PREDICTION OF RAINFALL IN DKI JAKARTA PROVINCE BASED ON THE FOURIER SERIES ESTIMATOR

Nur Chamidah[*], Diana Ulya, Qumadha Zainal Abidin, Syifa'un Nadhiro, Zidni 'Ilmatun Nurrohmah

Department of Mathematics, Faculty of Science and Technology, Airlangga University, Indonesia

*e-mail: *nur-c@fst.unair.ac.id*

Abstract: Rainfall is the height of rainwater in a rain gauge on a flat place that does not seep and flow, where rainfall is measured in millimeters (mm). This study aims to estimate and model the rainfall for DKI Jakarta Province from January 2016 to December 2021 using the Fourier series estimation. Based on the results of the study, a model with a minimum GCV value of 21909,4, at the 7th λ 43,78972. This model shows that the predictor variable can explain the diversity of response variables by 94,14%.

1. INTRODUCTION

Weather is one aspect that affects human life. Humans have various activities that depend on the weather both individually and in companies. Bad weather conditions not only affect human activities when outdoors, but also affect indoor activities because mobilization from one building to another can be hampered. Extreme changes in weather such as high heat waves and rainfall can also negatively affect physical systems, the environment, and people's lives. Disasters with high intensity occurring in Southeast Asian countries, especially Indonesia, such as: floods, droughts, hurricanes continue to increase in tandem with climate change. Natural disasters that occur due to extreme weather changes can affect the agricultural sector, water resources, and infrastructure [1]. Currently there is a phenomenon of a shift in the rainy season, this can be identified from the uncertain timing of rain in various regions in Indonesia. In addition to causing floods and landslides, high rainfall has also resulted in damage to infrastructure such as bridges and roads, thus disrupting the economic activities of residents.

Jakarta as the center of the economy in Indonesia has a very crucial role. Bad weather has the potential to disrupt Jakarta's economy, such as floods, heat waves, and strong winds. Based on the official website of the Central Bureau of Statistics for DKI Jakarta Province, every district/city in DKI Jakarta has flood-prone sub-districts except Pulau Seribu, and each administrative city has flood-prone sub-districts, there are 5 sub-districts that have high potential for flooding in each city. The administrative divisions are Tanah Abang District in Central Jakarta, Mampang Prapatan District in South Jakarta, Makassar District in East Jakarta, Cengkareng District in West Jakarta, and Penjaringan District in North Jakarta. On February 2 2020, DKI Jakarta received an accumulated rainfall of 1,043.3 mm which is the accumulated daily rainfall from the monitoring of the BMKG observation station. Meanwhile, according to publications on the Jakarta Open Data website in 2020, the achievements of flooded areas have increased compared to 2013, namely from 263 sub-districts to 603 sub-districts. This figure

occupies the second highest position since the rain in January 2014. In line with this news, the Central Bureau of Statistics also recorded that the average rainfall in 2020 reached 235.958 mm. Therefore, to ensure that economic activities run well, it is necessary to predict Jakarta's rainfall using time series data to see the data patterns formed so that appropriate anticipatory steps can be taken. As for one method of forecasting the level of rainfall that can use non-parametric regression analysis using the Fourier series estimator. Analysis with this method can be used to estimate functions or curves from data with no known patterns, especially for data that tends to have seasonal or time series [2].

2. LITERATURE REVIEW
2.1. Rainfall

Rain is a water cycle that is useful in maintaining the balance of water in the universe. The water cycle or hydrologic cycle is the never-ending circulation of water from the atmosphere to the earth and back to the atmosphere. The process can occur through condensation, precipitation, evaporation, and transpiration. The cycle of absolute rain occurs every year and makes water a very important resource for the continuation of life on earth.

Rainfall is the height of rainwater that collects in a flat place, does not evaporate, does not seep, and does not flow [3]. Rainfall is also defined as the height of water (mm) received by the surface before experiencing runoff, evaporation, and infiltration into the soil. Air humidity is the amount of water vapor in the air (atmosphere) at a certain time and place. Air humidity is determined by the amount of water vapor contained in the air. Air temperature is the state of hot or cold air [3].

According to the Meteorology, Climatology and Geophysics Agency (BMKG), one mm of rainfall is equivalent to one liter of rainwater in an area of one square meter. A dry condition occurs when the rainfall is less than 50mm/10 days. Conversely, the rainy season will occur when rainfall reaches more than or equal to 50 mm/10 days. High and evenly distributed rainfall throughout the year will be a good source of water and cause water supply to fluctuate. Data regarding rainfall have characteristics that that occurred in the past and can describe the characteristics of rainfall events that will occur in the future

2.2. Nonparametrc Regression

Nonparametric regression is a method used to determine the pattern of relationship between the response variable and the predictor if the relationship pattern of the two is not clearly known [4]. Suppose there is paired observation data (x_i, y_i), then the nonparametric regression model for observations i = 1, 2, …, n is shown in equation (1)

$$y_i = m(x_i) + \varepsilon_i ; \ \varepsilon_i \sim IIDN(0, \sigma^2) \qquad (1)$$

Where is the response variable, is the predictor variable for the nonparametric regression and the function $m(x_i)$ is the unknown function and is estimated by the functions in the nonparametric regression. Meanwhile, ε_i it is a random error which is assumed to be identically independent, and normally distributed with mean, varians σ2 [4].

In nonparametric regression, the researcher looks for the curve itself without being influenced by subjectivity. The nonparametric regression function only assumes smooth. One of the nonparametric regression approaches is the Fourier series [5].

2.3. Fourier Series Estimator

Fourier series is one of the approaches used in nonparametric regression. The Fourier series is a model that has a trigonometric polynomial function that has a high degree of flexibility which is generally used for data whose pattern is unknown and there is a tendency for a seasonal pattern [4].

Suppose given observational data (t_r, y_r) that follow the general regression model as follows:

$$y_r = m(t_r) + \varepsilon_r, \quad r = 1, 2, \dots, n \tag{2}$$

Where the form of the regression function $m(t_r)$ is unknown and will be estimated by a nonparametric regression approach using the Fourier series estimator and it is assumed that $m(t_r) \in L_2[a, b]$ which is a Hilbert space. From equation (2) $m(t_r)$ can be expressed as:

$$m(t_r) = \sum_{j=1}^{\infty} \beta_j x_j(t_r) + \varepsilon_r, \quad r = 1, 2, \dots, n \tag{3}$$

With β_j is a scalar so that the model in (3) becomes:

$$y_r = \sum_{j=1}^{\infty} \beta_j x_j(t_r) + \varepsilon_r, \quad r = 1, 2, \dots, n \tag{4}$$

If n infinite then the regression $m(t_r)$ can be approximated by equation

$$m(t_r) = \sum_{j=1}^{\lambda} \beta_j x_j(t_r), \quad r = 1, 2, \dots, n \tag{5}$$

with λ is an integer, then equation (5) becomes:

$$y_r = \sum_{j=1}^{\lambda} \beta_j x_j(t_r) + \varepsilon_r, \quad r = 1, 2, \dots, n \tag{6}$$

assuming $t_1, t_2, \dots, t_n$ equidistant $[a, b]$.

The method of estimating the unknown Fourier coefficients is to determine the optimal value with the λ optimal expressing the number of Fourier coefficients β_j hat determine the smoothness of the regression curve. The Fourier series estimator for $\hat{m}(t_r)$ can be written as follows:

$$\hat{m}(t_r) = \hat{\beta}_0 + \sum_{j=1}^{\lambda} \left[a_j \cos \cos \left(\frac{2\pi j(r-1)}{n} \right) + b_j \right. \\ \left. \sin \sin \left(\frac{2\pi j(r-1)}{n} \right) \right] \tag{7}$$

with

$$a_j = \frac{2}{n}\sum_{r=1}^{n} y_r \; coscos \; cos\left(\frac{2\pi j(r-1)}{n}\right) \qquad (8)$$

$$b_j = \frac{2}{n}\sum_{r=1}^{n} y_r \; sin \; sin\left(\frac{2\pi j(r-1)}{n}\right) \qquad (9)$$

2.4. Optimal Bandwith Determination

Bandwidth selection is considered very important because it influences the Fourier series nonparametric regression model to be selected. The optimal bandwidth value (λ) indicates the number of Fourier coefficients that can determine the smoothness of the function or regression curve. There are two strategies for choosing a good bandwidth. The first strategy is to choose a relatively small amount of bandwidth, while the second strategy is the opposite, namely using a relatively large amount of bandwidth. Between the two strategies, the second strategy is more widely used in models that pay close attention to the mathematical patterns in the data. While the first strategy is more directed at reasons of model simplicity.

Choosing a bandwidth that is too small will result in an undersmoothing curve that is very rough and very volatile, and conversely a bandwidth that is too wide will produce an over-smoothing curve that is very smooth but does not match the data pattern. Determining different bandwidth locations will produce different Fourier series nonparametric regression models. The location of the bandwidth will affect the criterion value of the nonparametric Fourier series regression model formed.

One of the optimal bandwidth point selection methods is Generalized Cross Validation (GCV). The appropriate Fourier series nonparametric regression model is related to the optimal bandwidth point obtained from the minimum GCV value. The GCV function is defined as follows:

$$GCV(\lambda) = \frac{MSE\,(\lambda)}{\left(n^{-1}trace(I-A)\right)^2} \qquad (10)$$

$$GCV(\lambda) = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n}$$

2.5. Mean Absolute Percentage Error (MAPE)

According to Wei [6] one of the criteria for determining the accuracy of forecasting results is to use the *Mean Absolute Percentage Error* (MAPE) value which is formulated as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\dfrac{\left|Z_t - \hat{Z}_t\right|}{Z_t}}{n} \times 100\% \qquad (11)$$

with Z_t is the actual value at time-t, $\hat{Z}_t$ is the predicted value at time-t, and n is the number of observations. MAPE values can be interpreted into four categories, namely:

Table 1. MAPE Criteria

MAPE Range	Interpretation
$< 10\%$	Very good forecasting result
$10 - 20\%$	Good forecasting result
$20 - 50\%$	Fair forecasting result
$> 50\%$	Bad forecasting result

3.   METHODOLOGY
3.1 Data Source

This study uses secondary data obtained from the Central Statistics Agency (BPS) regarding monthly rainfall in DKI Jakarta from January 2016 to June 2021. It uses data from January 2016 to October 2020 as in sample, while the rest is out sample.

3.2 Research Variables
Variables used in this study include:

Table 2. Research Variables

Variable Type	Variable Name
Predictor Variable	Time (January 2016-June 2021)
Response Variable	Rainfall

3.3 Data Analysis Procedure
1. Modeling monthly rainfall data in DKI Jakarta Province with nonparametric regression based on the Fourier series approach.
   a. Seeing the observation plots and determining the estimation model by finding optimal lambda.
   b. After obtaining the best model, the equations are described based on the best knot points and interpreted.
2. Predict or forecast rainfall for the next eight periods and calculate the MAPE value

4. RESULTS AND DISCUSSION

Based on the data sourced from Central Bureau of Statistics, the general description of

rainfall in DKI Jakarta Province as follows

Table 3. Descriptive Statistics

Variable	Mean	StDev	Minimum	Maximum
Rainfall	181,2	176,2	0,8	1043,2

Table 3 shows the average rainfall in DKI Jakarta Province from January 2016 to October 2020 was 181,2 mm. The lowest rainfall was 0,8 mm which occurred in August 2017. Meanwhile, the highest rainfall reached 1043,3 mm which occurred in February 2019.
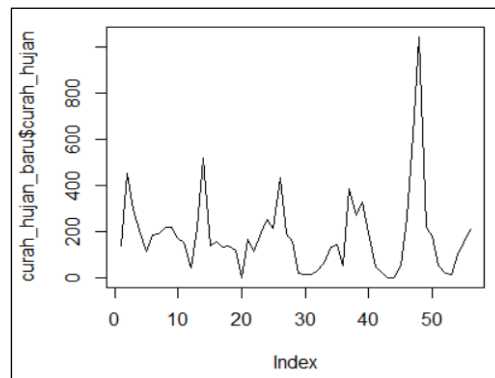
**Figure 1.** Rainfall Plot in DKI Jakarta Province

Figure 1 shows that the rainfall data has a fluctuating and periodic trend pattern. So that, Fourier series estimation can be used for determining the best model in this research. In Fourier series estimation, an optimal bandwidth value (λ) states the number of Fourier coefficients that determine the smoothness of regression curve which can be obtained by GCV method. The results of optimal GCV value using R software are presented in Table 4.

Table 4. Fourier Coefficients and GCV Value

λ	GCV Value
1	33115,2
2	35192,0
3	37310,9
4	39627,6
5	26959,9
6	25217,8
7	21909,4

8	22350,8

From the GCV value table above, a plot can also be formed to make it easier to determine the optimal lambda value which shown in Figure 2.
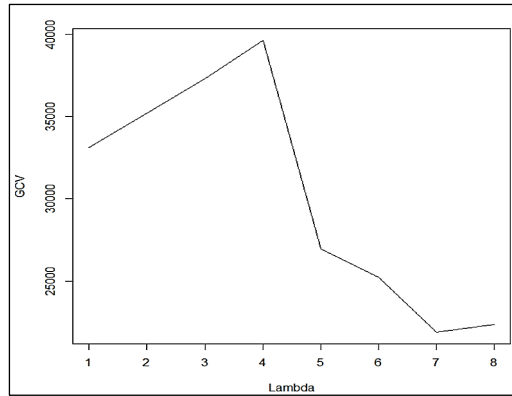

Figure 2. Plot of Fourier Coefficients and GCV Value

Based on the plot in Figure 2, it can be seen that the GCV value is at the minimum point when λ equal to 7 was chosen. The calculation results presented in the Table 4 also show that the smallest or minimum GCV value is 21909,4. By using $\lambda = 7$, the best model can be found and the estimation model equation is obtained as follows

$$
\begin{aligned}
m(t_r) = {} & 181{,}2473 + 39{,}38398 \cos 2\pi t_r - 8{,}877956 \sin 2\pi t_r - 20{,}46878 \cos 4\pi t_r \\
& - 22{,}99694 \sin 4\pi t_r - 26{,}99081 \cos 6\pi t_r + 22{,}71039 \sin 6\pi t_r \\
& - 34{,}51583 \cos 8\pi t_r + 10{,}65884 \sin 8\pi t_r + 59{,}58739 \cos 10\pi t_r \\
& + 132{,}1334 \sin 10\pi t_r + 70{,}44903 \cos 12\pi t_r - 10{,}83298 \sin 12\pi t_r \\
& + 46{,}03793 \cos 14\pi t_r - 64{,}25866 \sin 14\pi t_r
\end{aligned}
$$

where $t_r = \frac{t_i - 1}{n}$.
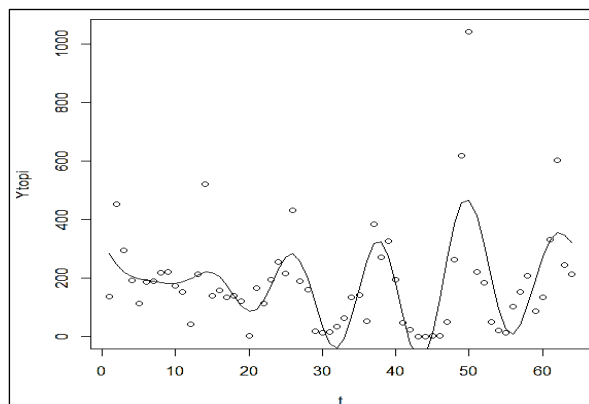
Figure 3. Plot of Fourier Series Estimation Model

The plot presented in Figure 3 shows the estimation results which are not too smooth, but not too rough either. In addition, an R-Square value of 94,14% and an MSE value of 43,78972

were obtained. Furthermore, the estimation model that has been obtained can be used to predict data for several future periods. In this study, predictions were made for the next eight periods of rainfall data. The following table shows a comparison between the out-sample data and the results of rainfall predictions.

Table 5. Prediction Results of DKI Jakarta Rainfall

Period	Out-Sample	Data Prediction	APE
59	87,3	418,6207	3,795197
60	134,7	536,17967	2,980547
61	332,8	585,5999	0,759615
62	604,4	554,50503	0,082553
63	244,1	451,4727	0,849539
64	213,9	303,49392	0,418859
65	203,6	148,13766	0,272408
66	79,1	22,850187	0,711123
	MAPE		14,95%

Based on the results in Table 5, the out-sample MAPE value generated based on the Fourier series estimator model is 14,95% which means that the Fourier series estimation model is a good model for predicting rainfall in DKI Jakarta Province.

5. CONCLUSION

Based on the analysis and prediction of rainfall in DKI Jakarta Province using the Fourier series approach, the best model is obtained with a minimum GCV value of 7 and an MSE value of 43,78972 and a GCV value of 21909,4. This model also shows that the predictor variable can explain the diversity of the response variables by 94,14%. The estimated model equation for Rainfall in DKI Jakarta Province is

$$\begin{aligned}
m(t_r) = {} & 181,2473 + 39,38398 \cos 2\pi t_r - 8,877956 \sin 2\pi t_r - 20,46878 \cos 4\pi t_r \\
& - 22,99694 \sin 4\pi t_r - 26,99081 \cos 6\pi t_r + 22,71039 \sin 6\pi t_r \\
& - 34,51583 \cos 8\pi t_r + 10,65884 \sin 8\pi t_r + 59,58739 \cos 10\pi t_r \\
& + 132,1334 \sin 10\pi t_r + 70,44903 \cos 12\pi t_r - 10,83298 \sin 12\pi t_r \\
& + 46,03793 \cos 14\pi t_r - 64,25866 \sin 14\pi t_r
\end{aligned}$$

MAPE value obtained as an out-sample result of predicting rainfall using a nonparametric regression model with a Fourier series estimator is 14.95% which means the estimation model above is a good model. Rainfall is a crucial natural factor and has the potential

to disrupt the economy if it occurs abnormally. Therefore, forecasting rainfall is expected to increase public awareness in dealing with natural situations in the future. Of course, predicting rainfall is not limited to using the Fourier series approach only. In future research, it is hoped that predictions can be made using different methods, for example SARIMA or others and can use and/or add factors that can influence rainfall.

REFERENCES

[1]   Handmer, J., Honda, Y., Kundzewicz, Z. W., Arnell, N., Benito, G., Hatfield, J., et al,"Changes in Impacts of Climate Extremes: Human Systems and Ecosystems. In C.B. Field, V. Barros, T. F. Stocker, & Q. Dahe (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation," Cambridge:Cambridge University Press, pp. 231–290, 2012

[2]   N. Chamidah, S. D. Febriana, R. A. Ariyanto, and R. Sahawaly , "Fourier series estimator for predicting international market price of white sugar," AIP Conference Proceedings 2329, 060035, 2021

[3]   Desvina, A. P., dan Ratnawati,"Penerapan model vector autoregressive (VAR) untuk peramalan curah hujan kota pekanbaru. Jurnal Sains, Teknologi dan Industri," 11(2), pp. 151–159, 2014

[4]   Eubank. R, L, "Nonparametric Regression and Spline Smoothing Second Edition," New York: Marcel Deker, 1999

[5]   Bilodeau. M, "Fourier Smootherand Additive Models, TheCanadian of Statistic," 3, pp. 257-259, 1992

[6]   Wei, W. W. S. "Time Series Analysis, Univariate and Multivariate Methods, Second Edition," United State of America: Pearson Education, Inc, 2006