Semiparametric Time Series Regression Using Exponential Complex Fourier Series for Temperature Prediction in the Tropical Rainforest of Samarinda

Regresión semiparamétrica de series temporales utilizando series de Fourier complejas exponenciales para la predicción de temperatura en la selva tropical de Samarinda

Andrea Tri Rian Dani [1,2], Nur Chamidah [3,4*], I Nyoman Budiantara [5], Budi Lestari [6], Ratna Kusuma [7], Dursun Aydin [8].

Highlights

- Semiparametric Time Series Regression Exponential Complex Fourier Series (STSR-ECFS) provides a flexible approach for modeling complex temperature patterns in tropical rainforest climates
- The model achieves high accuracy, with MAPE consistently less than 2% for both in-sample and out-of-sample data.
- The STSR-ECFS supports climate adaptation strategies and early warning systems in regions with strong seasonal variability.

Innovaciencia ISSN: 2346-075X

E- ISSN: 2346-075X
Innovaciencia2025;13(1); e5364
http//dx.doi.org/10.15649/2346075X.5364

ORIGINAL RESEARCH

How to cite this article:
Dani, A. T. R., Chamidah, N., Budiantara, I. N., Lestari, B., Kusuma, R., Aydin, D. Regresión semiparamétrica de series temporales utilizando series de Fourier complejas exponenciales para la predicción de temperatura en la selva tropical de Samarinda. Innovaciencia. 2025;13(1); e5364
http//dx.doi.org/10.15649/2346075X.5364

Received: June 19, 2025
Accepted: November 28, 2025
Published: December 17, 2025

Keywords:

Climate & Temperature, Semiparametric Fourier Series, Time Series Regression.

Palabras clave:

Clima y temperatura, series de Fourier semiparamétricas, regresión de series temporales.

ABSTRACT

Introduction. Forecasting temperature in tropical rainforest regions is challenging due to complex seasonality and nonlinear trends that conventional models often fail to capture. Addressing this, advanced modeling approaches are required for more accurate climate predictions. **Objectives.** This study aims to construct and apply the STSR-ECFS model for temperature forecasting in Samarinda, East Kalimantan, and to assess its predictive performance using standard accuracy metrics. **Materials and Methods.** The STSR-ECFS model combines an autoregressive parametric component with a nonparametric component constructed using the exponential complex Fourier series. The model was trained and validated on monthly temperature data from 2015 to 2024, with the optimal number of oscillations determined by Generalized Cross-Validation (GCV). Performance was assessed using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) for both in-sample and out-of-sample data. **Results and Discussion.** The model demonstrated high predictive accuracy, with MAPE consistently less than 10% for both in-sample and out-of-sample. It effectively captured both seasonal fluctuations and long-term warming trends in Samarinda's temperature data. **Conclusions.** The STSR-ECFS is a flexible and accurate model for temperature forecasting in tropical rainforest climates, with potential applications in climate adaptation strategies, early warning systems, and other climate variables with similar data characteristics. The model could inform policy decisions on climate adaptation, aiding local governments and environmental agencies in managing risks and formulating mitigation strategies. Its integration into national climate action plans can enhance decision-making for sustainable development and disaster risk reduction.

RESUMEN

Introducción. La predicción de la temperatura en regiones de selva tropical es un desafío debido a la compleja estacionalidad y a las tendencias no lineales que los modelos convencionales suelen no captar adecuadamente. Para enfrentar este problema, se requieren enfoques de modelación avanzados que permitan obtener predicciones climáticas más precisas. **Objetivos.** Este estudio tiene como objetivo construir y aplicar el modelo STSR-ECFS para la predicción de temperatura en Samarinda, Kalimantan Oriental, y evaluar su desempeño predictivo mediante métricas estándar de precisión. **Materiales y Métodos.** El modelo STSR-ECFS combina un componente paramétrico autorregresivo con un componente no paramétrico construido mediante series de Fourier complejas exponenciales. El modelo se entrenó y validó con datos mensuales de temperatura de 2015 a 2024, determinándose el número óptimo de oscilaciones mediante Validación Cruzada Generalizada (GCV). El rendimiento se evaluó mediante el Error Cuadrático Medio (MSE) y el Error Porcentual Absoluto Medio (MAPE) tanto para los datos dentro de la muestra como fuera de ella. **Resultados y Discusión.** El modelo mostró una alta precisión predictiva, con valores de MAPE consistentemente inferiores al 10% tanto en los datos de entrenamiento como en los de validación. Capturó de manera efectiva las fluctuaciones estacionales y las tendencias de calentamiento a largo plazo en los datos de temperatura de Samarinda. **Conclusiones.** El STSR-ECFS es un modelo flexible y preciso para la predicción de temperatura en climas de selva tropical, con potencial de aplicación en estrategias de adaptación climática, sistemas de alerta temprana y otras variables climáticas con características de datos similares. El modelo podría orientar decisiones de política pública sobre adaptación al cambio climático, apoyando a gobiernos locales y agencias ambientales en la gestión de riesgos y en la formulación de estrategias de mitigación. Su integración en planes nacionales de acción climática puede fortalecer la toma de decisiones para el desarrollo sostenible y la reducción del riesgo de desastres.

[1]Doctoral Study Program MIPA, Faculty of Science and Technology, Airlangga University, andreatririandani@fmipa.unmul.ac.id
[2]Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, andreatririandani@fmipa.unmul.ac.id
[3]Department of Mathematics, Faculty of Science and Technology, Airlangga University, nur-c@fst.unair.ac.id
[4]Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, nur-c@fst.unair.ac.id
[5]Department of Statistics, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology, i_nyoman_b@statistika.its.ac.id
[6]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Jember University, lestari.statistician@gmail.com
[7]Department of Biology, Faculty of Mathematics and Natural Sciences, Mulawarman University, ratnakusumafmipa@gmail.com
[8]Department of Statistics, Faculty of Sciences, Mugla Sitki Kocman University, duaydin@mu.edu.tr

INTRODUCTION

Kalimantan's tropical rainforest is crucial in maintaining global climate balance and supporting biodiversity [1]. Temperature is a key factor influencing this ecosystem, affecting biological processes, species distribution, and environmental stability [2-4]. However, recent temperature fluctuations driven by deforestation and climate change pose significant risks, including habitat loss and increased vulnerability to extreme weather events [5]. Accurate temperature analysis and prediction are essential for climate adaptation planning, ecological conservation, and disaster risk reduction [6-8].

Time series regression models like Autoregressive Integrated Moving Average (ARIMA) are widely used in climate data analysis to predict variables such as temperature. The ARIMA model, developed by Box and Jenkins, forms the basis for many forecasting models [9]. Swain et al. (2018) used ARIMA to predict monthly rainfall in Khordha District, India [10]. Narayanan et al. (2013) analyzed pre-monsoon rainfall trends in Western India, showing an increase in several regions [11]. Based on the literature review, it is known that parametric models such as ARIMA have limitations in capturing nonlinear and complex patterns that are often found in temperature data [12-14]. This model assumes a fixed relationship between variables, which does not always correspond to temperature data patterns influenced by various external factors. Therefore, a more flexible approach, such as semiparametric time series regression, is a better choice because it can combine parametric and nonparametric elements in the analysis [15-16].

Semiparametric regression can handle linear and nonlinear relationships, making it more flexible than purely parametric or nonparametric models [17-18] . The parametric component captures linear relationships, while the nonparametric part models complex patterns without assuming a specific functional form [19-20]. This study applies a modified semiparametric partially linear model introduced by Gao for time series data with seasonal characteristics and nonlinear trends [21-22]. The model maintains a linear structure for well-understood relationships while using a smoothing approach for complex patterns. The nonparametric component is estimated using the Fourier series, which effectively captures seasonal patterns in climate data [23-25]. Fourier series are particularly useful for representing oscillatory patterns in temperature data. This approach enables the model to account for linear trends and complex recurring seasonal influences in temperature analysis [26]. In addition, this approach provides flexibility in adjusting the number of Fourier components used based on the complexity of the patterns in the data [26-27]. By utilizing the exponential complex Fourier series (ECFS), this model can handle sharper oscillations and high variability in temperature data, thus providing more accurate and stable estimates than other methods.

Several studies have applied semiparametric regression to time series data. Fibriyani et al. (2024) used a local polynomial estimator to model COVID-19 CFR in Pasuruan [28]. Fitriyah et al. (2025) introduced using a Least Square Spline Estimator to model fluctuating rice production in Indonesia [29]. These models combine parametric and nonparametric approaches to handle linear and nonlinear patterns. However, no studies have specifically combined the Exponential Complex Fourier Series (ECFS) within a semiparametric framework for temperature prediction in tropical rainforest areas. This model can simultaneously capture sharp seasonal fluctuations and long-term warming trends, delivering higher prediction accuracy compared to ARIMA and conventional semiparametric models. To address these emerging challenges, this study develops a Semiparametric Time Series Regression model that employs the Exponential Complex Fourier Series (STSR-ECFS) to predict temperature variations in the tropical rainforest region of Samarinda.

The main contribution of this research is the first application of the STSR-ECFS model to tropical temperature data, which has been shown to provide highly accurate predictions while flexibly capturing both seasonal patterns and long-term trends. The key advantage of the Exponential Complex Fourier Series lies in its ability to capture higher-frequency components, allowing the model to effectively identify sharp temperature fluctuations and subtle climate changes, something that is challenging for other temperature prediction models, such as ARIMA. Unlike conventional semiparametric models that rely on rigid linear assumptions, this approach can handle the irregularities in temperature data more effectively, providing a deeper understanding of tropical temperature dynamics. This modeling approach is expected not only to deliver a more accurate representation of temperature dynamics but also to contribute significantly to improved climate monitoring, data-driven environmental policymaking, and more effective disaster risk management strategies.

MATERIALS AND METHODS

Regression Analysis

Mathematically, the general model of regression analysis is written in Equation (1).

$$y_i=f(x_i)+\varepsilon_i \quad (1)$$

where as the response variable, is the regression curve to be estimated, and is assumed to be IIDN). Regression analysis based on how we approach the shape of the curve is divided into three parts: parametric regression, nonparametric regression, and semiparametric regression [30, 31].

Fourier Series Estimator

Fourier series is a trigonometric polynomial function with a high flexibility level. The Fourier series estimator contains Sine and Cosine functions [23]. Based on the perspective of regression analysis, the Fourier series estimator can be used to estimate functions or curves from data whose patterns are unknown, especially for data that tend to have recurring patterns (seasonal) [26, 32]. This seasonal pattern usually occurs in temperature time series. Suppose given paired data (V_t,y_t) that follow the general model of regression analysis.

$$y_t=m(V_t)+\varepsilon_t \quad ,t=1,2,\ldots,T \quad (2)$$

In Equation (2), y_t denotes the response being modeled at time t, while V_t represents the vector of explanatory variables (regressors) at time t. The regression function $m(V_t)$ is of unknown form and will be estimated using the Fourier series estimator using a nonparametric regression approach. Suppose it is assumed that the regression function $m(Vt)$ is contained in the Hilbert space L_2 [a,b], namely $m(Vt) \in L_2$ [a.b] so that $m(Vt)$ can be expressed as a linear combination of the basis elements of L_2 [a,b]. The Fourier series function with one predictor variable is given in Equation (3).

$$m(V_t) = \alpha_0 + \gamma V_t + \sum_{k=1}^{\lambda}\left[\alpha_k \cos\left(\frac{2\pi k(V_t - 1)}{n}\right) + \vartheta_k \sin\left(\frac{2\pi k(V_t - 1)}{n}\right)\right] \tag{3}$$

where $\hat{\alpha}_0 = \frac{1}{n}\sum_{t=1}^{T} y_t;\ \hat{\alpha}_k = \frac{2}{n}\sum_{k=1}^{\lambda} y_t \cos\left(\frac{2\pi k(V_t-1)}{n}\right); \hat{\vartheta}_k = \frac{2}{n}\sum_{k=1}^{\lambda} y_t \sin\left(\frac{2\pi k(V_t-1)}{n}\right)$.

Data, Data Source, and Research Variables

The data utilized in this study are secondary data obtained from the NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), encompassing temperature records in Samarinda from January 2015 to December 2024. The specific location corresponds to the coordinates of Aji Pangeran Tumenggung Pranoto International Airport: Latitude -0.37361° and Longitude 117.25556°. The variables employed in this study include both response and predictor variables, as detailed in **(Table 1)**.

Table 1. Research Variables

Variable	Notation	Notes	Operational Definition
Response		Temperature	The average temperature recorded in month (, measures in degrees Celsius (°C)
Predictors		Previous observation temperature	The temperature observed in the previous month (, used to capture temporal dependence and trend
		Time	Time index showing the sequence of observations from the first month to month T.

Research Stages

The steps in implementing the semiparametric time series regression model using exponential complex Fourier series (STSR-ECFS) are detailed as follows:

1. Conducting data exploration with a time series graph of monthly temperature data in Samarinda.
2. Divide the data into in sample and - data with a proportion of 90:10.
3. Creating a scatter plot between each predictor variable and the response variable using in-sample data.
4. Identifying the predictor variables of the parametric components and nonparametric components of the Fourier series using in-sample data.
5. Modeling monthly temperature data in Samarinda with semiparametric time series regression using exponential complex Fourier series (STSR-ECFS).
6. Obtaining optimal oscillations using the Generalized Cross-Validation (GCV) method.
7. Calculate the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) for both in-sample and out-of-sample data.
8. Make temperature predictions for 12 periods based on best model STSR-ECFS.

RESULTS

Theoretical Study of the Semiparametric Time Series Regression-Exponential Complex Fourier Series (STSR-ECFS) Model

Suppose we are given a time-series dataset consisting of a response variable . The predictor variables include an autoregressive component , and a time index with following the STSR-ECFS model. Assume that the pattern of relationship between the predictor variables and the response variable follows a semiparametric regression model:

$$y_t=f(y_{(t-1)})+m(V_t)+\varepsilon_t \quad (4)$$

where ε_t is random error $E(\varepsilon_t)=0$ and $Var(\varepsilon_t)=\sigma^2$

The parametric component is defined as , which the Linear estimator will approximate, while the nonparametric component is defined as , which will be approximated by the complex exponential basis Fourier series estimator. The parametric component in Equation (4) is assumed to be known, namely linear, while the nonparametric component is unknown. The first thing in estimating the nonparametric component is to assume the parameters of the parametric component β are known with $y_t^*=y_t-\beta y_{t-1}$, so that:

$$y_t^*=m(V_t)+\varepsilon_t. \quad (5)$$

The nonparametric component in Equation (5) is approximated by the complex exponential basis Fourier series estimator written in Equation (3). If samples are taken at a time, then the complete STSR-ECFS model.

$$y_2 = \beta_0 + \beta_1 y_1 + \frac{1}{2}\alpha_0 + \gamma V_2 + \sum_{k=1}^{\lambda}\left[\alpha_k \cos\left(\frac{2\pi k(V_2-1)}{n}\right) + \vartheta_k \sin\left(\frac{2\pi k(V_2-1)}{n}\right)\right]$$

$$y_3 = \beta_0 + \beta_1 y_2 + \frac{1}{2}\alpha_0 + \gamma V_3 + \sum_{k=1}^{\lambda}\left[\alpha_k \cos\left(\frac{2\pi k(V_3-1)}{n}\right) + \vartheta_k \sin\left(\frac{2\pi k(V_3-1)}{n}\right)\right]$$

$$\vdots$$

$$y_T = \beta_0 + \beta_1 y_{T-1} + \alpha_0 + \gamma V_T + \sum_{k=1}^{\lambda}\left[\alpha_k \cos\left(\frac{2\pi k(V_T-1)}{n}\right) + \vartheta_k \sin\left(\frac{2\pi k(V_T-1)}{n}\right)\right]$$

So it can be written in matrix notation in Equation (6).

$$y=X\beta+W\gamma+\varepsilon \qquad (6)$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \quad \boldsymbol{\gamma} = \begin{bmatrix} \alpha_0 \\ \gamma \\ \alpha_1 \\ \vdots \\ \alpha_\lambda \\ \vartheta_1 \\ \vdots \\ \vartheta_\lambda \end{bmatrix}; \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix}; \quad \boldsymbol{X} = \begin{bmatrix} 1 & y_1 \\ 1 & y_2 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{bmatrix}; \text{ and}$$

$$\boldsymbol{W} = \begin{bmatrix} 1/2 & V_2 & \cos\left(\frac{2\pi(V_2-1)}{n}\right) & \cdots & \cos\left(\frac{2\pi\lambda(V_2-1)}{n}\right) & \sin\left(\frac{2\pi(V_2-1)}{n}\right) & \cdots & \sin\left(\frac{2\pi\lambda(V_2-1)}{n}\right) \\ 1/2 & V_3 & \cos\left(\frac{2\pi(V_3-1)}{n}\right) & \cdots & \cos\left(\frac{2\pi\lambda(V_3-1)}{n}\right) & \sin\left(\frac{2\pi(V_3-1)}{n}\right) & \cdots & \sin\left(\frac{2\pi\lambda(V_3-1)}{n}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1/2 & V_t & \cos\left(\frac{2\pi(V_T-1)}{n}\right) & \cdots & \cos\left(\frac{2\pi\lambda(V_T-1)}{n}\right) & \sin\left(\frac{2\pi(V_T-1)}{n}\right) & \cdots & \sin\left(\frac{2\pi\lambda(V_T-1)}{n}\right) \end{bmatrix}$$

Next, based on Equation (6), each of the two sides is operated with parametric components so that it becomes,

$$y-X\beta=W\gamma+\varepsilon$$
$$y^*=W\gamma+\varepsilon \qquad (7)$$

where $y^*=y-X\beta$

The γ estimator can be obtained using the Least Squares (LS) method. The sum of the squares of the errors is presented in the Equation (8).

$$\varepsilon\hat{}T\ \varepsilon=(y^{*}-W\gamma)\hat{}T\ (y^{*}-W\gamma) \quad (8)$$

The estimated value for the parameter vector is obtained by partially differentiating Equation (8) with respect to and then sets it to zero. The LS estimator for the parameter is written in Equation (9).

$$\gamma\hat{}=(W^{T}W)^{-1}W^{T}y^{*} \quad (9)$$

Based on Equation (5), the form of the Fourier series estimator can be expressed in Equation (10).

$$m\hat{}(V^{t})=W\gamma\hat{}$$

$$m\hat{}(T^{t})=W((W\hat{}T\ W)\hat{}(-1)\ W\hat{}T\ y\hat{}*)$$

$$m\hat{}(T^{t})=Fy\hat{}* \quad (10)$$

If $y^{*}=y-X\beta$, then $y=X\beta+F(y-X\beta)+\varepsilon$. Hence, we obtain error in Equation (11).

$$\varepsilon = y-(X\beta+F(y-X\beta))$$

$$\varepsilon= (I-F)(y-X\beta)$$

$$(11)$$

Based on Equation (11), the form of the sum of squares of errors can be written in Equation (12).

$$\varepsilon^{T}\varepsilon=(y^{T}(I-F)^{T}(I-F)y)-2\beta^{T}X^{T}(I-F)^{T}(I-F)y+(X^{T}\beta^{T}(I-F)^{T}(I-F)X\beta) \quad (12)$$

The value of is obtained by partially differentiating Equation (12) with respect to and then sets it to zero, so that it can be written in Equation (13).

$$\beta\hat{}=(X^{T}(I-F)^{T}(I-F)X)^{-1}X^{T}(I-F)^{T}(I-F)y \quad (13)$$

Next, we substitute Equation (13) into Equation (10), such that we obtain

$$\begin{aligned}\hat{y} &= X\hat{\beta}+F(\hat{y}-X\hat{\beta})\\ &= X\left(\left(X^{T}(I-F)^{T}(I-F)X\right)^{-1}X^{T}(I-F)^{T}(I-F)y\right)\\ &\qquad + F\left(\hat{y}-X\left(\left(X^{T}(I-F)^{T}(I-F)X\right)^{-1}X^{T}(I-F)^{T}(I-F)y\right)\right)\end{aligned}$$

where

$$B_{par}=X\left(\left(X^{T}(I-F)^{T}(I-F)X\right)^{-1}X^{T}(I-F)^{T}(I-F)\right);\text{ and}$$

$$B_{nonpar}=W\left(\left(W^{T}W\right)^{-1}W^{T}\left(I-B_{par}\right)\right).$$

The general form of the prediction in the STSR-ECFS model, which is expressed in Equation (14).

$$\hat{y}=(B_{par}+B_{nonpar})y \quad (14)$$

where $B_{STSR\text{-}ECFs}=B_{par}+B_{nonpar}$

Equation (14) illustrates the general form of prediction in the STSR-ECFS model, where the predicted value is the sum of two components: the parametric component ($\boldsymbol{B}_{par}$ and the nonparametric component $\boldsymbol{B}_{nonpar}$. The parametric component accounts for the linear relationships in the data, while the non-parametric component captures the non-linear relationships, together providing a more comprehensive and flexible prediction model. The method used to determine the optimal number of oscillations in the STSR-ECFS model using the Generalized Cross-Validation (GCV) method. The GCV formula for the STSR-ECFS model is written in Equation (15).

$$GCV(\lambda_{opt}) = \frac{n^{-1}\left(y^T(\boldsymbol{I} - \boldsymbol{B}_{STSR-ECFs})\right)^T\left((\boldsymbol{I} - \boldsymbol{B}_{STSR-ECFs})\boldsymbol{y}\right)}{\left(n^{-1}tr(\boldsymbol{I} - \boldsymbol{B}_{STSR-ECFs})\right)^2} \quad (15)$$

In Equation (15), n denotes the number of observations, and $tr(I\text{-}B_{STSR\text{-}ECFs})$ represents the trace of the smoothing matrix $tr(I\text{-}B_{STSR\text{-}ECFs})$, which reflects the model's level of complexity. The minimum value of GCV will provide the STSR-ECFS model with the optimal number of oscillations. To evaluate the model's prediction accuracy, we compute MSE and MAPE for both in-sample and out-of-sample datasets using Equations (16) and (17).

$$MSE(\hat{\lambda}_{opt}) = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \quad (16)$$

$$MAPE(\hat{\lambda}_{opt}) = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \times 100. \quad (17)$$

where is the estimated value from the STSR-ECFS model using the optimal to generate the prediction.

Application of Temperature Data in Samarinda

Exploration Data

An exploratory analysis of Samarinda's temperature data reveals substantial fluctuations and a clear seasonal pattern over the observed timeframe.

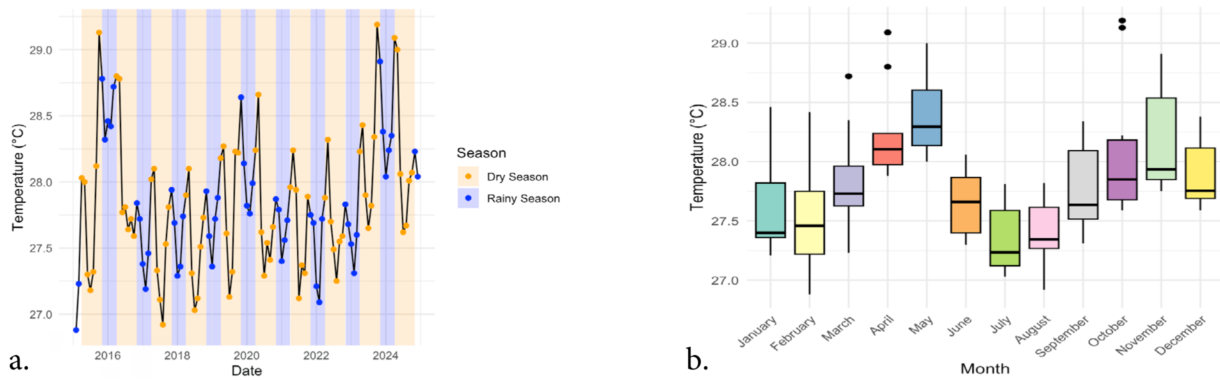

Figure 1. Descriptive Statistics of Temperature in Samarinda. (a) Times series plot, (b) Box plot

Based on monthly temperature data recorded from January 2015 to December 2024 in Samarinda, clearly recurring seasonal pattern is evident **(Figure 1a)**. Temperatures are consistently decreased during the rainy season (November–March) and increased during the dry season (April–October). The alternating colored bands visually reinforce this cyclical trend in the time series plot, distinguishing the two primary climate periods. The boxplot analysis **(Figure 1b)** further substantiates this observation, showing that the median temperatures for dry season months, particularly from July to October, are not only higher but also exhibit greater variability, as reflected by longer whiskers and a higher frequency of outliers.

In addition to these established seasonal cycles, there is clear evidence of a gradual upward trend in average temperatures, especially after 2020. This warming trend is consistent with reports from the Intergovernmental Panel on Climate Change (IPCC, 2021), which indicates that Southeast Asia has experienced an average temperature increase of approximately 0.15–0.2°C per decade since the 1960s. Local factors like urbanization and deforestation, alongside global phenomena like El Niño, likely contribute to the observed rise and the amplification of temperature extremes and variability, particularly in equatorial regions.

A notable finding from the boxplot is the contrast in temperature variability between seasons. The dry season demonstrates a wider interquartile range and frequent outliers, suggesting that temperature extremes are more common outside the rainy months. This increased variability may be attributed to reduce cloud cover, greater solar radiation, and episodic weather events characteristic of the dry season. In contrast, the rainy season is marked by more stable and lower temperatures, likely due to persistent cloud cover, higher humidity, and frequent precipitation acting as natural cooling mechanisms.

Furthermore, outliers and anomalies, especially in February, March, and September, could indicate the influence of extraordinary climatic events or interannual climate variability. Such fluctuations underscore the importance of robust modeling approaches capturing both regular seasonal dynamics and unexpected deviations. This nuanced understanding is critical for developing adaptive early warning systems and effective climate risk mitigation strategies for Samarinda and other tropical regions.

The STSR-ECFs modeling outputs for temperature in Samarinda are illustrated in the scatter plots presented below, including the relationship between the previous temperature observation and the current value, as well as the association between time and temperature

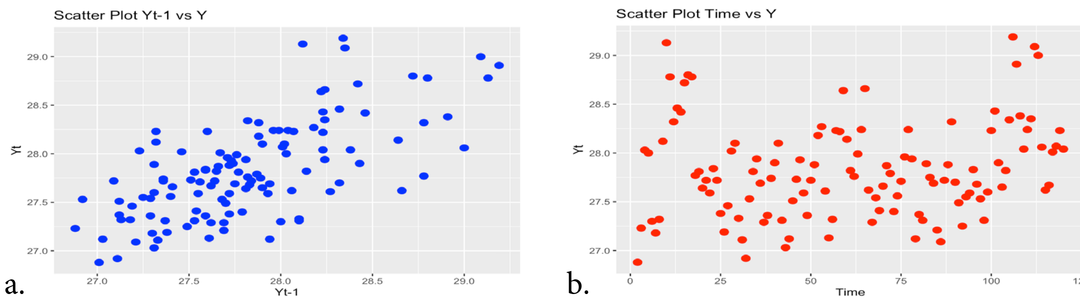

Figure 2. Scatter Plots of (a) Previous observation temperature with temperature and (b) Time with temperature

The scatter plot between previous observation temperature (y_{t-1}) and temperature (y_t) demonstrates a moderately linear relationship, despite some dispersion around the trend. It suggests that the temperature in each month is influenced by the temperature in the preceding month, thereby justifying the potential use of time series models with autoregressive components, such as the Autoregressive (AR) model. These models leverage the dependence structure between past and present values to improve forecasting accuracy. In contrast, the scatter plot between time (V_t) and temperature (y_t) reveals cyclical fluctuations that recur at regular intervals, indicating the presence of a seasonal pattern in the temperature data. Based on (Figure 2), these patterns suggest that both autoregressive and seasonal components are relevant for capturing the underlying structure of the temperature time series in Samarinda.

The STSR-ECFs modeling is carried out by creating R-Code with R software. In this study, the oscillation chosen minimizes the GCV value and considers the parsimony aspect of the model.

The model selection and computational analysis for the STSR-ECFs approach applied to temperature data in Samarinda are summarized in **(Figure 3)**. The GCV values across different oscillation orders (K) are shown in **(Figure 3a)**. The curve indicates that the lowest GCV value of 14.692 is obtained at K = 2, demonstrating the best trade-off between model complexity and generalization. When K is less than 2, the model tends to underfit, failing to capture important seasonal and nonlinear patterns in the data. Conversely, greater values of K (i.e., K > 2) result in increased GCV, suggesting overfitting and reduced generalizability due to excessive model flexibility. This confirms that K = 2 is the optimal choice for accurate temperature prediction using the STSR-ECFs model.

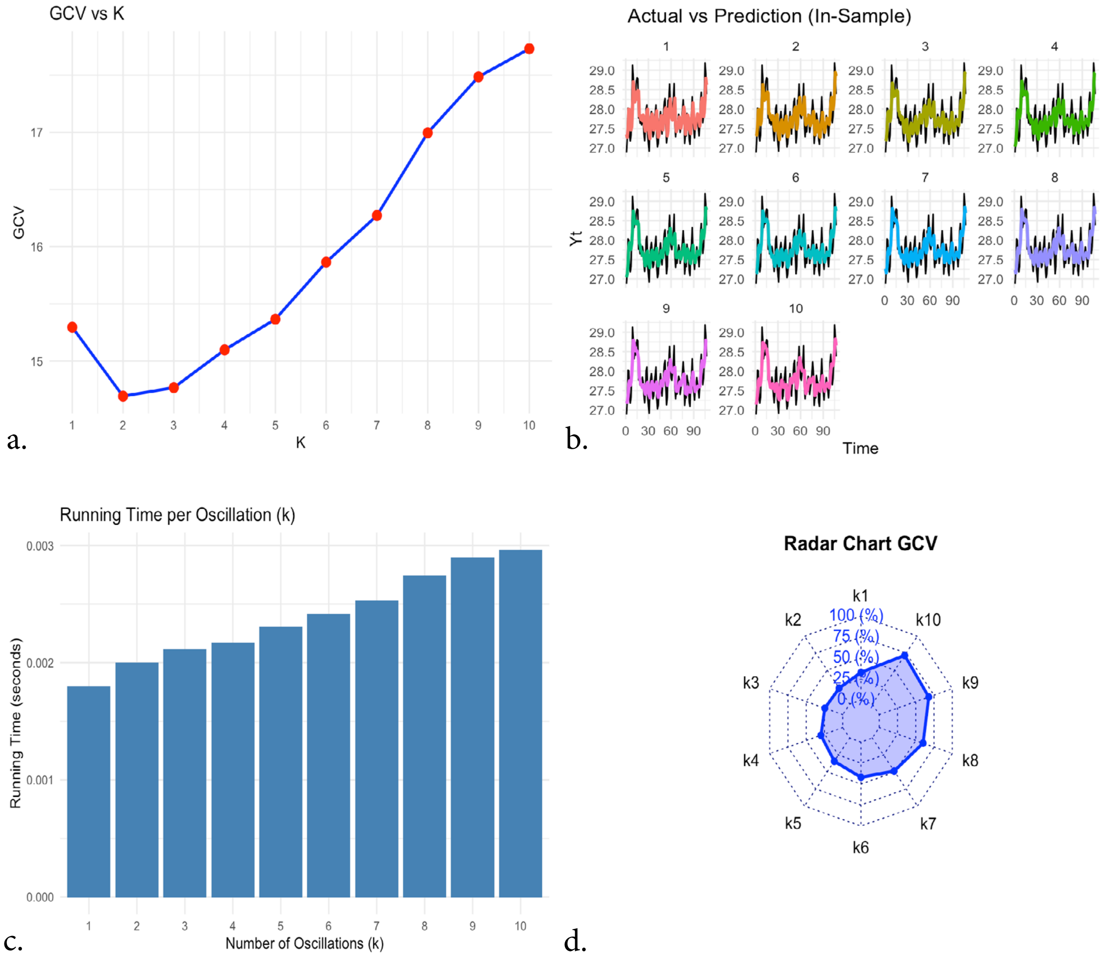

Figure 3. STSR-ECFs modeling results for Samarinda City, including: (a) GCV versus oscillation, (b) actual versus predicted values for in-sample data, (c) running time for each oscillation, and (d) radar chart based on GCV.

The comparison between actual and predicted values (in-sample) for each oscillation order K is presented in **(Figure 3b)**, visualized in a faceted plot. Each subplot illustrates the performance of the STSR-ECFs model for a specific K value. Predictions for K = 2 most closely follow the actual data patterns, further supporting the selection of K = 2 as the optimal model order. For K values much smaller or much larger than 2, the predicted patterns either fail to capture the observed dynamics or become excessively wiggly, indicating underfitting and overfitting, respectively.

The computational running time for each oscillation order is shown in **(Figure 3c)**. The bar chart reveals a clear upward trend, with running time increasing almost linearly as the number of oscillations grows. This aligns with the higher computational burden associated with fitting more complex models that include additional Fourier components. Nevertheless, the overall running times remain practical, supporting the feasibility of implementing the STSR-ECFs approach even when moderate model complexity is required.

The GCV values for each K are visualized in a radar-chart format in **(Figure 3d)** , providing an alternative and intuitive summary of model performance across oscillation orders. The radar chart highlights the minimum

GCV at K = 2 and the relative increases for higher K, visually reinforcing the findings from the line plot. This visualization helps quickly identify the optimal model parameter and communicate the relationship between model complexity and performance to both technical and non-technical audiences.

These analyses underscore the importance of balancing model fit and complexity in semiparametric time-series modeling. Selecting the optimal K based on the minimum GCV ensures that the model can accurately and efficiently capture temperature dynamics in Samarinda, minimizing the risks of overfitting while avoiding unnecessary computational costs. This balanced approach is essential for producing robust, interpretable, and actionable forecasts in real-world climate applications.

Table 2. Accuracy Measure

Data	MSE	MAPE (%)	Interpretation
In-Sample	0.119	1.025	Highly Accurate Prediction
Out of Sample	0.577	2.435	Highly Accurate Prediction
Overall	0.126	1.034	Highly Accurate Prediction

The accuracy results in **(Table 2)** indicate that the STSR-ECFs model exhibits consistently high performance across in-sample and out-of-sample data, with MAPE values are less than 10%. The STSR-ECFs obtained is considered highly accurate. This demonstrates that STSR-ECFs can effectively model the temporal dynamics of temperature data and generalize well to testing data. Such results validate the model's reliability for short-term forecasting and seasonal pattern analysis. The best STSR-ECFs model is written in Equation (18).

$$\hat{y}_t = 25.814 + 5.181 \times 10^{-1} y_{t-1} + 0.009V_t + 0.230\cos\left(\frac{2\pi(V_t-1)}{n}\right) + 0.292\sin\left(\frac{2\pi(V_t-1)}{n}\right) + 0.216\cos\left(\frac{4\pi(V_t-1)}{n}\right) + 0.085\sin\left(\frac{4\pi(V_t-1)}{n}\right) \quad (18)$$

Visualization of the prediction results of the STSR-ECFs model on out-of-sample data is given in **(Figure 4)**.

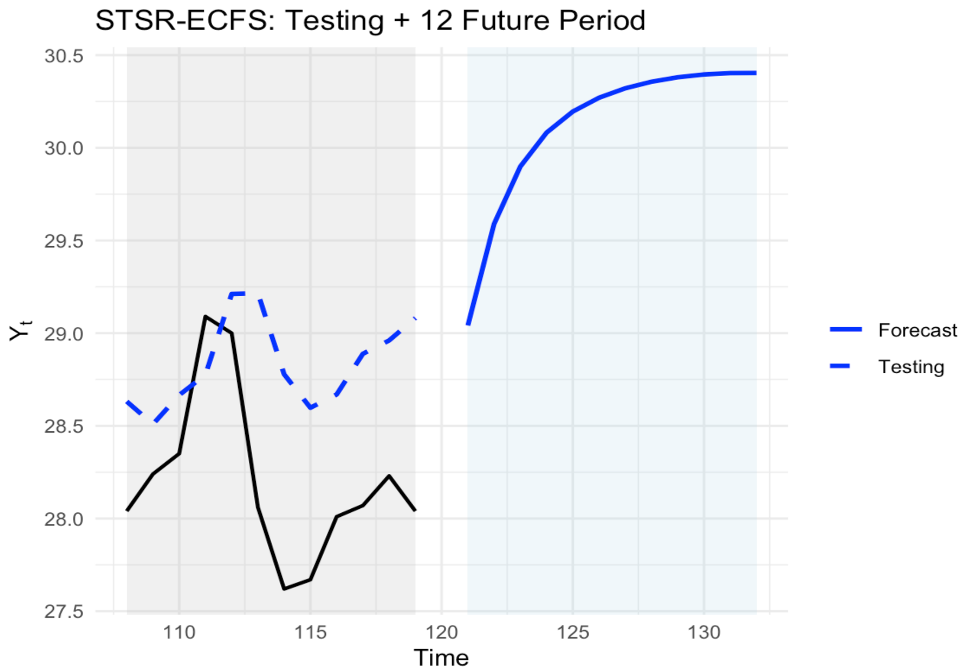

Figure 4. STSR-ECFS for Out of sample Data and Forecast for 12 Future Period

The gray area in the STSR-ECFS plot **(Figure 4)** marks the testing period where the model predictions (dashed blue line) closely follow the observations. Based on the STSR-ECFS projection for the following 12 periods, the average monthly temperature in Samarinda is expected to increase from around 29.2°C in the first period to 29.5°C at the end of the 12th period, with a seasonal peak reaching 30.0–30.2°C in the middle of the year (periods 5–7), decreasing to 29.4–29.6°C during the early rainy season (periods 8–10), then strengthening again to approach 30°C at the horizon cover. This projection serves as an early warning system, emphasizing the urgency of reducing carbon emissions, expanding green open spaces, and implementing fundamental strategies, such as reflective roofs, tree corridors, cooling centers, and low-emission transportation so that Samarinda becomes a city that is friendly and cares about the global climate crisis.

DISCUSSION

This study highlights the effectiveness and flexibility of the Semiparametric Time Series Regression with Exponential Complex Fourier Series (STSR-ECFS) in modeling climate-based data characterized by strong seasonality and nonlinear trends. The STSR-ECFS framework, which combines a parametric autoregressive component, and a nonparametric component based on a complex Fourier series, offers a significant advancement over traditional methods, often limited in capturing the whole dynamics of climate data [17].

This study presents a novel application of the STSR-ECFS model, incorporating explicit Fourier series components within a semiparametric regression framework to temperature data from a tropical rainforest region. Previous research in semiparametric modeling has typically used other nonparametric estimators, such as local linear [28,33-34] or spline methods [29], for the nonparametric component. However, using an exponential complex Fourier-based estimator for the nonparametric component remains unexplored mainly in climate studies for tropical rainforest environments. This methodological innovation allows the model to flexibly capture strong seasonality and nonlinear trends inherent in such data.

Theoretically, the STSR-ECFS model is highly generalizable and can be applied to various domains where time series data exhibit linear (autoregressive) and nonlinear (seasonal/periodic) structures. This opens opportunities for its use in climate studies and fields such as hydrology, finance, epidemiology, and energy demand forecasting. Empirical results based on Samarinda's temperature data demonstrate that the STSR-ECFS model achieves outstanding predictive performance for both in-sample and out-of-sample scenarios. The model produced MAPE which is less than 2% for both in-sample and out-of-sample predictions. This high accuracy indicates that the model can fit historical data effectively while maintaining excellent generalization to some periods ahead.

In summary, the STSR-ECFS model represents an innovative, flexible, and empirically validated approach for climate time series analysis, particularly for regions with complex seasonal and nonlinear characteristics such as tropical rainforests. Its ability to deliver highly accurate predictions in both in-sample and out-of-sample contexts underlining its potential for broader application in environmental monitoring and other fields with similar data characteristics.

CONCLUSIONS

Theoretical studies on the Semiparametric Time Series Regression with Exponential Complex Fourier Series (STSR-ECFS) model have been successfully conducted. This approach effectively captures both seasonal cycles and long-term trends in temperature data with accelerating computation process which is in average less than one second. When we apply the STSR-ECFS to Samarinda's climate data, the model delivered highly accurate predictions and flexibly accommodating the full range of observed fluctuations. It also supported by with MAPE value which is less than 10%. However, the STSR-ECFS has limitation when the pattern has no seasonal patterns.

This research highlights the novelty of integrating an exponential complex Fourier series into a semiparametric regression framework for tropical rainforest climate modeling where this approach has not previously explored in the literature. Looking ahead, enriching the parametric component by incorporating autoregressive terms up to the p-th lag would enable STSR-ECFS to model more intricate serial dependencies, further boosting predictive accuracy for tropical temperature series. In addition, incorporating confidence intervals through simulation would provide a clearer assessment of uncertainty and strengthen the interpretability of the model's estimates. These results validate the STSR-ECFS model as an adaptable framework for climate forecasting and decision-making support in regions facing complex seasonal and nonlinear climate variability. The results of this research support the 13rd Sustainable Development Goals (SDGs) point, namely climate change.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP), under the Ministry of Finance of the Republic of Indonesia, for the financial support provided through the Doctoral Program Scholarship.

FUNDING

This work was funded by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP), Ministry of Finance of the Republic of Indonesia, through the Doctoral Program Scholarship, as specified in Decree No. SKPB-10349/LPDP/LPDP.3/2024.

ETHICAL CONSIDERATIONS

This study did not involve human participants or animals. The research was conducted exclusively using secondary temperature data obtained from official sources for the period 2015–2024. Therefore, informed consent and ethics committee approval were not required. The study complies with ethical principles applicable to research using publicly available secondary data and does not include personal information or identifiable subjects.

DECLARATION OF COMPETING INTEREST

The authors have declared no conflict of interest.

REFERENCES

1. **Lestari F, Sudaryo MK, Djalante R, Adiwibowo A, Kadir A, Zakianis, et al.** Estimating the flood, landslide, and heavy rainfall susceptibility of vaccine transportation after 2021 flooding in South Kalimantan Province, Indonesia. *Sustainability*. 2024;16:1554. https://doi.org/10.3390/su16041554

2. **Niko N.** Dayak Benawan Indigenous Futures: Tropical Rainforest Knowledge in Kalimantan, Indonesia. *ETropic*. 2025;24:218-39. https://doi.org/10.25120/etropic.24.1.2025.4144

3. **Manandhar S, Dev S, Lee YH, Winkler S, Meng YS.** Systematic study of weather variables for rainfall detection. *Int Geosci Remote Sens Symp*. 2018:3027-30. https://doi.org/10.1109/IGARSS.2018.8517667

4. **Baljon M, Sharma SK.** Rainfall prediction rate in Saudi Arabia using improved machine learning techniques. *Water*. 2023;15:826. https://doi.org/10.3390/w15040826

5. **Marjenah, Ramadani, Farahdita WL, Kiswanto.** Species diversity and carbon storage of Tanah Grogot urban forest in Paser Regency of East Kalimantan. *IOP Conf Ser Earth Environ Sci*. 2025;1447:012012. https://doi.org/10.1088/1755-1315/1447/1/012012

6. **Munandar D, Ruchjana BN, Abdullah AS, Pardede HF.** Literature review on GSTARIMA and deep neural networks for climate forecasting. *Mathematics*. 2023;11:2975. https://doi.org/10.3390/math11132975

7. **Sanikhani H, Nikpour MR, Jamshidi F.** Advanced framework for predicting rainfall-runoff. *Water Resour Manag*. 2025;:1-??. https://doi.org/10.1007/s11269-025-04106-9

8. **Sudradjat A, Muntalif BS, Marasabessy N, Mulyadi F, Firdaus MI.** Relationship between chlorophyll-a, rainfall, and climate phenomena in estuarine waters. *Heliyon*. 2024;10:e25812. https://doi.org/10.1016/j.heliyon.2024.e25812

9. **Box GEP, Jenkins GM.** Time series analysis: forecasting and control. *Holden-Day*. 1976.

10. **Swain S, Nandi S, Patel P.** Development of an ARIMA model for monthly rainfall forecasting. *Adv Intell Syst Comput*. 2018;708:325-31. https://doi.org/10.1007/978-981-10-8636-6_34

11. **Narayanan P, Basistha A, Sarkar S, Kamna S.** Trend analysis and ARIMA modelling of pre-monsoon rainfall. *Comptes Rendus Geosci*. 2013;345:22-7. https://doi.org/10.1016/j.crte.2012.12.001

12. **Aydın D, Mammadov M.** A comparative study of hybrid neural networks and nonparametric regression models. *WSEAS Trans Math*. 2009;8:593-603.

13. **Unnikrishnan P, Jothiprakash V.** Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall. *Water Resour Manag*. 2020;34:3609-23. https://doi.org/10.1007/s11269-020-02638-w

14. **Khosravi K, Farooque AA, Bateni SM, Jun C, Dhiman J.** Prediction of rainfall characteristics using hybrid learners. *Results Eng*. 2025;25:103840. https://doi.org/10.1016/j.rineng.2024.103840

15. **Ahmed SE, Aydin D, Yilmaz E.** Semiparametric time-series model using local polynomial. *J Risk Financ Manag*. 2022;15:0141. https://doi.org/10.3390/jrfm15030141

16. **Gao J.** Semiparametric regression smoothing of non-linear time series. *Scand J Stat*. 1998;25:521-39. https://doi.org/10.1111/1467-9469.00118

17. Gao J, Hawthorne K. Semiparametric estimation and testing of temperature trends. *Econom J*. 2006;9:332-55. https://doi.org/10.1111/j.1368-423X.2006.00188.x

18. Roozbeh M, Arashi M. New ridge regression estimator in semiparametric models. *Commun Stat Simul Comput*. 2016;45:3683-715. https://doi.org/10.1080/03610918.2014.953685

19. Harezlak J, Ruppert D, Wand MP. Semiparametric regression with R. *Springer*. 2018. https://doi.org/10.1017/cbo9781139058414.011

20. Lin DY, Ying Z. Semiparametric regression analysis of longitudinal data. *J Am Stat Assoc*. 2001;96:103-13. https://doi.org/10.1198/016214501750333018

21. Gao J. Nonlinear time series: semiparametric and nonparametric methods. *Chapman & Hall*. 2007.

22. Gao J, Phillips PCB. Semiparametric estimation in triangular system equations. *J Econom*. 2013;176:59-79. https://doi.org/10.1016/j.jeconom.2013.04.018

23. Bilodeau M. Fourier smoother and additive models. *Can J Stat*. 1992;20:257-69. https://doi.org/10.2307/3315313

24. Ming WY, Huang LJ. Fourier series neural networks for regression. *IEEE Int Conf Appl Syst Innov*. 2018:716-9. https://doi.org/10.1109/ICASI.2018.8394358

25. Mariati NPAM, Budiantara IN, Ratnasari V. Combination estimation of smoothing spline and Fourier series. *J Math*. 2020;2020:4712531. https://doi.org/10.1155/2020/4712531

26. Chamidah N, Febriana SD, Ariyanto RA, Sahawaly R. Fourier series estimator for predicting international market price of white sugar. *AIP Conf Proc*. 2021;2329:1-?. https://doi.org/10.1063/5.0042287

27. Chamidah N, Lestari B, Budiantara IN, Aydin D.. Estimation of multiresponse multipredictor nonparametric regression model. *Symmetry*. 2024;16:0386. https://doi.org/10.3390/sym16040386

28. Fibriyani V, Chamidah N, Saifudin T. Estimating semiparametric regression model for inflation. *J King Saud Univ Sci*. 2024;36:103549. https://doi.org/10.1016/j.jksus.2024.103549

29. Fitriyah AT, Chamidah N, Saifudin T. Prediction of paddy production using semiparametric regression. *Data Metadata*. 2025;4:527. https://doi.org/10.56294/dm2025527

30. Ratnasari V, Budiantara IN, Dani ATR. Nonparametric regression mixed estimators of truncated spline and Gaussian kernel. *Int J Adv Sci Eng Inf Technol*. 2021;11:2400-6. https://doi.org/10.18517/ijaseit.11.6.14464

31. Dani ATR, Ratnasari V, Budiantara IN. Optimal knots point and bandwidth selection in modeling mixed estimator nonparametric regression. *IOP Conf Mater Sci Eng*. 2021;1115:012020. https://doi.org/10.1088/1757-899X/1115/1/012020

32. Amri IF, Chamidah N, Saifudin T, Purwanto D, Fadlurohman A, Fitriyana Ningrum A, et al. Prediction of extreme weather using nonparametric regression approach with Fourier series estimators. *Data Metadata*. 2024;4:319. https://doi.org/10.56294/dm2024319

33. Fibriyani V, Chamidah N. Prediction of inflation rate in Indonesia using local polynomial estimator. *J Phys Conf Ser*. 2021;1776:012065. https://doi.org/10.1088/1742-6596/1776/1/012065

34. Fibriyani V, Chamidah N, Saifudin T. Modeling case fatality rate of COVID-19 in Indonesia using semiparametric regression. *Commun Math Biol Neurosci*. 2024;2024:1-22. https://doi.org/10.28919/cmbn/8379