BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1459–1470
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1459-1470

MODELING DHF IN CENTRAL JAVA USING HYBRID NONPARAMETRIC SPLINE TRUNCATED-FOURIER SERIES APPROACH

Tiani Wahyu Utami[1*], Endang Tri Wahyuni Maharani[2], Alwan Fadlurohman[3]

[1]Department of Statistics, Universitas Muhammadiyah Semarang
[2]Department of Chemistry Education, Universitas Muhammadiyah Semarang
[3]Department of Data Sains, Universitas Muhammadiyah Semarang
Jl. Kedungmundu No.18, Kota Semarang, Jawa Tengah 50273, Indonesia

*Corresponding author's e-mail: * **tianiutami@unimus.ac.id***

Article History:

Received: 2nd November 2023
Revised: 2nd January 2024
Accepted: 12th June 2024
Published: 1st September 2024

Keywords:

DHF;
Fourier Series;
GCV;
Nonparametric Regression;
Spline Truncated.

ABSTRACT

Regression analysis aims to determine the relationship and influence of predictor variables on response variables through regression curve. The problem with nonparametric regression research so far is that it only uses one approach, causing the estimation results to be biased, even though each data sub-pattern has its own suitability depending on the approach method used. Therefore, the hybrid method emerged as a development of nonparametric regression. Hybrid models are models that combine approach methods, with the hope of increasing accuracy in modeling analysis. This research was carried out using two non-parametric approaches, namely Spline Truncated and Fourier Series. Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus. DHF is endemic and occurs throughout the year, especially during the rainy season because mosquitoes reproduce optimally. The aim of this research is to estimate the Hybrid Nonparametric Spline Truncated -Fourier Series model and apply the estimation results to data on DHF cases in Central Java. The data used to apply the hybrid nonparametric Spline Truncated-Fourier series regression model is DHF in the city/districts of Central Java. Estimation smoothing parameters uses the GCV (Generalized Cross Validation) method. The best model is selected based on largest R-Square and the smallest MSE. Modeling the disease of DHF cases in Central Java using the Spline Truncated-Fourier Series hybrid estimator produced the best model from the Spline Truncated model with two knot points for each predictor and the Fourier Series model with K value of 9. Based on the results obtained, it can be compared that the Truncated Spline-Fourier Series hybrid model is better than the Spline Truncated model, this can be seen from the largest R-square, namely 99.94% and the smallest MSE.

How to cite this article:

T. W. Utami, E. T. W. Maharani and A. Fadlurohman, "MODELING DHF IN CENTRAL JAVA USING HYBRID NONPARAMETRIC SPLINE TRUNCATED-FOURIER SERIES APPROACH," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1459-1470, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is an endemic disease transmitted by the Aedes aegypti and Aedes albopictus mosquitoes. This disease has experienced an increase in cases of up to 30 times throughout the world. Dengue Hemorrhagic Fever (DHF) is often found in tropical and sub-tropical areas. Since 1968, the World Health Organization (WHO) has recorded Indonesia as the country with the highest dengue fever cases in Southeast Asia [1]. DHF Diseases in 2022 have increased again compared to the previous year. In 2022, the number of DHF Diseases has increased again to 131,265 cases with a death rate of 1,135 people. The province in Indonesia that will be number two with the highest number of DHF is Central Java in 2022 [2]. The number diseases of Dengue Hemorrhagic Fever (DHF) in Central Java are currently increasing. Throughout 2022, from January to June, there were 6,699 cases of dengue fever recorded, with an incidence rate of around 19.9 percent. Several districts/cities in Central Java experienced a significant increase in dengue fever cases. The people of Central Java are advised to be aware of DHF because there is still high rainfall in various areas where standing water appears, so the potential for the spread of DHF will also be greater [3].

Regression analysis is made to ensure the relationship and impact of the predictor variables on the respondent variables. We can create nonparametric regression models, which is flexible statistical approach without any assumptions. Nonparametric regression modeling experiences development every period. Nonparametric regression is used to examine data trends and estimate future values using historical data [4]. The problem with Nonparametric Regression research so far is that it only uses one approach, causing the estimation results to be biased, even though each data sub-pattern has its own suitability depending on the approach method used. Therefore, the Hybrid method emerged as a development of Nonparametric Regression. Hybrid models are models that combine approach methods, with the hope of increasing accuracy in modeling analysis. Two methods were used in this research nonparametric regression approach: Spline Truncated and Fourier Series.

The aim of this research is to estimate the Hybrid Nonparametric Spline Truncated-Fourier Series model and apply the estimation results to data on dengue fever cases in Central Java. Determine parameter estimates using the GCV (Generalized Cross Validation) method. By estimating the parameters, one can derive the regression curve estimator [4]. The GCV method is frequently utilized. Because it is invariant to transformations, has excellent asymptotic properties, and does not require knowledge of the population variance [5]. The GCV outperforms the UBR method [6]. Selected model based on largest R-Square and the smallest MSE. The urgency of this research is that there are no estimation results using the hybrid nonparametric Truncated Spline-Fourier Series to modeling DHF in Central Java. The results of this modeling are able to describe better than other modeling. The modeling results aim to reduce the increase in DHF sufferers in districts/cities in Central Java. The application of nonparametric models requires the help of statistical software. One of the opensource data analysis packages is R software which is based on the S programming language [7]. Several previous studies related to nonparametric regression using Fourier Series and Spline estimators include modeling the Courage of field extension officers using Spline [8], modeling of the tidal using Fourier Series approach [9], modeling of human development index (HDI) using Spline Truncated [10], while research on selecting optimum knots in spline regression [11]. Previous research regarding the number of DHF in Central Java includes DHF modeling using the spatial Durbin model method [12] and DHF modeling using the Geographically Weighted Negative Binomial Regression (GWNBR) method [13].

The nonparametric regression strategy is utilized to establish a pattern of relationship between the independent variable (x) and the dependent variable (y) in situations where the regression curve's shape is unknown [14]. Trigonometric polynomial function with certain degree of flexibility is the Fourier series. This is thus because the sine and cosine functions are represented as a curve in the Fourier series [15]. The degree of smoothness of the Fourier series estimator is determined by the choice of smoothing parameter K. The smaller the smoothing parameter K, the smoother the estimate and the larger the smoothing parameter K, the less smooth the estimate of f. Therefore, it is necessary to choose the optimum K [16]. Spline regression is a polynomial piece function that has segmented and continuous properties. Spline Truncated Regression allows for various orders so that linear, quadratic, cubic or m order Spline regression can be formed [17].

2 RESEARCH METHODS

2.1 Subsection of Material

Dengue Hemorrhagic Fever (DHF) Diseases in 2022 have increased again compared to the previous year. In 2022, the number of DHF Diseases has increased again to 131,265 cases with a death rate of 1,135 people. The province in Indonesia that will be number two with the highest number of DHF is Central Java in 2022. The following is a graph of DHF diseases in Indonesia:

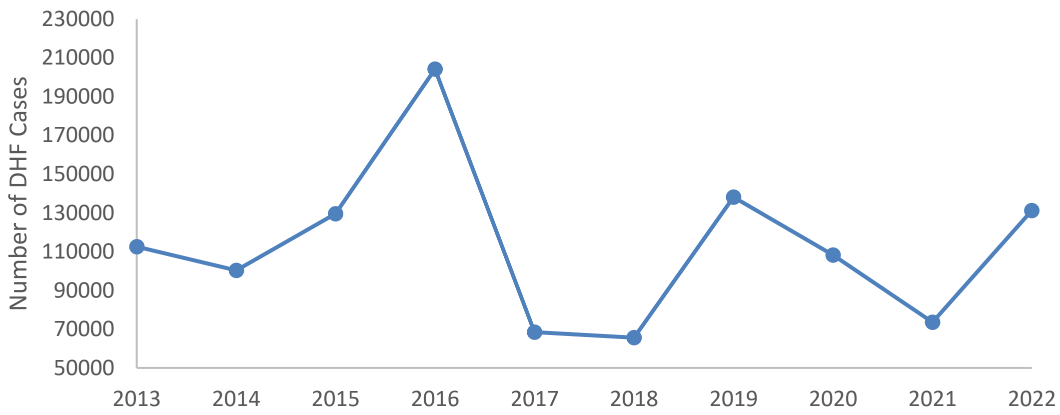

Figure 1. Graph of DHF Disease in Indonesia for the 2013-2022 period

This research uses secondary data on DHF in 2022 obtained from the Central Java Provincial Health Service. The data includes 4 variables for DHF in 2022 with the research objects observed being districts or cities in Central Java which consists of 29 districts and 6 cities. The research uses 4 variables with DHF indicators based on the Central Java Provincial Health Service. The variables used in this research are the number of dengue sufferers (y), the Indicator Rate (IR) ($X1$), the percentage of Clean and Healthy Living Behavior ($X2$), Mean Years of Schooling ($X3$). Variables used in research:

a. Number of DHF sufferers: Dengue hemorrhagic fever (DHF) is a disease caused by the dengue virus and transmitted through the bite of the Aedes Aegypti or Aedes albopictus mosquito.
b. Indicate Rate (IR): the Indicate Rate (IR) is the number of new sufferers of a disease found in a certain period of time (generally 1 year) compared to the number of residents who may be affected by the new disease in the middle of the period concerned [18].
c. Percentage of Clean and Healthy Living Behavior (PCHB) is a house that meets the minimum criteria: access to drinking water, access to a healthy toilet, flooring, ventilation and lighting calculated cumulatively from the previous year [19].
d. Mean Years of Schooling (MYS) is defined as the number of years of study for people aged 15 years and over who have completed formal education (excluding repeat years). The population coverage calculated by MYS is the population aged 25 years and over with the assumption that at the age of 25 years the education process has ended.

2.2 Method

The main problem that will be solved in this research is the development and application of hybrid model on data on DHF in Central Java and the factors that influence it as well as the development of an opensource program using *R* Software. This research uses two approach methods, namely Spline and Fourier Series. The following includes several research stages. Hybrid nonparametric modeling using Spline Truncated- Fourier Series on data of DHF in Central Java

1. Estimation of the Spline Truncated-Fourier Series Nonparametric Regression Hybrid Model. Finding the relationship between the independent and dependent variables are done using regression analysis. The regression equation model is as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \tag{1}$$

where the independent variable is x and the dependent variable is y_i.

2. Create a program to determine optimum knots using the GCV method in the Spline Truncated model. The general form of the Spline Truncated model is presented in the equation:
$$s(t) = \sum_{i=0}^{p} a_i x^i + \sum_{j=0}^{k} \beta_j (x - knot)_+^p \tag{2}$$
where p is the degree of the polynomial and k is the number of knot points in the truncated function, and ε is an independent random error with mean zero and variance σ^2.
3. Use optimum knots to determine the Spline Truncated regression parameter estimation
4. Calculate the residual value of the Spline Truncated regression model
5. Estimate the Fourier Series with the response variable being the residual value of the Spline Truncated regression model. The following is the Fourier Series function:
$$f\,(t) = \frac{1}{2}\,a_0 + \gamma t + \sum_{k=1}^{K} a_k \cos\left(\frac{2\pi kt}{2L}\right) \tag{3}$$
6. Create a program to determine the optimum K in the Fourier Series model
7. Use K optimum to determine parameter estimates for the Fourier Series regression model
8. Calculate the MSE and R-square of the Hybrid model. Selection of the best model based on largest R-square and smallest MSE value.

3. RESULTS AND DISCUSSION

3.1 Estimation of Truncated Spline-Fourier Series Nonparametric Regression Hybrid Model

Given data on the response variable (y_i) and predictor variables (x_i) consisting of a number of m predictor variables so that if this is modeled using a regression model as follows [20]:
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + +\beta_v x_{vi} + \cdots + +\beta_m x_{mi} + \varepsilon_i \tag{4}$$
The hybrid approach is carried out using Spline Truncated and Fourier Series estimators. The hybrid approach is a combination of two estimators with different functional forms. This modeling consists of a combination based on predictors and residuals, meaning that the predictor variables are approached using a Spline Truncated while the residuals are approached using Fourier Series [20]. The equation can be written as follows:
$$y_i = S(x_i) + \varepsilon_i, i = 1,2, \dots, n \tag{5}$$
Where the function $S(x_i)$ is approached using Spline Truncated, while ε_i is approached using a Fourier series, so if approached using Spline Truncated then Equation (5) becomes:
$$y = \beta_0 + \sum_{j=1}^{m} \beta_j x^j + \sum_{k=1}^{n} \beta_{j+k} (x - \mathrm{Knot}_k)_+^m + \varepsilon \tag{6}$$
This equation is elaboration of a Spline Truncated for each predictor so that if the predictor is approached using a Spline, each predictor will have its own knot point value and polynomial order.

The residual value is approached using the Fourier Series estimator, while the residual value is obtained from Nonparametric Spline Truncated modeling, so that in Equation (5) there should be two functions as follows: $y_i = s(x_i) + \varepsilon_i$ and $\varepsilon_i \approx f(x_i)$ so that$y_i = s(x_i) + f(x_i) + e_i$, with the function $f(x_i)$ being the residual value of Spline Truncated estimation modeling results. The curve $f(x_i)$ will be approximated using the Fourier Series and e_i is the residual value from the combined or hybrid model. The MSE value calculation is obtained from the MSE value of a combined or hybrid model between Spline Truncated and Fourier Series [20].
$$\varepsilon_i \approx f\,(x_{ij}) = \frac{1}{2}\,a_0 + \gamma x_{ij} + \sum_{k=1}^{K} a_{jk} \cos\left(\frac{2\pi k x_{ij}}{2L}\right) \tag{7}$$

$$y_i = s(x_i) + \frac{1}{2} a_0 + \gamma x_{ij} + \sum_{k=1}^{K} a_{jk} \cos\left(\frac{2\pi k x_{ij}}{2L}\right) + e_i \tag{8}$$

3.2. Modeling DHF Cases in Central Java City/Districts using a hybrid Truncated Spline-Fourier Series regression approach

Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus. DHF is transmitted through the bite of *Aedes* genus mosquitoes, especially *Aedes Aegypti*. In tropical and subtropical areas, DHF is endemic and occurs throughout the year, especially during the rainy season because mosquitoes reproduce optimally. The data used to apply the hybrid nonparametric kernel Spline Truncated-Fourier series regression model is DHF in the City/Districts of Central Java. Before obtaining non-parametric hybrid regression modeling using the Spline Truncated - Fourier Series approach, create a scatterplot of the number of DHF sufferers in the City/Districts of Central Java. The following is scatterplot of DHF in Central Java City/Districts in 2022:

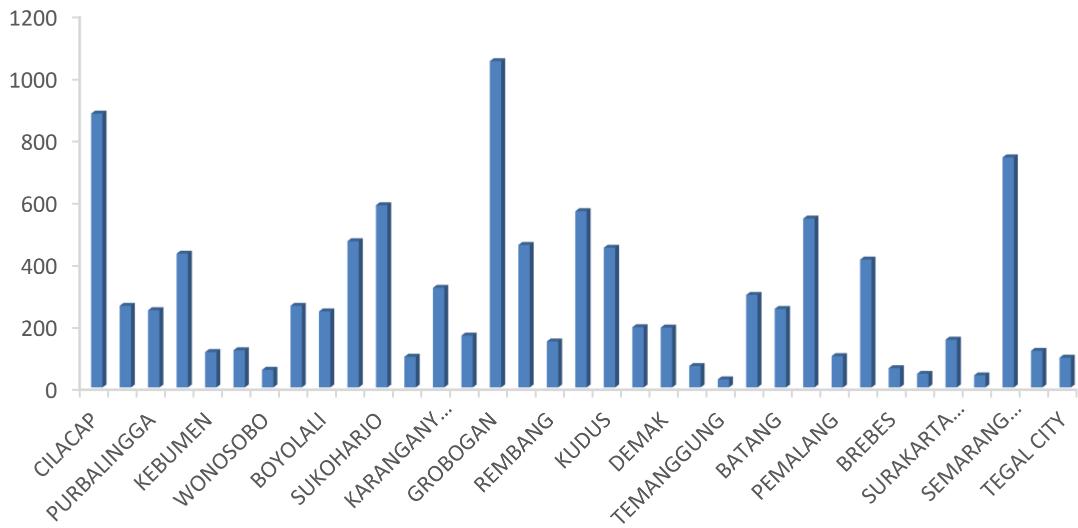

Figure 2. Graph of DHF in Central Java City/Districts 2022

The highest number of DHF in Central Java in 2022 occurred in Grobogan Districts with 1,051 DHF sufferers. In this research, the response variable is DHF in the Central Java Province. Factors causing high levels of Dengue Hemorrhagic Fever include are: population density, lack of clean and healthy living behavior, low public knowledge and education, late information from hospitals, and insufficient health workers [21]. Meanwhile, predictor variables in this research are the Indicate Rate (IR), Percentage of Clean and Healthy Living Behavior, Mean Years of Schooling. All predictor variables are assumed not to form a particular distribution pattern. DHF in Central Java Province for each predictor variable produces an unformed data pattern, therefore the approach taken is nonparametric, namely using Spline Truncated-Fourier Series approach. Data sources used in research comes from the 2022 Central Java Health Service consisting of 35 cities/districts in Central Java Province.

a) Modeling the Number of DHF Cases Using the Spline Truncated Approach

Knot are points of change in data behavior at certain sub-intervals. Determining the best knot point using the GCV method. The best knot is obtained with the lowest GCV value. The following is the determination of optimum knot points using third order at each knot point. The nonparametric spline regression model with one knot of DHF sufferers in Central Java is estimated as follows:

$$\hat{S}(x) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 {X_1}^2 + \hat{\beta}_3 (X_1 - K_1)_+^2 + \hat{\beta}_4 X_2 + \hat{\beta}_5 {X_2}^2 + \hat{\beta}_6 (X_2 - K_2)_+^2 + \hat{\beta}_7 X_3 + \hat{\beta}_8 {X_3}^2 + \hat{\beta}_9 (X_3 - K_3)_+^2$$

The following **Table 1** shows several GCV values that are around the minimum GCV using one knot point:

Table 1. GCV Values for One Knot Point

GCV	$X1$	$X2$	$X3$
30025	35	90	10
28719	26	83	10
27493	55	95	10
25184	**55**	**94**	**10,5**

Based on **Table 1**, the minimum GCV for the Spline Truncated model is 25184. Based on the minimum GCV results, all predictors have their own knot points. The Indicate Rate (IR) variable ($X1$) has optimum knots of 55, the Percentage of Clean and Healthy Living Behavior variable ($X2$) has an optimum knot of 94, while the Average Years of Schooling variable ($X3$) has an optimum knot of 10.5 The next step is to determine the two optimum knots for each variable. The following is Spline Truncated model of the DHF in Central Java with two knot points.

$$\hat{S}(x) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 {X_1}^2 + \hat{\beta}_3 {X_1}^3 + \hat{\beta}_4 (X_1 - K_1)_+^3 + \hat{\beta}_5 (X_1 - K_2)_+^3 + \hat{\beta}_6 X_2 + \hat{\beta}_7 {X_2}^2 + \hat{\beta}_8 {X_2}^3 + \hat{\beta}_9 (X_2 - K_3)_+^3 + \hat{\beta}_{10} (X_2 - K_4)_+^3 + \hat{\beta}_{11} X_3 + \hat{\beta}_{12} {X_3}^2 + \hat{\beta}_{13} {X_1}^3 + \hat{\beta}_{14} (X_3 - K_5)_+^3 + \hat{\beta}_{15} (X_3 - K_6)_+^3$$

The following **Table 2** shows four GCV values that are around the minimum GCV using two knot points:

Table 2. GCV for Two Knot Point

GCV	$X1$		$X2$		$X3$	
	Knot1	Knot2	Knot1	Knot2	Knot1	Knot2
26134	50	55	80	91	8	10
26058	51	55	83	94	8,4	10
25160	36	45	85	95	10	10,5
24377	**46**	**53**	**90**	**93**	**9**	**10,4**

Based on **Table 2**, the minimum GCV for the Spline model is 24377. Based on the minimum GCV results, all predictors have their own knot points. The Indicate Rate (IR) variable ($X1$) has two optimum knots, including 46 and 53, the variable Percentage of Clean and Healthy Living Behavior ($X2$) has two optimum knot points, namely 90 and 93, while the variable Average Years of Schooling ($X3$) consists of three optimum knot points including 9 and 10.4.

The next step is to determine the three optimum knots for each variable. The Spline Truncated model with three knot points is as follows:

$$\hat{S}(x) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 {X_1}^2 + \hat{\beta}_3 {X_1}^3 + \hat{\beta}_4 (X_1 - K_1)_+^3 + \hat{\beta}_5 (X_1 - K_2)_+^3 + \hat{\beta}_6 (X_1 - K_3)_+^3 + \hat{\beta}_7 X_2 + \hat{\beta}_8 {X_2}^2 + \hat{\beta}_9 {X_2}^3 + \hat{\beta}_{10} (X_2 - K_4)_+^3 + \hat{\beta}_{11} (X_2 - K_5)_+^3 + \hat{\beta}_{12} (X_2 - K_6)_+^3 + \hat{\beta}_{13} X_3 + \hat{\beta}_{14} {X_3}^2 + \hat{\beta}_{15} {X_3}^3 + \hat{\beta}_{16} (X_3 - K_7)_+^3 + \hat{\beta}_{17} (X_3 - K_8)_+^3 + \hat{\beta}_{18} (X_3 - K_9)_+^3$$

The following **Table 3** shows the GCV values that are around the minimum GCV using three knot points:

Table 3. GCV Values for Three Knot Point

GCV	$X1$			$X2$			$X3$		
	Knot1	Knot2	Knot3	Knot1	Knot2	Knot3	Knot1	Knot2	Knot3
31338	36	43	50	85	90	96	8,6	9,3	10,2
29528	36	45	53	88	93	94	8,8	9	10,4
28057	44	48	54	86	92	95	8,5	9,5	10,5
27806	**46**	**48**	**54**	**85**	**94**	**95**	**8,5**	**9,5**	**10,5**

The selection of the best Spline model is based on knot points. Choosing the optimum knot point is very important in determining the best model. Different knot point locations will produce different models. The optimum knot point is seen based on the minimum GCV criteria. The following is a table that shows the optimum knot points and minimum optimum GCV for each variable.

Table 4. Comparison of GCV

Number of Knot Points	GCV
1	25184
2	**24377**
3	27806

The best knot in Spline Truncated model is determined from smallest GCV. Table 4 presents the values of one knot, two knots, and three knots which will be compared based on the minimum GCV value. From this table we know that the minimum GCV for two knots is 24377 with an MSE of 3742.56. Therefore, the best spline model for the DHF in Central Java uses two knots. The non-parametric spline regression model equation with two knots is presented below:

$$\hat{S}(x) = -161.98 - 0.475X_1 + 0.481{X_1}^2 - 0.006{X_1}^3 - 3003.78(X_1 - 46)_+^3 + 56.364(X_1 - 53)_+^3 + 0.265X_2 + 10449.56{X_2}^2 - 1303.596{X_2}^3 + 53.496(X_2 - 90)_+^3 + 0.041(X_2 - 93)_+^3 + \hat{\beta}_{11}X_3 - (1.39 \times 10^{-13}){X_3}^2 - 0.828(X_3 - 9)_+^3 - 6418.706(X_3 - 10.4)_+^3$$

Therefore, it can be concluded that the results of modeling DHF in Central Java in each city/district using the Spline Truncated estimator obtained an MSE value of 3742.56 and a coefficient of determination of 82.56%. Based on the Spline Truncated model, the residuals are very different, so the scatterplot results of the residuals from the Spline Truncated nonparametric model are as follows:

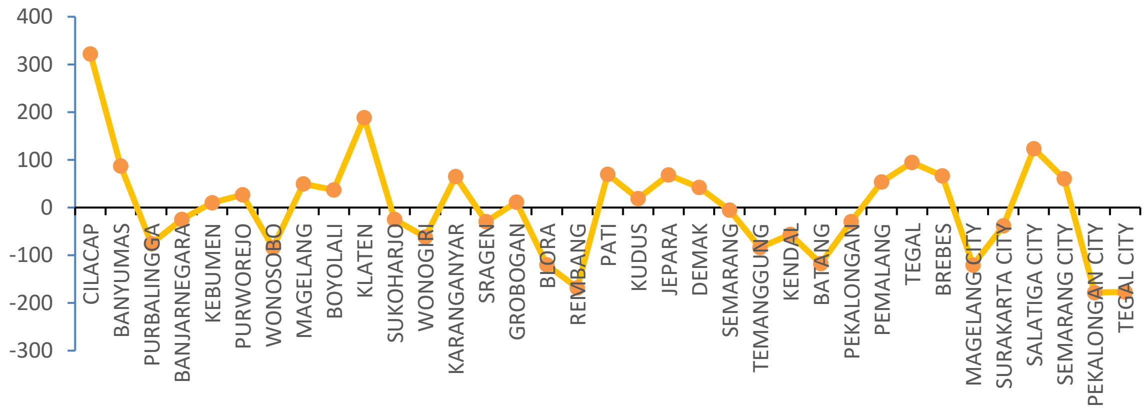

Figure 3. Scatterplot of The Residuals of Spline Truncated Nonparametric Model

b) Estimating Fourier Series Nonparametric Hybrid Regression Model Based on Spline Model Residual Values

The next step is to determine the model using Fourier Series estimator based on the residual values of the Spline model by finding the optimum K value by running the optimum K determination program that was created previously on case data on DHF in Central Java. The K value is a positive integer. Determination of the optimum K value in this research using the GCV method. The results obtained from testing K on the residual data of the Fourier Series model are as follows:

Table 5. GCV for Each Value of K

Value of K	GCV	Value of K	GCV
1	5.507×10^5	7	9.099×10^3
2	1.794×10^5	8	4.395×10^3
3	9.136×10^4	9	2.921×10^3
4	5.222×10^4	10	5.016×10^1
5	2.508×10^4	11	1.887×10^{-17}
6	1.264×10^4	12	9.751×10^{-20}

Table 5, it is presented that the GCV at $K = 12$ is the most optimum K because the GCV value is the smallest among the other GCV. If the value $K = 12$ is used then the estimated parameters are determined as many as 40 parameters, this is obtained based on Equation (8). Next, determining the best model can be done through the R-Square and MSE values for each K value. The following are the R-Square and MSE values for $K = 8$ to $K = 11$:

Table 6. The R^2 and MSE values for each K are optimum

Value of K	R^2	MSE
8	0.947 (94.7%)	219.0122
9	**0.992(99.2%)**	**33.57091**
10	0.997 (99.7 %)	9.882
11	0.999 (99.9%)	4.119×10^{-19}

Based on **Table 6**, it can be seen that the value of K = 9 produces high R^2, while the value of K = 8 gives R^2 result of 94.7%. If the K chosen to be used in modeling is K = 9 then parameters that must be estimated is 31. The model selection criteria are a model with a large R-Square, small MSE value and a parsimonious (simple) model, so the model selected is a model that has an optimum K value of 9. The model obtained from estimating residual values from Spline modeling using the Fourier Series estimator is as follows:

$$f(x_i) = 252.0305 - 4.412x_{i1} + 7.176\cos x_{i1} - 43.199\cos 2x_{i1} + \cdots - 1.618\cos 9x_{i1} + 44.791x_{i2} - 7.206\cos x_{i2} - 60.863\cos 2x_{i2} + \cdots + 46.72\cos 9x_{i2} + 44.791x_{i3} - 73.424\cos x_{i3} + 6.538\cos 2x_{i1} + \cdots + 2.554\cos 9x_{i1}$$

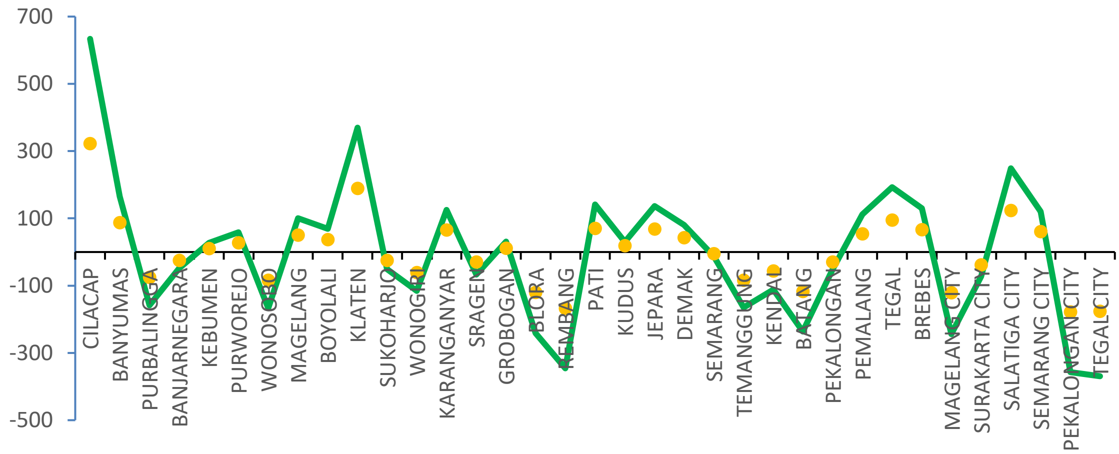

Figure 4. Scatterplot of Estimated Residual Values for the Spline Truncated Model Using the Fourier Series Approach

After obtaining estimates from the Spline Truncated model $\hat{s}(x)$ and the results of model estimation using the Fourier Series approach $\hat{f}(x_i)$, the two models are combined to become Spline Truncated-Fourier Series Nonparametric Regression Hybrid model as follows:

$$\hat{y}_i = \hat{s}(x) + \hat{f}(x_i)$$

when,

$$\hat{s}(x) = -161.98 + (-0.475)X_1 + 0.481{X_1}^2 + (-0.006){X_1}^3 + (-3003.78)(X_1 - 46)_+^3 + 56.364(X_1 - 53)_+^3 + 0.265X_2 + 10449.56{X_2}^2 + (-1303.596){X_2}^3 + 53.496(X_2 - 90)_+^3 + 0.041(X_2 - 93)_+^3 + \hat{\beta}_{11}X_3 + (-1.39\text{X}10^{-13}){X_3}^2 + (-0.828)(X_3 - 9)_+^3 + (-6418.706)(X_3 - 10.4)_+^3$$

$$\hat{f}(x_i) = 252.0305 - 4.412x_{i1} + 7.176\cos x_{i1} - 43.199\cos 2x_{i1} + \cdots - 1.618\cos 9x_{i1} + 44.791x_{i2} - 7.206\cos x_{i2} - 60.863\cos 2x_{i2} + \cdots + 46.72\cos 9x_{i2} + 44.791x_{i3} - 73.424\cos x_{i3} + 6.538\cos 2x_{i1} + \cdots + 2.554\cos 9x_{i1}$$

The following are the estimation model results of the Spline Truncated-Fourier Series estimator to modelling DHF in Central Java which are presented together with actual data:

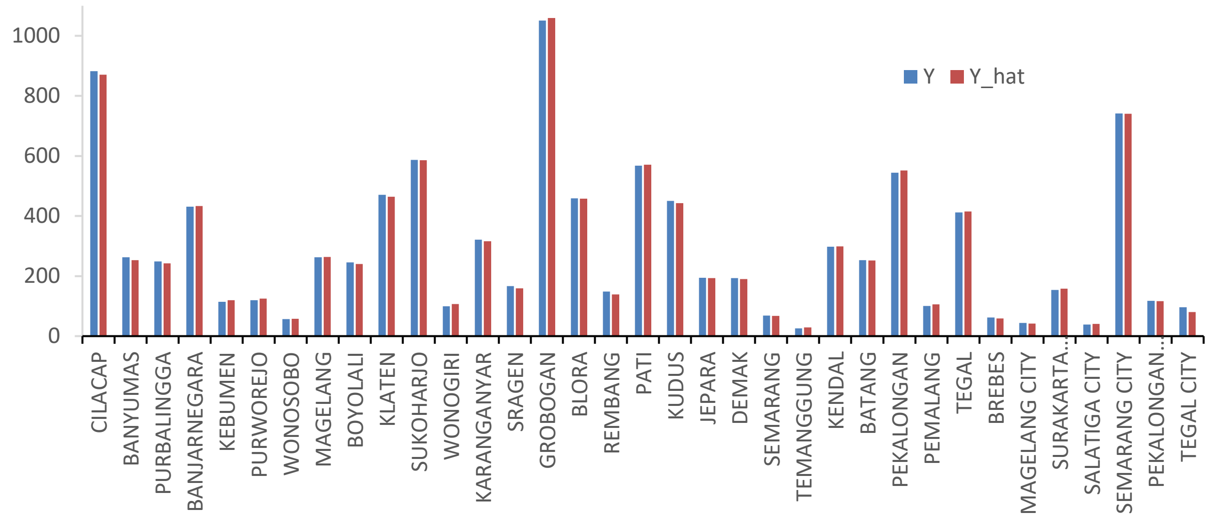

Figure 5. Graph of the Spline Truncated- Fourier Series Hybrid Model

Based on **Figure 5**, the graph showing the blue color is the actual data, while the red graph shows the results obtained from the estimated model. The results obtained from the model estimation show that the number of DHFs in the actual data produces values that are not much different from the estimated results. The modeling obtained can be said to be suitable for predicting DHF cases in Central Java, this can be seen from the R-Square value approaching 100% and the smallest MSE. The modeling results concluded that the highest DHF disease in Central Java was in Grobogan District, while the lowest DHF in Central Java was in Temanggung District.

The ability to evaluate model accuracy can be expressed through R-Square and MSE. The model is said to be better if the R-Square value approaching 100% and the smallest MSE. Results processed using hybrid Spline Truncated-Fourier Series regression applied to data on the number of DHF disease in Central Java show R-Square of 0.9994 (99.94%) and MSE of 33.586. R-Square is 99.94%, which means that the predictor variable is able to explain 99.94% of DHF disease in Central Java, while the remaining 0.06% of DHF disease is influenced by factors not used in this research. The following is a comparison table between the Spline Truncated model and the Truncated Spline-Fourier Series Hybrid model:

Table 7. Comparison of the Spline Truncated with the Spline Truncated - Fourier Series Hybrid Model

Nonparametric Regression	MSE	R-Square
Spline Truncated Model	10527.27	82.55%
Spline Truncated -Fourier Series Hybrid Model	**33.586**	**99.94%**

Based on **Table 7**, it can be compared that the Spline Truncated-Fourier Series hybrid model is better than the Spline Truncated model, this can be seen from the largest R-square, namely 99.94% and the smallest MSE. It can be concluded that the hybrid model can further minimize the error value so that the resulting R-square is close to 100%.

4. CONCLUSIONS

Based on the results obtained, it can be compared that the Truncated Spline-Fourier Series hybrid model is better than the Spline Truncated model, this can be seen from the largest R-square, namely 99.94% and the smallest MSE. It can be concluded that the hybrid model can further minimize the error value so that the resulting R-square is close to 100%. The results of Spline Truncated-Fourier Series Hybrid modeling to the number of DHF disease in Central Java can be concluded that the pattern between the actual data and the data estimated by the model is the same. The modeling obtained is suitable for predicting DHF disease in Central Java. The modeling results concluded that the highest DHF disease in Central Java was in Grobogan District, while the lowest DHF in Central Java was in Temanggung District.

The suggestion from the results of this research is that an optimal Knot selection method other than the GCV method can be used. Apart from that, research can be carried out to develop hybrid methods using other estimators such as local polynomials or kernels. Adding predictor variables that have not been used in the model can be used for further research.

ACKNOWLEDGMENT

Thank you to the Universitas Muhammadiyah Semarang for support and funding through the LPPM based on the "Penelitian Institusional" Program implementation contract letter number 085/UNIMUS.L/PT/PJ.INT/2023.

REFERENCES

[1] H. Harapan, A. Michie, M. Mudatsir, R. T. Sasmono, and A. Imrie, "Epidemiology of dengue hemorrhagic fever in Indonesia: analysis of five decades data from the National Disease Surveillance," *BMC Res Notes*, vol. 12, no. 1, p. 350, 2019, doi: 10.1186/s13104-019-4379-9.

[2] M.H. Atmoko, "Kemenkes catat kasus dengue sampai minggu ke-39 2022 capai 94.355", *ANTARA* 17 Oktober 2022 [Online]. Tersedia: https://www.antaranews.com/berita/3184081/kemenkes-catat-kasus-dengue-sampai-minggu-ke-39-2022-capai-94355 [Diakses: 22 Desember 2023].

[3] D. Sri Rejeki, N. Nurhayati, and B. Aji, "A spatiotemporal analysis of dengue hemorrhagic fever in Banyumas, Indonesia," *International Journal of Public Health Science (IJPHS)*, vol. 10, p. 231, Jan. 2021, doi: 10.11591/ijphs.v10i2.20713.

[4] T. W. Utami, A. Prahutama, A. Karim, and A. R. F. Achmad, "Modelling rice production in Central Java using semiparametric regression of local polynomial kernel approach," in *Journal of Physics Conference Series*, in Journal of Physics Conference Series, vol. 1217. May 2019, p. 12108. doi: 10.1088/1742-6596/1217/1/012108.

[5] D. Aydin, M. Memmedli, and R. E. M. Omay, "Smoothing Parameter Selection for Nonparametric Regression Using Smoothing Spline," *European Journal of Pure and Applied Mathematics*, vol. 6, no. 2, pp. 222–238, 2013.

[6] T. W. Utami, M. Haris, A. Prahutama, and E. Purnomo, "Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods," *Journal of Physics: Conference Series*, vol. 1446, p. 12049, Jan. 2020, doi: 10.1088/1742-6596/1446/1/012049.

[7] S. Wood, *Generalized Additive Models: An Introduction With R*, Boca Raton: Chapman and Hall/CRC, 2006. doi: 10.1201/9781315370279.

[8] D. A. Widyastuti, A. A. R. Fernandes, and H. Pramoedyo, "Spline estimation method in nonparametric regression using truncated spline approach," *Journal of Physics: Conference Series*, vol. 1872, no. 1, p. 12027, May 2021, doi: 10.1088/1742-6596/1872/1/012027.

[9] T. W. Utami, I. M. Nur, and Ismawati, "Fourier series nonparametric regression for the Modelizing of the Tidal," in *Proc. the 3rd International Seminar on Educational Technology: Global Education through Network Learnin*, pp. 119–123, 2017.

[10] D. R. Sari Saputro, K. R. Demu, and P. Widyaningsih, "Nonparametric truncated spline regression model on the data of human development index (HDI) in indonesia," *Journal of Physics: Conference Series*, vol. 1028, no. 1, p. 012219, 2018, doi: 10.1088/1742-6596/1028/1/012219.

[11] T. W. Utami and A. Prahutama, "Regresi semiparametri spline truncated dengan Software R," in *Seminar Nasional Pendidikan, Sains dan Teknologi*, 2017.

[12] D. Ispriyanti, A. Prahutama, and A. Taryono, "Modelling space of spread Dengue Hemorrhagic Fever (DHF) in Central Java use spatial durbin model," *Journal of Physics: Conference Series*, vol. 1025, p. 12112, Jan. 2018, doi: 10.1088/1742-6596/1025/1/012112.

[13] B. Sumargo, S. J. Kirana, and S. R. Rohimah, "Dengue hemorrhagic fever modeling using geographically weighted negative binomial regression," *AIP Conf. Proc.*, vol. 2588, no. 1, p. 050013, Jan. 2023, doi: 10.1063/5.0112673.

[14] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd Edition. New York: Marcel Dekker, 1999.

[15] A. Iriany and A. A. R. Fernandes, "Hybrid Fourier series and smoothing spline path non-parametrics estimation model," *Frontiers in Applied Mathematics and Statistics*, vol. 8, 2023. doi: 10.3389/fams.2022.1045098.

[16] A. Prahutama, Suparti, and T. W. Utami, "Modelling fourier regression for time series data- a case study: modelling inflation in foods sector in Indonesia," *Jounal of Physics: Conference Series*, vol. 974, no. 1, p. 12067, Mar. 2018, doi: 10.1088/1742-6596/974/1/012067.

[17] Sifriyani, A. R. M. Sari, A. T. R. Dani, and S. Jalaluddin, "Bi-Response truncated spline nonparametric regression with optimal knot point selection using generalized Cross-Validation in diabetes mellitus patient's blood sugas levels," *Communications in Mathematical Biology and Neuroscience*, vol. 2023, 2023, doi: 10.28919/cmbn/7903.

[18] I. Prasetyowati, S. Km, and M. Kes, "Population density and Incidence Rate of Dengue Hemorrhagic Fever (DHF) in Bondowoso distric, East Java," *The Indonesia Journal of Health Science*, vol. 5, no. 2, pp. 1–12, 2015.

[19] H. Adeswastoto and B. Setiawan, "Sebaran rumah sehat dan gambaran perilaku penghuni rumah tempat tinggal desa Sitorajo Kari kecamatan Kuansing Tengah," *Prepotif: Jurnal Kesehatan Masyarakat*, vol. 4, no. 1, pp. 85–92, Apr. 2023, [Online]. Available: http://journal.universitaspahlawan.ac.id/index.php/prepotif/article/view/6334

[20] Suparti, A. Prahutama, and R. Santoso, "Mix local polynomial and spline truncated: the development of nonparametric regression model," *Journal of Physics: Conference Series*, vol. 1025, no. 1, p. 12102, May 2018, doi: 10.1088/1742-6596/1025/1/012102.

[21] E. Madeira, A. Yudiernawati, N. Maemunah, "Hubungan Perilaku Hidup Bersih dan Sehat (PHBS) Ibu dengan Cara Pencegahan Demam Berdarah Dengue," *Nursing News Journal*, vol. 4, no. 1, pp. 288-299, 2019, doi: https://doi.org/10.33366/nn.v4i1.1549