DOI: 10.14421/ijid.2022.3300
IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 11, No. 1, 2022, Pp. 142-151

Fourier Series Nonparametric Regression Modeling in the Case of Rainfall in West Java Province

Anatansyah Ayomi Anandari
Department of Mathematics
UIN Sunan Kalijaga
Yogyakarta, Indonesia
17106010029@student.uin-suka.ac.id

Epha Diana Supandi
Department of Mathematics
UIN Sunan Kalijaga
Yogyakarta, Indonesia
epha.supandi@uin-suka.ac.id

Muhammad Wakhid Musthofa
Department of Mathematics
UIN Sunan Kalijaga
Yogyakarta, Indonesia
muhammad.musthofa@uin-suka.ac.id

Article History
Received Dec 9th, 2021
Revised Feb 11th, 2022
Accepted Feb 11th, 2022
Published July, 2022

Abstract— **The Fourier series is a trigonometric polynomial that has flexibility, so it adapts effectively to the local nature of the data. This Fourier series estimator is generally used when the data used is investigated for unknown patterns and there is a tendency for seasonal patterns. This study aims to determine the results of the best Fourier series nonparametric regression model and the level of accuracy of the Fourier series nonparametric regression model on rainfall data by month in West Java Province in 2015-2019. This research is about a nonparametric regression model of Fourier series which is estimated using *Ordinary Least Square* method. Nonparametric regression using the Fourier series approach was applied to Rainfall data in West Java Province in 2015-2019. The independent variables used were the average air humidity, air pressure, wind speed, and air temperature. The model used to model the amount of rainfall in West Java Province is a nonparametric Fourier series. The nonparametric regression model is the best Fourier series with K =13 values obtained Generalized Cross Validation, Mean Square Error, and R^2 respectively at 549.92; 462.09; and 97.30%. The results showed that the variables of air humidity and air pressure had a significant effect on rainfall.**

Keywords— *accuracy; seasonal patterns; Ordinary Least Square method; Generalized Cross Validation; Mean Square Error*

1 INTRODUCTION

In Indonesia, rain is a very important climate element due to its high diversity. In general, the nature of rain is that the shorter the rainfall, the higher the intensity and the greater the return period, the higher the intensity. According to the conditions in Indonesia, where the temperature is not so much and it changes quickly. The average amount of rain that falls each month and year in a place is not always the same.

Rain can be beneficial for living things, as explained in the word of Allah SWT. An-Nahl verse 10 as follows:

هُوَ الَّذِيْٓ اَنْزَلَ مِنَ السَّمَاۤءِ مَاۤءً لَّكُمْ مِّنْهُ شَرَابٌ وَّمِنْهُ
شَجَرٌ فِيْهِ تُسِيْمُوْنَ

"It is He who has sent down water (rain) from the sky for you, some of it as drink and some of it (fertilizing) plants, on it you graze your livestock."

High rainfall can also be the cause of natural disasters such as floods, landslides, and so on. Floods occurred in Bogor Regency, Indramayu Regency, Sukabumi City, Purwakarta Regency, and West Java in January 2020 which resulted in damaged houses and several people being killed because they were carried away by currents during floods. The presence of high rainfall will also trigger landslides as it has happened in several regencies and cities, including Sumedang, Cimahi, Bandung, Ciamis, Sukabumi, and Bogor in January 2020. Meteorology, Climatology and Geophysics (BMKG) stated that every province in Indonesia, especially in Java, that West Java Province ranks first for data on the highest amount of rainfall on Java Island in 2019. Therefore, it is very interesting to model the amount of rainfall that can be estimated and will be useful for practitioners and researchers.

This study is using the Fourier series approach. The Fourier series is a trigonometric polynomial that has flexibility, so it can adapt effectively to the local nature of the data. According to Suparti, et al. The Fourier series is very good for data that forms a distribution of sine and cosine waves. This Fourier series estimator is generally used if the data used is investigated for unknown patterns and there is a tendency for seasonal patterns [1]. This study will use a Fourier series model approach because this model is proven to be suitable for modeling and predicting the amount of rainfall. Research on Fourier series regression modeling in multivariable nonparametric regression refers to previous research conducted by Alan Prahutama [2], Ni Putu Ayu Mirah Mariati [3], Fatmawati Nurjanah, Tiani Wahyu Utami, Indah Benefiti Nur [4], and Intaniah Ratna Nur Wisisono, et al. [5], and Suparti, Rukun Santoso, Alan Prahutama, Hasbi Yasin, Alvita Rachma Devi [1]. The purpose of this study was to find the results of the best Fourier series nonparametric regression model for rainfall data by month in West Java Province in 2015-2019 and to determine the level of model accuracy.

West Java Province ranks first for data on the highest amount of rainfall on Java Island in 2019. Based on these problems, the aim of the research is trying to find the modeling rainfall data by month in West Java Province and detecting in which month the highest rainfall occurs. Thus, this study takes the title Fourier Series Nonparametric Regression Modeling in The Case of Rainfall in West Java Province, so that there are several criteria that must be considered to form the best nonparametric regression model with the *Generalized Cross Validation* (GCV) *K* optimal

The benefit of this research is that people can understand the application of the Fourier series in the field of life so that it can help the representation of rainfall by month in West Java Province and the variables or factors that influence it. This research can be used as a reference for information for local governments and the central government in tackling and alerting the occurrence of natural disasters in West Java Province.

Rainfall is the height of rainwater that collects in a flat place, does not evaporate, does not seep, and does not flow. Rainfall data is data with seasonal pattern and it is suitable to be modeled using nonparametric regression analysis with Fourier series approach [6]. The background of this research is because research on rainfall models and predictions with several independent variables has not been carried out. Research on Fourier series regression modeling in multivariable nonparametric regression refers to previous research conducted by Mariati which concluded that predictions using nonparametric regression with a Fourier series approach were good for data whose patterns were unknown and tended to repeat themselves [3]. The Fourier series is also very good for random data and it is not detected seasonally because the Fourier series is a nonparametric regression where its characteristics match the data [2]. In addition, there is also research conducted by Nurjanah, F., et al, which explains the application of a Fourier series nonparametric regression approach using Rainfall Data Patterns in Semarang City [4]. The purpose of this research is to analyze the rainfall pattern in the city of Semarang by producing a sine or cosine curve.

Wisisono, et al. examine Nonparametric Regression with the Fourier Series Approach, in which the nonparametric regression approach in this study was carried out using the Fourier series and explained the estimation of the Fourier series using OLS (Ordinary Least Square) and the GCV [5]. Application of nonparametric regression approach using Fourier series on Citarum river water discharge data. Then, Suparti, et al., examines the Analysis of Indonesian Inflation Data Using the Fourier and Wavelet Methods Multiscale Autoregressive, in the research conducted modeling the value of inflation in Indonesia from January 2007 to August 2017 [1]. The response variable from his research is the inflation rate, while the predictor variable is time with the aim of the study comparing the efficiency between the wavelet method and the Fourier method.

2 METHOD

2.1 Research Data

The object in this study is secondary data obtained from the annual publication report of the Central Statistics Agency (BPS) of West Java Province which contains a complete statistical description of the geographical, climatic, economic, social conditions in West Java Province, and so on. The data used by the author in this study is about rainfall data by month in West Java Province in 2015-2019 and several predictor variables which are factors that influence rainfall. Table 1 shows rainfall data by month in West Java Province in 2015-2019.

Table 1. Rainfall Data by Month in West Java Province in 2015-2019

Time	*Y*	Time	*Y*	Time	*Y*
2015	188	2017	65	2019	231
	189.1		199.3		269
	318.6		389.3		223
	285.2		220		299
	322.4		222.3		243
	58		106.4		27
	0.3		39.1		13
	6.9		48		0
	43.2		91		55
	37.9		345		84
	455		442		271
	311.5		129.9		316
2016	48	2018	191	2020	207.6
	34.5		239		336.6
	86.3		292		292.5
	112.6		298		271.4
	74.7		124		292.3
	38.2		33		30.3
	44.6		0		63.7
	49.7		39		41.6
	46.8		41		87.7
	83.5		125		327.3
	87.1		483		207.3
	15.2		324		262.1

Dependent variable of the amount of rainfall by month in West Java Province (mm^3) such as Average Air Humidity (%) (X_1), Average Air Pressure (mb) (X_2), Average Wind Speed (Km)/Hour) (X_3), and Average Air Temperature (°C) (X_4) [7].

2.2 Descriptive Statistics

Data used in this case is rainfall data by month in West Java Province in 2015-2019, with a lot of data n = 60, the dependent variable is the amount of rainfall. The average rainfall in West Java Province is 157.577 mm^3 with a variance of 10757.515 (Table 2). The average value of 157.577 mm^3 shows that the amount of rainfall by month in 2015-2019 is the same as for 60 months in West Java Province which is 157.577 mm^3 of the total rainfall in each month. The high fluctuating variance value is quite high as indicated by July which has the lowest average rainfall of 19.4 mm^3 and November which has the highest average rainfall of 347.62 mm^3 in West Java Province.

Table 2. Descriptive Statistics

Variable	Mean	Variance	Minimum	Maximum
Y	157.577	10757.515	19.4	347.62
X_1	74.558	24.645	65.68	80.08
X_2	924.425	0.228	923.8	925.22
X_3	4.198	0.152	3.66	4.76
X_4	4.453	0.082	24	24.98

The following will analyze the data pattern between the dependent variable: (1) the amount of rainfall with each independent variable, (2) the average air humidity, air pressure, wind speed, and average air temperature. The first factor that is thought to influence the dependent variable is the amount of rainfall is the average humidity of the air. Figure 1 is a picture that shows the pattern of the relationship between the dependent variable and the average air humidity variable.

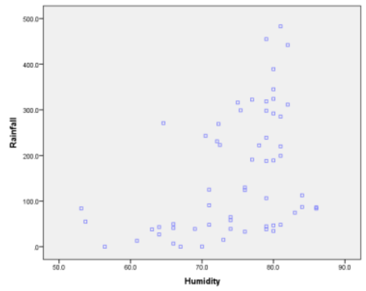

Figure 1. Scatter plot of rainfall data against average air humidity

The second factor that is thought to influence the dependent variable is air pressure. Figure 2 is a picture that shows the pattern of the relationship between the dependent variable and the air pressure variable.

Figure 3 is a picture that shows the pattern of the relationship between the dependent variable and the wind speed variable.

The fourth variable that is thought to influence the dependent variable on the amount of rainfall is the average air temperature. Figure 4 is a picture that shows the pattern of the relationship between the dependent variable and the average air temperature variable.

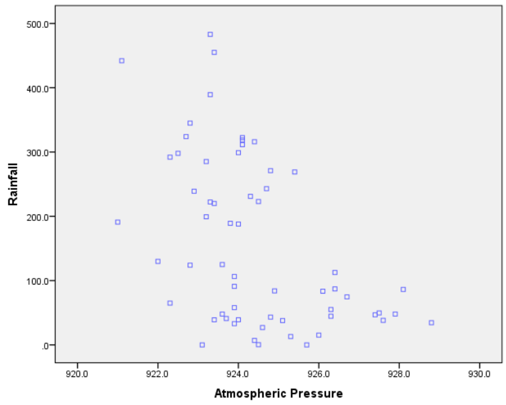

Figure 2. Scatter plot of rainfall data against air pressure

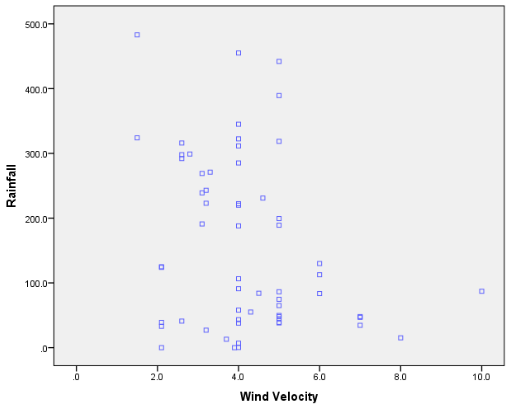

Figure 3. Scatter plot of rainfall data against wind speed

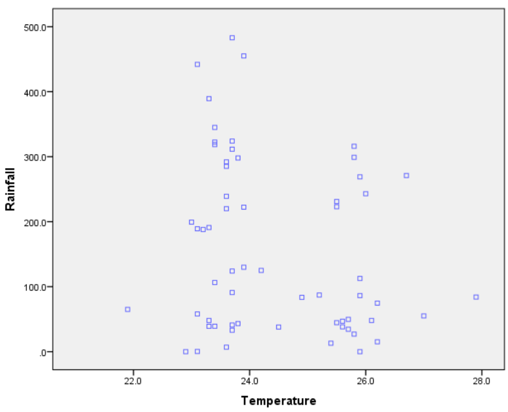

Figure 4. Scatter plot of rainfall data against average air temperature

In the data analysis method, the steps in the analysis of research data are described as follows (see Figure 5):

1) Creating *Scatter Plot*

   Through *a scatter diagram*, two kinds of information can be obtained, they are the pattern and estimation equation regarding the relationship between the two variables. The pattern of the relationship between the two variables is shown from the picture/curve obtained from the trend pattern of the point spread. While the estimation equation that shows the relationship between the two variables can be determined through the approximate curve obtained from the point distribution. A Scatter *Plot* between the amount of rainfall and the factors that influence it is made to determine the relationship between variables. *Scatter Plot* is used to identify the regression model used.

2) Fourier Coefficient (*K* Optimal)

   If *K* is large, then the curve of the estimator of the function *f* will be *smoother*, but the ability to approach data patterns is not good. Conversely, if *K* is small, then the curve of the estimator of the function *f* will be rougher. Therefore, it is necessary *K* that is neither too large nor too small so that it will produce a good curve estimator of the function *f*. Thus, the selection of *K* is very necessary. One method that can be used to determine the optimal smoothing parameter is the *Generalized Cross Validation* smallest or minimum

3) Parameter Estimation of Fourier Series Regression Model

   Estimating the parameters of the Fourier series regression model using the OLS method which can determine the parameter value by minimizing the number of squares of *errors*.

4) Creating a Fourier Series Nonparametric Regression Model

   After obtaining the estimator and *K* , then the factor influence model on the amount of rainfall in West Java Province uses Fourier series regression with the *K* optimal

5) Selecting the Best Model

   Several steps that have been taken previously, obtained several Fourier series regression models with *K* optimal *Akaike Information Criterion* (AIC) and *Bayesian Information Criterion* (BIC) which had a minimum value.

6) Testing the Feasibility of the Model

   Testing the significance of the model parameters can be used as a tool to determine whether or not there is a parameter relationship in the regression model. Parameter significance test is carried out simultaneously (simultaneously) or individually (partial). The feasibility test of the model aims to determine the significance of the regression model parameters together.

7) Testing *Error*

   Assumption test *error* can be used to find out whether classical assumptions are fulfilled, such as normal distribution, multicollinearity, heteroscedasticity, and autocorrelation. For normality test using Kolmogorov-Smirnov test, multicollinearity and heteroscedasticity test using *Rank* Spearman test, while for autocorrelation test using Durbin-Watson test.

8) Application of the Best Model to Rainfall Data

   Several steps have been taken, after obtaining the best Fourier series regression model, then testing the model parameters together and individually, and meeting the classical assumptions and the absence of deviations, the model obtained can be used to predict rainfall data in West Java Province.

The research steps are described in the form of a flowchart as follows (see Figure 5). The steps are added as clearer as possible.

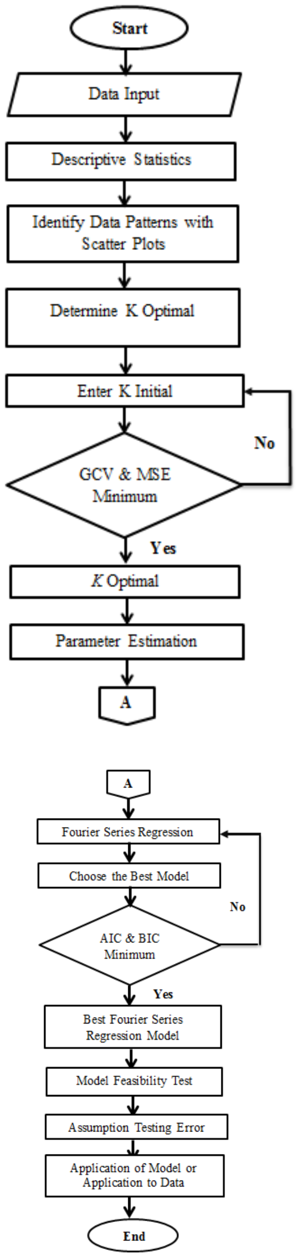

Figure 5. Flowchart of the analysis of research data

3 RESULT AND DISCUSSION

3.1. *Fourier Series*

The series was introduced by Joseph Fourier (1768-1830) to solve the problem of heat equations in metal plates. In mathematics, the Fourier series is the decomposition of periodic functions into the sum of oscillating functions, namely sine and cosine functions or complex exponentials [8]. According to Octavanny, et al. Fourier series is useful for describing curves showing sine and cosine waves, generally used when the data pattern is unknown and there is a tendency to iterate [9]. Chamidah states that the Fourier series is a trigonometric polynomial function that has a high degree of flexibility [10]. This is because the Fourier series is a curve that shows the sine cosine function [11]. The advantage of the Fourier series estimation is that it is able to handle data characters that follow repeated patterns at certain trend intervals and have good statistical interpretations [3].

3.2. *Definition of Periodic Functions and Fourier Series*

Function $f(x)$ is said to be periodic if the value of the function repeats itself at regular intervals. The regular interval between repetitions is the period of the oscillation, where the oscillation is the periodic variation with time of a measurement result.

According to Damanik, a function $f(x)$ [8] can be said to have a period of T or periodic with a period of $T > 0$ if for all x applies, as formulated in Formula 1:

$$f(x+T) = f(x) \tag{1}$$

Based on Formula 1 above, we can count as follows:

$$\begin{aligned} f(x+2T) &= f((x+T)+T) \\ &= f(x+T) \\ &= f(x) \end{aligned}$$

Thus, it can be concluded that $f(x+nT) = f(x)$.

Suppose a function $f(x)$ defined in the interval (-*L,L*) and outside this interval by $f(x+2L) = f(x)$, so $f(x)$ that it has a period of *2L*. Every function $f(x)$ which is periodic can be made a Fourier series, as can be seen on Formula 2 as follows:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right)\right) \tag{2}$$

where $a_0, a_n, and\ b_n$ are the Fourier coefficients.

The value of the Fourier coefficients can be obtained from the following Formula 3:

$$\begin{aligned} a_0 &= \frac{1}{L}\int_{-L}^{L} f(x)\,dx, \\ a_n &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx, \\ b_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx \end{aligned} \tag{3}$$

If we change the integral limit by letting the period *2L*, then we can use the following relation to determine the Fourier coefficients of a periodic function expanded into a Fourier series (Formula 4):

$$a_0 = \frac{1}{L}\int_c^{c+2L} f(x)\,dx,$$
$$a_n = \frac{1}{L}\int_c^{c+2L} f(x)\cos\left(\frac{n\pi x}{L}\right)dx, \quad (4)$$
$$b_n = \frac{1}{L}\int_c^{c+2L} f(x)\sin\left(\frac{n\pi x}{L}\right)dx$$

where *c* is any real number.

3.3. *Odd Function and Even Function*

A function can be an odd function or an even function or a combination of both [8]. A function is called an odd function if it satisfies the property:

$$f(-x) = -f(x)$$

and is called an even function if it satisfies the property:

$$f(-x) = f(x)$$

To determine the Fourier coefficients a_0, a_n, dan b_n of the periodic function even and odd functions with a period $L = \frac{1}{2}T = \frac{1}{2}$ the following Formula 5 is used:

If $f(x)$ is even, then:

$$a_0 = \frac{2}{L}\int_0^L f(x)\,dx$$
$$a_n = \frac{2}{L}\int_0^L f(x)\cos\frac{n\pi x}{L}dx \quad (5)$$
$$b_n = 0$$

For this case, it is said the function $f(x)$ is even and decomposes in the cosine series ($b_n = 0$). The graph of an even function is symmetric about the Y axis. Formula 6 is used.

If $f(x)$ is odd, then:

$$a_0 = 0$$
$$a_n = 0 \quad (6)$$
$$b_n = \frac{2}{L}\int_0^L f(x)\sin\frac{n\pi x}{L}dx$$

For this case, it is said the function $f(x)$ is odd and decomposes in a sine series ($a_n = 0$). The graph of the odd function is symmetric about the origin.

According to Damanik, in the Fourier series, which deals with an odd function, there may only be sine terms, while the Fourier series, which deals with an even function, may only contain cosine terms or a constant which we view as a term cosine [8]. This is a consequence of the odd and even function properties of integration.

Based on the properties of even and odd functions, it can be seen that the expansion of an odd function into a Fourier series only produces coefficients that correspond to terms containing sine $(a_n = 0,\ b_n \neq 0)$. Meanwhile, for an even function if it is expanded into a Fourier series, it only contains cosine terms or a constant that can be viewed as cosine $(because\ a_{ij} \neq 0, b_{ij} = 0)$. To get an even function or only odd function from a given function $f(x)$, we need to expand the interval (interval) first to get an even or odd function.

3.4. *Estimating Nonparametric Regression Parameters Fourier Series*

Given a multivariable nonparametric regression model [12], can be used to count using Formula 7:

$$Y_i = m(X_{1i}, X_{2i}, \ldots, X_{qi}) + \varepsilon_i$$
$$= \sum_{j=1}^{q} f_j(X_{ji}) + \varepsilon_i,\quad i = 1, 2, \ldots, n \quad (7)$$

Description:
Y_i : response variable on the *i* observation
X_{ji} : Nonparametric components where $X=X_{1i}, X_{2i}, \ldots, X_{qi}$
$\mathcal{E}_i$: residual value
$F_j(X_{ij})$: unknown regression curve
i : *1, 2, 3, …, n*
n : number of sample data

In nonparametric regression estimation using Fourier series, the function *f* is assumed to be unknown and contained in a continuous function space. The residuals are assumed to be normally distributed with a mean of zero and for constant and independent variance [13].

The model of the Fourier series equation is given on Formula 8 as follows:

$$Y_i = f(X_{ij}) + \varepsilon_i \quad (8)$$

where,
i : *1, 2, 3, …, n*
j : *1, 2, 3, …, m*

Based on the character of the rainfall data which is never negative, the more precise form of the Fourier series function as a predictor of rainfall is an even function. Even if the function is expanded into a Fourier series, then it only contains cosine terms or a constant that can be considered as cosine (because $a_n \neq 0, b_n = 0$).

Serov states that to determine the Fourier coefficients a_0, a_n, b_n of even and odd periodic functions with $L = \frac{1}{2}T$ [15] the following Formula 9 is used:

If $f(x)$ is even, then:

$$a_0 = \frac{2}{L}\int_0^L f(x)\,dx$$
$$a_n = \frac{2}{L}\int_0^L f(x)\cos\frac{n\pi x}{L}dx \quad (9)$$
$$b_n = 0$$

In this case, the function $f(x)$ is said to be even and decomposes in a cosine series $(b_n = 0)$. The graph of an even function is symmetrical about the *Y* axis. The function f in equation (8) is a continuous function, so that *f* can be approximated by using the function *T* which is a function of the Fourier cosine series [2], as can be seen in Formula 10 and 11:

$$T(t)=\frac{1}{2}a_0+bt+\sum_{k}^{K}a_k\cos kt \quad (10)$$

$$Y=f(X)+\varepsilon \quad (11)$$

X is a nonparametric component with $X=X_{i1},X_{i2},...,X_{im}$, so that

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix};f(X)=\begin{bmatrix}f(X_{11},X_{12},\dots,X_{1m})\\f(X_{21},X_{22},\dots,X_{2m})\\\vdots\\f(X_{n1},X_{n2},\dots,X_{nm})\end{bmatrix}\ and$$

$$\varepsilon=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}$$

The function $f(X_{ij})$ can be approximated using a Fourier series so that it becomes as shown on Formula 12:

$$f(X_{ij})=\frac{1}{2}\alpha_0+b_jX_{ij}+\sum_{k=1}^{K}\alpha_{ik}\cos kX_{ij} \quad (12)$$

with,

I = 1, 2, …, n : n = number of observations
j = 1, 2, …, m : m = number of independent variables
k = 1, 2, …, K : K = sum of fourier coefficients
α_0,α_k,b = parmeters in the model

The Fourier series regression model in Formula (12) when substituted into Formula (11), it is obtained:

$$Y=X\beta+\varepsilon \quad (13)$$

$$X=\begin{bmatrix}1 & X_{11} & \cos X_{11} & \cos 2X_{11}\cdots\cos KX_{11} & X_{12} & \cos X_{12} & \cos 2X_{12}\cdots\cos KX_{12}\cdots & X_{1m} & \cos X_{1m} & \cos 2X_{1m}\cdots & \cos KX_{1m}\\1 & X_{21} & \cos X_{21} & \cos 2X_{21}\cdots\cos KX_{21} & X_{22} & \cos X_{22} & \cos 2X_{22}\cdots\cos KX_{22}\cdots & X_{2m} & \cos X_{2m} & \cos 2X_{2m}\cdots & \cos KX_{2m}\\\vdots & \vdots & \vdots & \vdots\ \ddots\ \vdots & \vdots & \vdots & \vdots\ \ddots\ \vdots & \ddots\ \vdots & \vdots & \vdots\ \ddots & \vdots\\1 & X_{n1} & \cos X_{n1} & \cos 2X_{n1}\cdots\cos KX_{n1} & X_{n2} & \cos X_{n2} & \cos 2X_{n2}\cdots\cos KX_{n2}\cdots & X_{nm} & \cos X_{nm} & \cos 2X_{nm}\cdots & \cos KX_{nm}\end{bmatrix}$$

$$\beta=\begin{bmatrix}a_0^* & b_1 & a_{11} & a_{12}\cdots a_{1k} & b_2 & a_{21} & a_{22}\cdots a_{2k}\cdots b_m & a_{m1} & a_{m2}\cdots a_{mk}\end{bmatrix}^T$$

for $a_0^*=\frac{1}{2}a_0$

$$Y_i=\alpha_0+b_1X_{1i}+\sum_{k=1}^{K}\alpha_{k1}\cos kX_{1i}+b_2X_{2i}+\sum_{k=1}^{K}\alpha_{k2}\cos kX_{2i}+b_3X_{3i}+\sum_{k=1}^{K}\alpha_{k3}\cos kX_{3i}$$
$$+b_4X_{4i}+\sum_{k=1}^{K}\alpha_{k4}\cos kX_{4i}$$

The estimator $\hat{\beta}$ is obtained using the Least Square method is shown in Formula 14 as follows:

$$\begin{aligned}\varepsilon^T\varepsilon&=(Y-X\beta)^T(Y-X\beta)\\&=(Y^T-\beta^TX^T)(Y-X\beta)\\&=Y^TY-Y^TX\beta-\beta^TX^TY+\beta^TX^TX\beta\\&=Y^TY-2\beta^TX^TY+\beta^TX^TX\beta\end{aligned} \quad (14)$$

To get an estimator $\hat{\beta}$, the value $\varepsilon^T\varepsilon$ in equation (14), is reduced $\hat{\beta}$ so that it is obtained $\frac{\partial\varepsilon^T\varepsilon}{\partial\beta}=0$. Equation (14) becomes:

$$\begin{aligned}-2X^TY+2X^TX\hat{\beta}&=0\\2X^TX\hat{\beta}&=2X^TY\\X^TX\hat{\beta}&=X^TY\\\hat{\beta}&=(X^TX)^{-1}X^TY\end{aligned}$$

3.5. *Determination of the Number of Fourier Coefficients (K)*

One method that can be used to determine the number of Fourier coefficients (*K* optimal) is GCV [14]. The GCV for the Fourier series is shown in Formula 15 as follows:

$$GCV=\frac{MSE}{\left(n^{-1}trace\left(I-H(X)\right)\right)^2} \quad (15)$$

with $MSE=n^{-1}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$

The smallest GCV value will produce the *K* optimal value.

3.6. *The Best Model Selection Criteria*

According to Qudratullah, there are two information criteria that are often used, they are AIC and BIC [15] as shown in Formula 16 and 17.

$$AIC_m=-2\ln\left(\frac{JKS_m}{n}\right)+2p \quad (16)$$

$$BIC_m=-2\ln\left(\frac{JKS_m}{n}\right)+p\ln n \quad (17)$$

where is,
JKS_m : the number of squares of error for the m model
n : the number of sample data
p : the number of parameters of the m model
The best model that can be chosen is the model that gives the minimum AIC or BIC value.

3.7. Rainfall Modeling in West Java Province Using Fourier Series

Estimation of the model using the Fourier series approach can be obtained using Formula 18 can be seen as follows:

$$Y_i = \beta_0 + b_1 X_{1i} + \sum_{k=1}^{K} \alpha_{k1} \cos kX_{1i} + b_2 X_{2i} + \sum_{k=1}^{K} \alpha_{k2} \cos kX_{2i} + b_3 X_{3i} + \sum_{k=1}^{K} \alpha_{k3} \cos kX_{3i} + b_4 X_{4i} + \sum_{k=1}^{K} \alpha_{k4} \cos kX_{4i} \quad (18)$$

In the nonparametric regression of the Fourier series, it is very dependent on the Fourier coefficient (K). A K larger value will result in a more complex model and the estimation curve will follow the actual data pattern, so that the bias is smaller and the variance is larger. The K optimal value that can be selected is the value K that produces a minimum GCV value. The results of the analysis for the value are K given in the Table 3 below:

Table 3. GCV Values for Each Fourier Coefficient (K)

Value of Fourier Coefficient (K)	GCV
1	380971.11
2	157097.28
3	82572.65
4	47686.17
5	31686.39
6	19305.72
7	10056.82
8	7632.17
9	5151.57
10	2882.89
11	1653.63
12	1027.73
13	**549.92**
14	256.73
15	1.034×10^{-6}
16	2.686×10^{-3}
17	1.869×10^{-3}

It is necessary to look at the values MSE (Mean Square Error) and R^2 for each K (Table 4).

Table 4. MSE and R^2 Values for Each K

Value of Fourier Coefficient (K)	MSE	R^2
13	**462.09**	**97.30%**
14	239.90	98.60%
15	1.00074×10^{-6}	100%

Based on the output results of R Studio 4.0.0, the values are obtained in the AIC and BIC Fourier series nonparametric regression model with K= 13, K= 14, and K= 15, the results are given on Table 5 as follows:

Table 5. Results of AIC and BIC Values

Value of Fourier Coefficient (K)	AIC	BIC
13	**101.7284**	**221.1061**
14	111.0396	238.7946
15	157.6295	293.7619

As seen on Table 5, a nonparametric Fourier series regression model can be selected with the smallest values of AIC and BIC so that the selected Fourier series model is a Fourier series model with the number of K= 13.

Based on Table 3, the value of K= 13 produces a GCV value of 549.92. Nonparametric regression Fourier series with K= 13 produces a value of R^2 amounting to 97.30%, which means that the diversity of response values capable explained by the predictor variables of 97.30%. If the value K= 13, then had to estimate the parameters as many as 57 parameters. Thus, the estimation of the Fourier series nonparametric regression model for the model of the amount of rainfall in West Java Province is written on Table 6:

Table 6. Parameter Value for K=13

Parameter	Value	Parameter	Value	Parameter	Value
β_0	113685.36	α_{42}	-8.34	α_{93}	5824.75
b_1	-6.84	α_{52}	-39.05	α_{103}	10230.40
α_{11}	46.34	α_{62}	53.82	α_{113}	8190.23
α_{21}	-22.55	α_{72}	-117.50	α_{123}	3832.43
α_{31}	23.58	α_{82}	-16.64	α_{133}	741.41
α_{41}	74.80	α_{92}	23.05	b_4	93.94
α_{51}	-179.87	α_{102}	4.25	α_{14}	314.24
α_{61}	20.68	α_{112}	87.23	α_{24}	-187.59
α_{71}	26.06	α_{122}	-145.72	α_{34}	-213.57
α_{81}	-40.74	α_{132}	50.43	α_{44}	101.81

α_{91}	154.33	b_3	218.70	α_{54}	-280.65
α_{101}	-152.09	α_{13}	-32.85	α_{64}	23.09
α_{111}	-73.50	α_{23}	8954.64	α_{74}	-13.89
α_{121}	10.35	α_{33}	12481.80	α_{84}	-165.06
α_{131}	-167.99	α_{43}	5644.73	α_{94}	49.67
b_2	-128.43	α_{53}	-7616.12	α_{104}	-41.04
α_{12}	-29.27	α_{63}	-17113.59	α_{114}	-163.78
α_{22}	179.13	α_{73}	-15810.70	α_{124}	-35.24
α_{32}	-105.36	α_{83}	-5420.20	α_{134}	-165.73

The nonparametric regression model of the Fourier series with obtained *K*= 13 is as follows:

$$\begin{aligned}\hat{y} = 113685.36 &- 6.84x_1 + 46.34cosx_1 - 22.55cos2x_1 + 23.58cos3x_1 + 74.80cos4x_1 - 179.87cos5x_1 + 20.68cos6x_1 \\ &+ 26.06cos7x_1 - 40.74cos8x_1 + 154.33cos9x_1 - 152.09cos10x_1 - 73.50cos11x_1 + 10.35cos12x_1 \\ &- 167.99cos13x_1 - 128.43x_2 - 29.27cosx_2 + 179.13cos2x_2 - 105.36cos12x_2 - 8.34cos4x_2 \\ &- 39.05cos5x_2 + 53.82cos6x_2 - 117.50cos7x_2 - 16.64cos8x_2 + 23.05cos9x_2 + 4.25cos10x_2 \\ &+ 87.23cos11x_2 - 145.72cos12x_2 + 50.43cos13x_2 + 218.70x_3 - 32.85cosx_3 + 8954.64cos2x_3 \\ &+ 12481.80cos3x_3 + 5644cos4x_3 - 7616.12cos5x_3 - 17113.59cos6x_3 - 15810.70cos7x_3 \\ &- 5420.20cos8x_3 + 5824.75cos9x_3 + 10230.40cos10x_3 + 8190.23cos11x_3 + 3832.43cos12x_3 \\ &+ 741.41cos13x_3 + 93.94x_4 + 314.24cosx_4 - 187.59cos2x_4 - 213.57cos3x_4 + 101.81cos4x_4 \\ &- 280.65cos5x_4 + 23.09cos6x_4 - 13.89cos7x_4 - 165.06cos8x_4 + 49.67cos9x_4 - 41.04cos10x_4 \\ &- 163.78cos11x_4 - 35.24cos12x_4 - 165.73cos13x_4\end{aligned}$$

Model Feasibility Test

The purpose of the simultaneous test is to find out whether the independent variables jointly affect the regression model or not [16].

1. Hypothesis
   H₀: There is no effect of the independent variable on the dependent variable
   H₁ : There is an effect of the independent variable on the dependent variable
2. Level of Significance
   $\alpha = 5\% = 0.05$
3. Test Statistics

$$F_{count} = \frac{MSR}{MSE} \quad (19)$$

$$F_{table} = F_{\alpha,(t-1\,;\,n-t+1)} \quad (20)$$

where:
t: the number of parameters
n: the number of data

After doing the test using the *software* program *R Studio 4.0.0.* The results obtained by the value F_{count} of 2.577862 and the value F_{table} obtained from table *F* is

$F_{0.05(56;4)}$ =2.54.

4. Rejection Area
   If $F_{count} > F_{table}$ it is rejected or there is no influence of the independent variable on the dependent variable. Meanwhile, if $F_{count} > F_{table}$ it is accepted, it means that there is an influence of the independent variable on the dependent variable.
5. Conclusion
   Due to the value $F_{count} = 2.577862 > F_{table} = 2.54$ so that it is rejected or there is an influence of the independent variable on the dependent variable.

Implementation of the data on the best Fourier series nonparametric regression model for the amount of rainfall in West Java Province based on average air humidity (X_1), air pressure (X_2), wind speed (X_3), and average temperature (X_4) is shown in Table 7.

Table 7. Original Data and Predicted Data with Fourier Series Nonparametric Regression Model

Year	Y	$\hat{Y}$	Error
2020	207.6	247.37	-39.77
	336.6	334.90	1.69
	292.5	247.37	45.12
	271.4	267.81	3.58
	292.3	289.11	3.19
	30.3	27.70	2.59
	63.7	62.79	0.91
	41.6	38.61	2.99
	87.7	85.91	1.78
	327.3	324.43	2.86

207.3	205.35	1.94
262.1	260.93	1.16

Based on the graphs and predictions of the data of Table 7, it can be seen that the model can be used to accurately predict the rainfall data for West Java Province. In the prediction data, the obtained error by looking at the MSE value of the model is 462.09 and the coefficient of determination is 97.30%.

4 CONCLUSION

Based on the analysis above, several conclusions were made from this study, they are:

1. The best nonparametric Fourier series regression model for data on the effect of rainfall by month in West Java Province in 2015-2019 is K= 13.
2. The level of model accuracy is evidenced by the MSE value of the Fourier series nonparametric regression model for the effect of rainfall by month in West Java Province in 2015-2019 is 462.09 and the coefficient of determination is 97.30% so that the model can be used for predicting rainfall data for West Java Province very accurately.
3. Air humidity (X_1) and air pressure (X_2) variables have a significant effect on rainfall. This is evidenced by the correlation value X_1 is 0.867 and for X_2 is -0.736.

AUTHOR'S CONTRIBUTION

As the first author, Anatansyah Ayomi Anandari contributed technically and wrote to this research. She did the research with the supervision from the second author namely Epha Diana Supandi for ideas and technical aspects. Last, the research got theoretically supervision from the third author with Muhammad Wakhid Musthofa.

COMPETING INTERESTS

Comply with the publication ethics of this journal, Anatansyah Ayomi Anandari, Epha Diana Supandi and Muhammad Wakhid Musthofa as the authors of this article declare that this article is free from Conflict Of Interest (COI) or Competing Interest (CI).

ACKNOWLEDGMENT

Thank you to all parties for all input, moral and material support, corrections, and assistance throughout the research process so that this research can be completed.

REFERENCES

[1] S. Suparti, R. Santoso, A. Prahutama, and ..., "Analisis Data Inflasi Indonesia Menggunakan Metode Fourier dan Wavelet Multiscale Autoregresive," *... (Venue Artik. ...*, 2018.

[2] A. Prahutama, "PROSIDING SEMINAR NASIONAL STATISTIKA MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN DERET FOURIER PADA KASUS TINGKAT PENGANGGURAN TERBUKA DI JAWA TIMUR," *Pros. Semin. Nas. Stat. UNDIP*, 2013.

[3] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, "Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression," *J. Math.*, vol. 2020, 2020, doi: 10.1155/2020/4712531.

[4] F. Nurjanah, T. W. Utami, and I. M. Nur, "MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN DERET FOURIER PADA POLA DATA CURAH HUJAN DI KOTA SEMARANG," *J. Stat. Univ. Muhammadiyah Semarang*, vol. 3, no. 2, 2015.

[5] I. R. Nur Wisisono, A. I. Nurwahidah, and Y. Andriyana, "Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Data Debit Air Sungai Citarum," *J. Mat. "MANTIK,"* vol. 4, no. 2, 2018, doi: 10.15642/mantik.2018.4.2.75-82.

[6] W. Agustin, "Pola Distribusi Hujan Jam-jaman di Sub DAS Keduang," Universitas Sebelas Maret, Surakarta, 2010.

[7] E. Aldrian and R. Dwi Susanto, "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature," *Int. J. Climatol.*, vol. 23, no. 12, 2003, doi: 10.1002/joc.950.

[8] A. Damanik, *Fungsi-fungsi Khusus dilengkapi dengan Deret Fourier & Transformasi Integral*. Yogyakarta: Graha Ilmu, 2010.

[9] M. Ayu, D. Octavanny, N. Budiantara, H. Kuswanto, and D. P. Rahmawati, "A New Mixed Estimator in Nonparametric Regression for Longitudinal Data," 2021, doi: 10.1155/2021/3909401.

[10] N. Chamidah, *Buku Ajar Analisis Regresi Nonparametrik Menggunakan Program R*. Surabaya: Airlangga University, 2020.

[11] Ridwansyah, "Deret & Transformasi Fourier Dalam Penentuan Domain Frekuensi Sinyal Gelombang Radio Melalui Program Matlab," Makassar, 2016.

[12] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd ed. CRC, 1999.

[13] A. Munawwir, "Estimasi Parameter Model Regresi Menggunakan Metode Weighted Least Square (WLS) dengan Fungsi Pembobot Huber," UIN Maulana Malik, Malang, 2014.

[14] Suparti, R. Santoto, A. Prahutama, and A. R. Devi, *Regresi Nonparametrik*. Wade Group National Publishing, 2018.

[15] M. F. Qudratullah, *Analisis Regresi Terapan: Teori, Contoh Kasus, dan Aplikasi dengan SPSS*. Yogyakarta: Andi, 2013.

[16] R. A. Malau, *Metode Statistika Nonparametrik*, First. Jakarta: Universitas Terbuka, 2007.

.