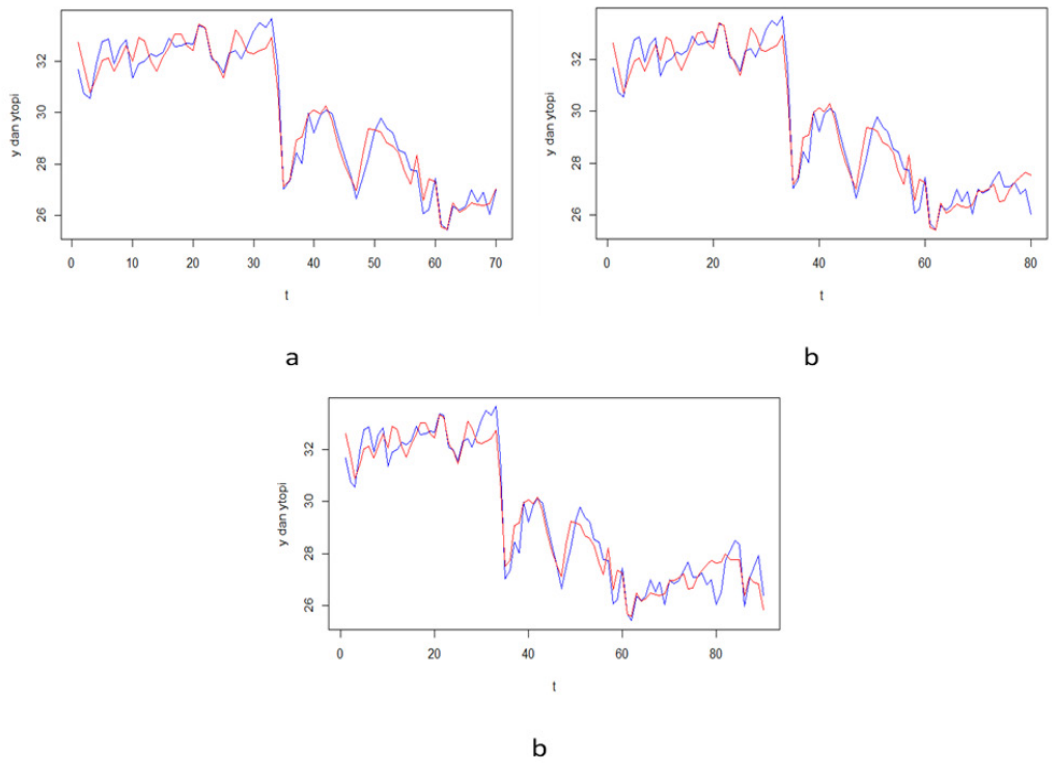


Figure 6. Plot of actual data and predictions for (a) N = 70. (b) N = 80. (c) N = 90

Based on figure 6, it can be seen that the forecasting data demonstrates good performance. For each N with the best Fourier coefficients, the model shows a strong ability to explain data variability and a low error rate in the testing data. Therefore, the semi-parametric regression model using the Fourier series estimator can be used to evaluate forecasting performance by comparing the predicted data with the testing data. The plots of the predicted data compared to the testing data based on the best semi-parametric regression model using the Fourier series estimator are presented in figure 7.

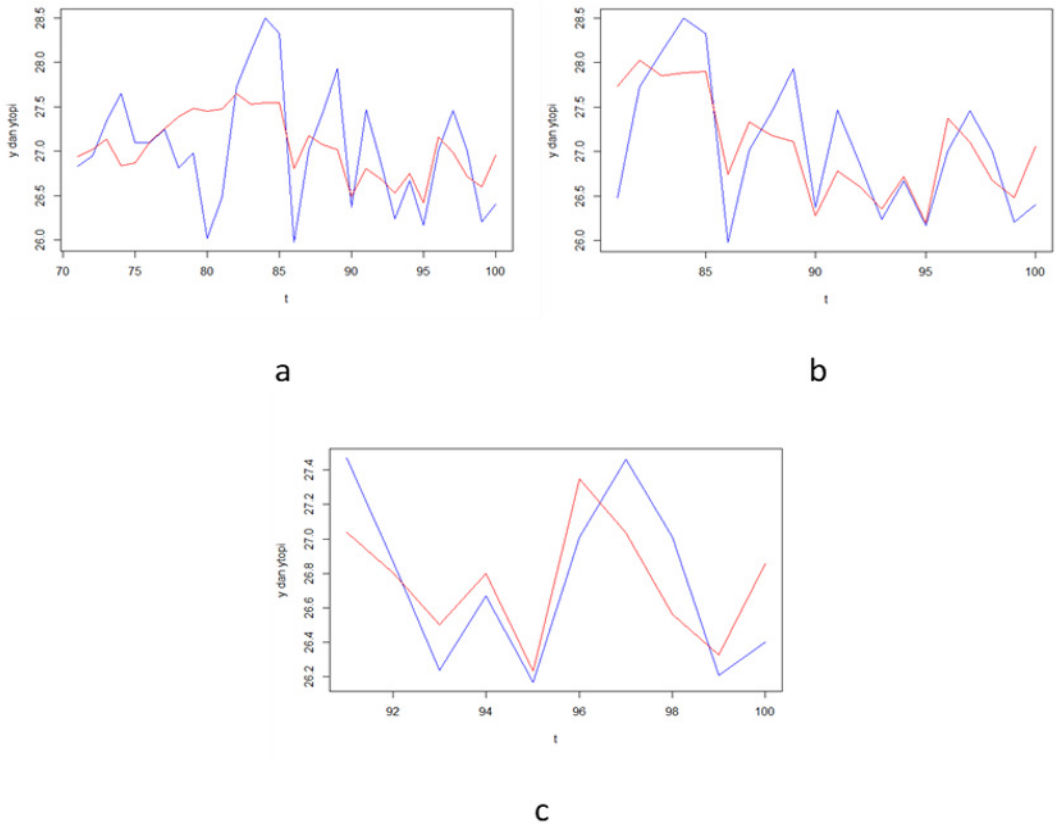


Figure 7. Plots of actual data and predictions for (a) N = 30. (b) N = 20. (c) N = 10

https://doi.org/10.56294/dm2025890

Table 5. MAPE values from the testing data for each N		
N=30	**N=20**	**N=10**
MAPE	**MAPE**	**MAPE**
1,606545	1,518221	1,018482

Table 6. Values of GCV for N=70, N=80, and N=90			
	N=70	**N=80**	**N=90**
K	**GCV**	**GCV**	**GCV**
1	0,3543943	0,3717146	0,3923186
2	0,3760212	0,4837378	0,4894241
3	0,4249463	0,4771473	0,4704768
4	0,4220604	0,4949033	0,493964
5	0,4066091	0,4661733	0,4538796
6	0,4248638	0,4897925	0,4908476
7	0,4193882	0,4931437	0,4871929
8	0,3789435	0,4746563	0,4955042
9	0,3772579	0,4432858	0,4465855
10	0,3626551	0,4015302	0,4356699
11	0,3681756	0,4822778	0,48666
12	0,4056925	0,477965	0,4877531
13	0,3946605	0,4412212	0,4501726
14	0,4279155	0,4926211	0,4951409
15	0,4176262	0,4969854	0,5017819

It can be seen in figure 7 that the forecasted data is not much different from the actual data, where this assumption is reinforced by obtaining the MAPE values given in table 5.

Additionaly, the GCV (Generalized Cross-Validation) values with lower limit, upper limit, and increment of zero for N=70, N=80, and N=90 are presented in table 6.

Next, in the following table (table 7), the RMSE and MAPE values are presented at the best Fourier coefficient (K), namely K=1, for N=10 with a lambda value of 0,035, for N=20 with a lambda value of 0,028, and for N=300 with a lambda value of 0,02 which is used for testing data.

Table 7. Values of RMSE and MAPE for N-30. N=20, and N=10					
N=30		**N=20**		**N=10**	
RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
0,5590762	1,606545	0,5074921	1,518221	0,3158802	1,018482

Furthermore, comparison of GCV (Generalized Cross-Validation) values based on Ordinary Least Squares (OLS) method and Penalized Least Squares (PLS) method for N=70, N=80, and N=90 is presented in the following table.

Table 8. Comparison of GCV values based on OLS and PLS methods for N=70, N=80, and N=90						
	N=70		**N=80**		**N=90**	
K	**GCV Ordinary Least Squares**	**GCV Penalized Least Squares**	**GCV Ordinary Least Squares**	**GCV Penalized Least Squares**	**GCV Ordinary Least Squares**	**GCV Penalized Least Squares**
1	0,3543943	0,3534871	0,3717146	0,3711413	0,3923186	0,3918924
2	0,3760212	0,3734924	0,4837378	0,4728418	0,4894241	0,4796277
3	0,4249463	0,4033425	0,4771473	0,4697023	0,4704768	0,467037
4	0,4220604	0,4032988	0,4949033	0,4729677	0,493964	0,4796707
5	0,4066091	0,3996436	0,4661733	0,4615144	0,4538796	0,4515693
6	0,4248638	0,4033412	0,4897925	0,4729007	0,4908476	0,4796329

https://doi.org/10.56294/dm2025890

7	0,4193882	0,403263	0,4931437	0,4729464	0,4871929	0,4787159
8	0,3789435	0,3762748	0,4746563	0,4681612	0,4955042	0,479689
9	0,3772579	0,3747158	0,4432858	0,4408315	0,4465855	0,4446873
10	0,3626551	0,3608138	0,4015302	0,4003833	0,4356699	0,4341343
11	0,3681756	0,3660474	0,4822778	0,4722082	0,48666	0,4785247
12	0,4056925	0,398974	0,477965	0,470282	0,4877531	0,4789971
13	0,3946605	0,3905162	0,4412212	0,4388478	0,4501726	0,4481078
14	0,4279155	0,4033847	0,4926211	0,4729398	0,4951409	0,4796829
15	0,4176262	0,403237	0,4969854	0,4729951	0,5017819	0,4797591

Finally, the difference curves of GCV values based on Ordinary Least Squares (OLS) and Penalized Least Squares (PLS) methods are presented in figure 8.

In the modeling patterns of cyclical climate data, Fourier series often demonstrates a clear advantage over methods like splines and kernels due to its ability to effectively represent periodicities. Fourier series excel at capturing the harmonic structure inherent in climate data, while splines and kernels can struggle to represent these periodicities without being prone to over- or under-fitting. The superiorities of Fourier series compared to splines and kernels are that the Fourier series has strengths in the following areas: (a). Decomposes cyclical patterns: Fourier series excels at breaking down time series data into a sum of sinusoidal functions with varying frequencies and amplitudes, directly reflecting the cyclical nature of climate data; (b). Efficient for periodic signals: When data exhibits a clear periodic trend (e.g., daily, monthly, or annual cycles), Fourier series provides a concise and accurate representation; and (c). Well-suited for long-term climate analysis: By capturing the dominant frequencies and amplitudes, Fourier series can identify and analyze long-term trends in climate data. In the contrary, the splines and kernels have limitations in these áreas: (a). Less efficient for cyclical patterns: Splines and kernels might not be as effective as Fourier series in capturing the underlying cyclical structures in climate data; and (b). Can be computationally expensive: Splines and kernel methods can be computationally intensive, especially for large datasets.[54] In summary, for climate data where cyclical patterns are dominant (e.g., seasonal temperatures, solar cycles), Fourier series provide a more efficient and effective approach compared to splines or kernels. Splines and kernels are better suited for capturing non-periodic variations or interpolating missing data.[54]

Figure 8. Difference curves of GCV values based on OLS and PLS methods

https://doi.org/10.56294/dm2025890

CONCLUSIONS

This study applies a semi-parametric regression model with a Fourier Series estimator based on PLS to examine the relationship between earth surface temperature and relative humidity in Sragen regency. By combining parametric and nonparametric components, the model effectively handles fluctuating data patterns in three training scenarios (N = 70, 80, 90) with optimal Fourier coefficients of 1, 1, 1 and lambda values of 0,035, 0,028, and 0,02, respectively. The estimation results show minimal differences between the actual and predicted values, indicating reliable performance. This model successfully made good predictions for testing data sizes of 30, 20, and 10, with MAPE values of 1,606545, 1,518221, and 1,018482, respectively. These results highlight the effectiveness of the Fourier-based semi-parametric approach in modeling the inverse relationship between earth surface temperature (EST) and relative humidity (RH) while maintaining accuracy in analyzing complex data. Future research could extend this approach to other regions or variables, enhancing its application for climate-related studies and policy development.

BIBLIOGRAPHIC REFERENCES

1. Khoirunnisa F, Rahmawati Y. Komparasi 2 metode cluster dalam pengelompokan intensitas bencana alam di Indonesia. J Informatika dan Teknik Elektro Terapan [Internet]. 2024 [cited 2025 April 14]; 12(1): 68-79. Available from: https://journal.eng.unila.ac.id/index.php/jitet/article/view/3619/1574.

2. Sari R L, Haeruddin H. Studi pendahuluan potensi carbon capture storage (CCS) melalui identifikasi sumber emisi CO2 beserta tinjauan geologi di daerah Probolinggo, Jember, dan Bondowoso, Jawa Timur. J Teknol Lingkungan Lahan Basah [Internet]. 2024. [cited 2025 April 14]; 12(1): 137–142. Available from: https://jurnal.untan.ac.id/index.php/jmtluntan/article/view/71804.

3. Subiyanto A D. Diplomasi iklim: upaya menyelamatkan bumi dari krisis iklim. PENDIPA: J Sci Edu [Internet]. 2024. [cited 2025 April 14]; 8(1): 27–34. Available from: https://doi.org/10.33369/pendipa.8.1.27-34.

4. Pamungkas G D, Iemaaniah Z M, Bustan B. Analisis karakteristik iklim dan hujan pada lahan pertanian di Kecamatan Kediri Kabupaten Lombok Barat. AGROTEKSOS [Internet]. 2023. [cited 2025 April 14]; 33(3): 855–66. Available from: https://agroteksos.unram.ac.id/index.php/Agroteksos/article/view/858.

5. Ardelia E. Proyeksi penerapan pajak karbon dalam upaya menekan emisi gas rumah kaca pada sektor pertanian dan perkebunan di Indonesia. Innovative: J Soc Sci Res [Internet]. 2023. [cited 2025 April 14]; 3(4): 9070–80. Available from https://j-innovative.org/index.php/Innovative/article/view/4686.

6. Panunggul V B, Yusra S, Khaerana K, Tuhuteru S, Fahmi D A, Laeshita P, et al. Pengantar Ilmu Pertanian. Penerbit Widina [Internet]. 2023. [cited 2025 April 17]. Available from: https://repository.penerbitwidina.com/publications/564795/pengantar-ilmu-pertanian.

7. Muhammad F, Maryono M, Hadiyanto H, Retnaningsih T, Hastuti R B. Reboisasi sebagai upaya konservasi di KHDTK dipoforest hutan penggaron Kabupaten Semarang. J Pasopati [Internet]. 2023. [cited 2025 April 17]; 5(1): 29–36. Available from: https://ejournal2.undip.ac.id/index.php/pasopati/article/view/17135/0.

8. Adib M. Pemanasan global, perubahan iklim, dampak dan solusinya di sektor pertanian. BioKultur [Internet]. 2014. [cited 2025 April 17]; 3(2): 420–9. Available from: https://journal.unair.ac.id/download-fullpapers-bkbbfe09eddcfull.pdf.

9. Pratama R. Efek rumah kaca terhadap bumi. Buletin Utama Teknik [Internet]. 2019. [cited 2025 April 17]; 14(2): 120–6. Available from: https://jurnal.uisu.ac.id/index.php/but/article/view/1096.

10. Rozci F. Dampak Perubahan iklim terhadap sektor pertanian padi. J Ilmiah Sosio Agribis [Internet]. 2024. [cited 2025 April 17]; 23(2): 108–16. Available from: https://journal.uwks.ac.id/index.php/sosioagribis/article/view/3476.

11. Dai X, Liu Q, Huang C, Li H. Spatiotemporal variation analysis of the fine-scale heat wave risk along the Jakarta-Bandung high-speed railway in Indonesia. Int J Environ Res Public Health [Internet]. 2021. [cited 2025 April 17]; 18(22): 12153. Available from: https://www.mdpi.com/1660-4601/18/22/12153.

12. Kadir E A, Rosa S L, Syukur A, Othman M, Daud H. Forest fire spreading and carbon concentration identification in tropical region Indonesia. Alexandria Eng J [Internet]. 2022. [cited 2025 April 17]; 61(2):

1551–61. Available from: https://www.sciencedirect.com/science/article/pii/S1110016821004191.

13. Jannah A N. Hubungan perubahan cuaca dengan indeks kecerahan matahari, suhu lingkungan dan kelembapan udara di Desa Karanganyar. Karst: J Pendidikan Fisika dan Terapannya [Internet]. 2021. [cited 2025 April 17]; 4(1): 27–32. Available from: https://ejournals.umma.ac.id/index.php/karts/article/view/929.

14. BNPB Kabupaten Sragen. Sragen kekeringan saat beberapa wilayah lain alami banjir. Minanews [Internet]. 2020. Available from: https://minanews.net/sragen-alami-kekeringan-saat-beberapa-wilayah-lain-alami-banjir/.

15. BPS Kabupaten Sragen. Kabupaten Sragen dalam angka 2021. Badan Pusat Statistik Kabupaten Sragen [Internet]. 2021. [cited 2025 April 17]. Available from: https://sragenkab.bps.go.id/id/publication/2021/02/26/2774bd589e365acf0be23cd2/kabupaten-sragen-dalam-angka-2021.html.

16. Aji N P W. Analisis urban heat island di Kabupaten Sragen tahun 2020. Geadidaktika [Internet]. 2023. [cited 2025 April 17]; 3(2): 167–82. Available from: https://jurnal.uns.ac.id/geadidaktika/article/view/76852.

17. Aripbilah S N, Suprapto H. Analisis kekeringan di Kabupaten Sragen dengan metode Palmer, Thornthwaite, dan Standardized Precipitation Index. J Sumber Daya Air [Internet]. 2021. [cited 2025 April 17]; 17(2): 111–24. Available from: https://jurnalsda.pusair-pu.go.id/index.php/JSDA/article/view/742.

18. Safitri A, Purnamasari E, Rahmawati F, Dasairy H F, Vebiani N A. Promosi kesehatan tentang pencegahan demam berdarah dengue pada remaja di Karang Taruna "Tunas Muda" Desa Sukomarto, Kecamatan Sidoharjo, Kabupaten Sragen. J Pengabdian Komunitas [Internet]. 2023. 2(03): 88–96. Available from: https://jurnalpengabdiankomunitas.com/index.php/pengabmas/article/view/56/70.

19. Awatara I G P D, Susila L N, Saryanti E. Pendampingan program kampung iklim (Proklim) di Desa Bon Agung, Kecamatan Tanon, Kabupaten Sragen. Wasana Nyata [Internet]. 2023. [cited 2025 April 17]; 7(2):93–7. Available from: https://e-journal.stie-aub.ac.id/index.php/wasana_nyata/article/view/1601.

20. Chamidah N, Lestari B. Analisis Regresi Nonparametrik dengan Perangkat Lunak R. Airlangga University Press (AUP). 2022. Available from: https://omp.unair.ac.id/aup/catalog/book/888.

21. Chamidah N, Lestari B, Saifudin T. Modeling of blood pressures based on stress score using least square spline estimator in bi-response nonparametric regression. Int J Innov Creat Change (IJICC) [Internet]. 2019. [cited 2025 April 21]; 5(3):1200–16. Available from: https://www.ijicc.net/images/Vol_5_Iss_3/Part_2_2020/5321_Chamidah_2019_E_R.pdf.

22. Fatmawati, Budiantara I N, Lestari B. Comparison of smoothing and truncated splines estimators in estimating blood pressure models. Int J Innov Creat Change (IJICC) [Internet]. 2019. [cited 2025 April 21]; 5(3):1177–99. Available from: https://repository.unair.ac.id/114278/1/C14.%20Fulltext.pdf.

23. Chamidah N, Lestari B, Massaid A, Saifudin T. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach. Commun Math Biol Neurosci [Internet].2020. [cited 2025 April 21]; 2020(72): 1–12. Available from: https://scik.org/index.php/cmbn/article/view/4963.

24. Chamidah N, Yonani Y S, Ana E, Lestari B. Identification the number of Mycobacterium Tuberculosis based on sputum image using local linear estimator. Bull Electric Eng Informatics (BEEI) [Internet]. 2020. [cited 2025 April 21]; 9(5): 2109 –16. Available from: https://doi.org/10.11591/eei.v9i5.2021.

25. Tohari A, Chamidah N, Fatmawati, Lestari B. Modelling the number of HIV and AIDS cases in East Java using biresponse multipredictor negative binomial regression based on local linear estimator. Commun Math Biol Neurosci [Internet]. 2021. [cited 2025 April 21]; 2021(73): 1–17. Available from: https://scik.org/index.php/cmbn/article/view/5652.

26. Lestari B, Chamidah N, Aydin D, Yilmaz E. Reproducing kernel Hilbert space approach to multiresponse smoothing spline regression function. Symmetry [Internet]. 2022. [cited 2025 April 21]; 14(11) 2227: 1–22. Available from: https://www.mdpi.com/2073-8994/14/11/2227.

27. Aydin D, Yilmaz E, Chamidah N, Lestari B, Budiantara I N. Right-censored nonparametric regression with measurement error. Metrika (Int J Theor Appl Stats) [Internet]. 2024. [cited 2025 April 21]; 87(3). Available from: https://doi.org/10.1007/s00184-024-00953-5.

28. Chamidah N, Lestari B, Saifudin T, Rulaningtyas R, Wardhani P, Budiantara I N, Aydin D. Determining the number of malaria parasites on blood smears microscopic images using penalized spline nonparametric Poisson regression. Commun Math Biol Neurosci [Internet]. 2024. [cited 2025 April 21]; 2024(60): 1-19. Available from: https://scik.org/index.php/cmbn/article/view/8578.

29. Chamidah N, Lestari B, Budiantara I N, Aydin D. (2024). Estimation of multiresponse multipredictor nonparametric regression model using mixed estimator. Symmetry [Internet]. 2024. [cited 2025 April 21]; 16(4) 386:1–25. Available from: https://www.mdpi.com/2073-8994/16/4/386.

30. Chamidah N, Lestari B, Larasati T N, Muniroh L. (2024). Designing Z-score standard growth charts based on height-for-age of toddlers using local linear estimator for determining stunting. AIP Conf Proc [Internet]. 2024. [cited 2025 April 21]; 3083(1) 030002. Available from: https://doi.org/10.1063/5.0225156.

31. Chamidah N, Lestari B, Susilo H, Alsagaff M Y, Budiantara I N, Aydin D. Spline estimator in nonparametric ordinal logistic regression model for predicting heart attack risk. Symmetry [Internet] 2024. [cited 2025 April 21]: 16(11) 1440: 1-23. Available from: https://www.mdpi.com/2073-8994/16/11/1440.

32. Chamidah N, Lestari B, Wulandari A Y, Muniroh L. Z-score standard growth chart design of toddler weight using least square spline semiparametric regression. AIP Conf Proc [Internet]. 2021. [cited 2025 April 21]; 2329: 060031. Available from: https://doi.org/10.1063/5.0225156.

33. Chamidah N, Lestari B, Budiantara I N, Saifudin T, Rulaningtyas R, Aryati A, Wardani P, Aydin D. Consistency and asymptotic normality of estimator for parameters in multiresponse multipredictor semiparametric regression model. Symmetry [Internet]. 2022. [cited 2025 April 21]; 14(2) 336:1-18. Available from: https://www.mdpi.com/2073-8994/14/2/336.

34. Chamidah N, Zaman B, Muniroh L, Lestari B. Multiresponse semiparametric regression model approach to standard growth charts design for assessing nutritional status of East Java toddlers. Commun Math Biol Neurosci [Internet] 2023. [cited 2025 April 21]; 2023(30): 1-23. Available from: https://scik.org/index.php/cmbn/article/view/7814.

35. Lestari B, Chamidah N, Budiantara I N, Aydin D. (2023). Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator. J King Saud Univ-Sci [Internet]. 2023. [cited 2025 April 21]; 35(5): 102664. Available from: https://doi.org/10.1016/j.jksus.2023.102664.

36. Aydin D, Yilmaz E, Chamidah N, Lestari B. Right-censored partially linear regression model with error in variables: application with carotid endarterectomy dataset. Int J Biostatistics [Internet]. 2023. [cited 2025 April 21]; 20(1): 1–34. Available from: https://doi.org/10.1515/ijb-2022-0044.

37. Utami T W, Chamidah N, Saifudin T, Lestari B, Aydin D. Estimation of biresponse semiparametric regression model for longitudinal data using local polynomial kernel estimator. Symmetry [Internet]. 2025. [cited 2025 April 21]; 17(3) 392: 1-22. Available from: https://www.mdpi.com/2073-8994/17/3/392.

38. Selingerova I, Katina S, Horova I. Comparison of parametric and semiparametric survival regression models with kernel estimation. J Stat Comput Simul [Internet]. 2021. [cited 2025 April 21]; 91(13): 2717–39. Available from: https://www.doi.org/10.1080/00949655.2021.1906875.

39. Lestari B, Fatmawati, Budiantara I N, Chamidah N. Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and kernel estimators approaches. J Phys: Conf Ser [Internet]. 2019. [cited 2025 April 21]; 1397(1):012064. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/1397/1/012064.

40. Eubank R L. Nonparametric Regression and Spline Smoothing. CRC press; Boca Raton, 1999. Available from: https://doi.org/10.1201/9781482273144.

41. Chamidah N, Zaman B, Muniroh L, Lestari B. Designing local standard growth charts of children in East Java province using a local linear estimator. Int J Innov Create Change (IJICC) [Internet]. 2020. [cited 2025 April 21]; 13(1): 45–67. Available from: https://www.ijicc.net/images/vol_13/13104_Chamidah_2020_E_R.pdf.

42. Chamidah N, Gusti K H, Tjahjono E, Lestari B. Improving of classification accuracy of cyst and tumor using local polynomial estimator. Telkomnika (Telecommunication Computing Electronics and Control) [Internet]. 2019. [cited 2025 April 21]; 17(3): 1492-500. Available from: https://telkomnika.uad.ac.id/index.php/TELKOMNIKA/article/view/12240.

43. Pane R, Budiantara I N, Zain I, Otok B W. Parametric and nonparametric estimators in Fourier series semiparametric regression and their characteristics. Appl Math Sci [Internet]. 2014. [cited 2025 April 21]; 8(102): 5053–64. Available from: http://dx.doi.org/10.12988/ams.2014.46472.

44. Khairunnisa L R, Prahutama A, Santoso R. Pemodelan regresi semiparametrik dengan pendekatan deret Fourier (Studi kasus: pengaruh indeks Dow Jones dan BI rate terhadap indeks harga saham gabungan). J Gaussian [Internet]. 2020. [cited 2025 April 21]; 9(1): 50–63. Available from: https://ejournal3.undip.ac.id/index.php/gaussian/article/view/27523.

45. Yao W, Weng Y, Catchmark J M. Improved cellulose X-ray diffraction analysis using Fourier series modeling. Cellulose [Internet]. 2020. [cited 2025 April 21]; 27(10): 5563–79. Available from: https://www.doi.org/10.1007/s10570-020-03177-8.

46. Chamidah N, Febriana S A, Ariyanto R A, Sahawaly R. Fourier series estimator for predicting international market price of white sugar. AIP Conf Proc. 2021. 2329(1): 060035. Available from: https://www.doi.org/10.1063/5.0042287.

47. Pasarella M D, Sifriyani S, Amijaya F D T. Nonparametrik regression model estimation with the Fourier series approach and its application to the accumulative Covid-19 data in Indonesia. BAREKENG [Internet]. 2022. [cited 2025 April 21]; 16(4): 1167–74. Available from: https://www.doi.org/10.30598/barekengvol16iss4pp1167-1174.

48. Amri I F, Chamidah N, Saifudin T, Purwanto D, Fadlurohman A, Ningrum A F, Amri S. Prediction of extrem weather using nonparametric regression approach with Fourier series estimators. Data and Metadata [Internet]. 2024. [cited 2025 April 21]; 4 (319): 1–12. Available from: https://www.doi.org/10.56294/dm2024319.

49. Gao J. Nonlinear Time Series: Semiparametric and Nonparametric Methods. CRC Press; Boca Raton, 2007. Available from: https://www.researchgate.net/publication/329483723_Nonlinear_time_series_Semiparametric_and_nonparametric_methods.

50. Fan J Q, Qi L, Tong X. Penalized least squares estimation with weakly dependent data. Sci China Math [Internet]. 2016. [cited 2025 April 21]; 59(12): 2335–54. Available from: https://www.doi.org/10.1007/s11425-016-0098-x.

51. Wang H, Li R, Tsai C L. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika [Internet]. 2007. [cited 2025 April 21]; 94(3): 553–68. Available from: https://doi.org/10.1093/biomet/asm053.

52. Moreno J J M, Pol A P, Abad A S, Blasco B C. Using the R-MAPE index as a resistant measure of forecast accuracy. Psicothema [Internet]. 2013. [cited 2025 April 21]; 25(4): 500–6. Available from: https://doi.org/10.7334/psicothema2013.23.

53. Bilodeau M. Fourier smoother and additive models. Canadian J Stats [Internet]. 2008. [cited 2025 April 20]; 20(3):257–69. Available from: https://doi.org/10.2307/3315313.

54. Baltazar J C, Claridge D. Study of cubic splines and Fourier series as interpolation techniques for filling in short periods of missing building energy use and weather data. Journal of Solar Energy Engineering [Internet]. 2002. [cited 2025 April 20]; 128(2). Available from: https://doi.org/10.1115/SED2002-1031.

ACKNOWLEDGMENT

The authors thank the Ministry of Education, Culture, Research and Technology, the Republic of Indonesia, for funding this research through the Doctoral Dissertation Research Grant (Hibah Penelitian Disertasi Doktor-PDD).

FINANCING

This research was funded by the Ministry of Education, Culture, Research and Technology (Mendikbudristek), the Republic of Indonesia through the Doctoral Dissertation Research Grant (Hibah Penelitian Disertasi Doktor-PDD) under contract No: 1749/B//UN3.LPPM/PT.01.03/2024 dated 13 June 2024.

CONFLICT OF INTEREST

The authors declare that no conflict of interest is associated with this research. The research process was conducted objectively and independently, without influence from any stakeholder that could benefit personally or institutionally.

AUTHORSHIP CONTRIBUTION

Conceptualization: Ihsan Fathoni Amri, Nur Chamidah, Toha Saifudin.
Data curation: Ihsan Fathoni Amri, Nur Chamidah, Toha Saifudin.
Formal analysis: Ihsan Fathoni Amri, Nur Chamidah, Budi Lestari.
Research: Ihsan Fathoni Amri, Nur Chamidah, Toha Saifudin.
Methodology: Ihsan Fathoni Amri, Nur Chamidah, Budi Lestari.
Project management: Ihsan Fathoni Amri, Nur Chamidah, Toha Saifudin.
Resources: Ihsan Fathoni Amri, Nur Chamidah.
Software: Ihsan Fathoni Amri, Toha Saifudin.
Supervision: Nur Chamidah, Toha Saifudin, Budi Lestari, Dursun Aydin.
Validation: Nur Chamidah, Toha Saifudin, Budi Lestari, Dursun Aydin.
Display: Ihsan Fathoni Amri.
Drafting - original draft: Ihsan Fathoni Amri.
Writing - proofreading and editing: Nur Chamidah, Budi Lestari.