Implementation of autoregressive integrated moving average model to predict total electron content from GPS satellite receiver in Bandung

Marifatul Nur Yuniati*, Agus Sugandha

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Jenderal Soedirman University

*E-mail: marifatul.yuniati@mhs.unsoed.ac.id

Abstract. Total electron content (TEC) or free-electron content that is in the ionosphere layer can disrupt telecommunication and navigation systems. This study discusses forecasting the daily average value of total electron content (TEC) during March 2019. The method used in this forecast is the autoregressive integrated moving average (ARIMA) model with the Box-Jenkins approach. The research results show that the best ARIMA model is ARIMA (1,1,1) with an MSE 2.453. The results showed that the daily average predicted TEC value increased every day and the maximum TEC value occurred in the 86th period (86th day) which was 18.4991 TECU (1 TECU =10¹⁵ electrons / m²).

1. Introduction
The ionosphere layer is part of the upper layer of the atmosphere where many free electrons affect the propagation of radio waves [1]. The GPS satellites which are useful for navigation by ionosphere researchers can be used as tools for understanding the ionosphere layer. GPS satellite signals will experience refraction and change in intensity as they pass through the ionosphere before being accepted by GPS signal receivers on earth. The changes that occur in the GPS signal as it passes through the ionosphere contain information on the state of the ionosphere. This information is the number or content of electrons in the ionosphere. The electron content in the ionosphere is called the total electron content (TEC)[2].

The propagation of radio waves through the ionosphere will experience a delay time due to interference with free electrons in the ionosphere. With this delay in the ionosphere, the measurement of the distance between the satellite and the GPS receiver based on the measurement of propagation time will experience an error, so that GPS positioning using the distance resection method will experience measurement errors. For high-precision positioning, the ionosphere TEC needs to be estimated so that it can be used for the correction of GPS satellite distance measurement errors [3].

The TEC value can increase, one of which is due to the influence of the seasonal pattern of the sun, wherein the equinox months (March-April, September-October) the sun is close to the Indonesian equator. If the TEC value is higher, it will affect the propagation of radio waves so that in everyday life the telecommunications and navigation systems will be disrupted [2]. In this study, the TEC daily average data was used from January to February 2019. Because the TEC value is daily time series data and has an up and down trend pattern, the data analyzed using time series analysis [4]. Time series
data can be used as the basis for making provisions to predict events that will occur in the future era. The autoregressive integrated moving average (ARIMA) method is a very popular model and is often used in time series data modeling. So that in this study using the ARIMA method assumption with the Box-Jenkins approach to predict the daily average TEC value in March 2019.

2. Methodology
Modeling and forecasting using the ARIMA method consist of several stages, included model identification, parameter estimation, diagnostic test, best model selection, and forecasting [5].

2.1. Model Identification
Model identification is carried out by the following steps:
- Plot the data to examine whether the data have seasonal patterns or not.
- Check the stationarity of the data (stationary in the mean and stationary in the variance). If it is nonstationary in the mean, perform the differencing, and if it is nonstationary in the variance, perform a transformation.
- Plot ACF and PACF of the data, to see whether the data has been stationary after transformation and differencing. The plot also can be used to decide the possible order of ARIMA models.

2.2. Parameter Estimation
The next step is to estimate the possible models and test their significances. The test is carried out to check whether the model is suitable for use or not. The process of estimating and testing the significance of the parameters can be performed using the Minitab 16 software.

2.3. Diagnostic Test
The diagnostic test consists of a normality test and a white noise test. Normality assumption can be tested using the Kolmogorov-Smirnov test and the white noise test or independence test can be performed using the Ljung-Box test [5].

2.3.1. Kolmogorov-Smirnov test. To find out the residuals that are normally distributed, the residual normality test can be done using the Kolmogorov-Smirnov test.
- Hypothesis
  \[ H_0 : F(x) = F_0(x) \text{ (residuals are normally distributed) } \]
  \[ H_1 : F(x) \neq F_0(x) \text{ (residuals are not normally distributed) } \]
- Significant Level:
  \[ \alpha = 5\% = 0.05 \]
- Area of rejection:
  Reject \( H_0 \) if \( p\)-value < \( \alpha \)
- Decision
  If \( p\)-value < \( \alpha = 0.05 \) then reject \( H_0 \), meanwhile if \( p\)-value > \( \alpha = 0.05 \) failed to reject \( H_0 \).

2.3.2. White Noise. To find out whether the autocorrelation of the residual value is zero or not, it is carried out through the Ljung-Box test.
- Hypothesis
  \[ H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0 \text{ (residuals is white noise) } \]
  \[ H_1 : \rho_k \neq 0, k = 1, 2, 3, \ldots, n \text{ (residuals is not white noise) } \]
- Significant Level
- Area of rejection:
  Reject \( H_0 \) if \( p\)-value < \( \alpha \)
• Decision
  If \( p\text{-value} \ < \alpha = 0.05 \) then reject \( H_0 \), meanwhile if \( p\text{-value} \ > \alpha = 0.05 \) failed to reject \( H_0 \).

2.4. Best Model Selection
There are several residual or error-based methods can be used to select the best model, one of which is the mean square error (MSE)[6]. MSE is formulated as

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (x_t - \hat{x}_t)^2
\]  

(1)

where

\( n \) : number of observations
\( x_t \) : observational data at time \( t \)
\( \hat{x}_t \) : forecast value at time \( t \)

3. Results And Discussion

3.1. Model Identification.
In this stage, two tests were carried out, the stationary test for variance and the stationary test for means using past data on the TEC value to identify it, at this stage a suitable ARIMA model will be sought and considered in accordance with the data obtained by the researcher.

3.1.1. Variance Stationary Test. The box-cox test is carried out to find out whether a data has fluctuations that tend to be constant or not, in a box-cox test data that tends to be constant or stationary will have a rounded value \((\lambda) = 1\).

![Figure 1. Box-cox plot of TEC value.](image1)

[Figure 1. Box-cox plot of TEC value.]

Box-cox test results on the total electron value data content (TEC) can be seen in Figure 1. It is known that the data is not stationary to the variance because the rounded value is \((\lambda) \neq 1\), so it needs to be transformed according to the lambda value \((\lambda)\). The data resulting from the box-cox transformation with the formula \(\sqrt{\lambda} \) causes the rounded value \((\lambda) = 1\), as shown in Figure 2, so it can be concluded that the data is stationary to variance.

3.1.2. Means Stationary Test. Identification of stationary means data includes identification of actual data plots, trend data plots, continued with autocorrelation function (ACF) pair plots and partial autocorrelation function (PACF) using TEC value data. The actual and trend data plots are shown in Figure 3.

![Figure 2. The results of the box-cox transformation test TEC.](image2)

[Figure 2. The results of the box-cox transformation test TEC.]
Figure 3. Trend analysis plot of TEC value.

It is known that based on Figure 3 shows that the data is not stationary because it still shows an uptrend. Therefore, a differentiation process is needed to assist in stationary data. Differencing is the process of calculating the change or difference in actual data on the nth data which is reduced by the previous data\[7\]. The graph of the first difference trend plot can be seen in Figure 4.

Figure 4. Graph of the first difference between the TEC value.

Based on the data, the first difference is stationary. The next step is to recreate the ACF and PACF plots using the data in the differencing results.

Figure 5. ACF graph for first difference \((d = 1)\).

Figure 6. PACF graph for first difference \((d = 1)\).

Based on the results of model identification through ACF and PACF plots, it can be seen that the possibility of the emergence of a temporary model, ARIMA \((2,1,1),(2,1,0)\) and \((0,1,1)\) model. According to the possibility of the emergence of other temporary ARIMA model is \((1,1,1)\) and \((1,1,0)\) model \[8\]. So the researchers decided to use the ARIMA \((2,1,1),(2,1,0),(1,1,1),(1,1,0),(0,1,1)\) models.

3.2. Determining Models Parameter Estimate

The next step is to find the best parameter estimate, the following is testing the significance of the parameters. The resume of test results is presented in the following Table 1.
Table 1. Parameter estimates of the ARIMA model

| ARIMA Model | Parameter | Coefficient | p-value | t-test | Decision |
|-------------|-----------|-------------|---------|--------|----------|
| (2,1,1)     | $\emptyset_1$ | 0.2527      | 0.084   | 1.76   | NS       |
|             | $\emptyset_2$ | 0.1517      | 0.300   | 1.05   | NS       |
|             | $\theta_1$ | 0.9608      | 0.000   | 13.59  | S        |
| (2,1,0)     | $\emptyset_1$ | -0.5006     | 0.000   | -3.71  | S        |
|             | $\emptyset_2$ | -0.1103     | 0.428   | -0.80  | NS       |
| (1,1,1)     | $\emptyset_1$ | 0.2979      | 0.039   | 2.12   | NS       |
|             | $\theta_1$ | 0.9584      | 0.000   | 15.74  | S        |
| (1,1,0)     | $\emptyset_1$ | -0.4531     | 0.000   | -3.75  | S        |
|             | $\theta_1$ | 0.7732      | 0.000   | 9.01   | S        |

Note: S = Significant and NS = Not significant

The decision is $H_0$ rejected for ARIMA(0,1,1) and ARIMA(1,1,0) because the $p$-value < $\alpha = 0.05$.

3.3. Diagnostic Test

In the model diagnosis, the residual normality test and independence test will be carried out. The normal distribution test for the residues used the Kolmogorov-Smirnov test and the independence test (white noise) using the Ljung-Box test. The resume of test results is presented in the following table.

Table 2. Comparison of Models Based on Assumptions

| ARIMA Model | White Noise | Residual normality test |
|-------------|-------------|-------------------------|
| (2,1,1)     | fulfilled   | fulfilled               |
| (2,1,0)     | not fulfilled | fulfilled               |
| (1,1,1)     | fulfilled   | fulfilled               |

3.4. Best Model Selection

Based on the results of the parameter assessment and diagnostics, the ARIMA (2,1,1) and ARIMA (1,1,1) models are suitable for forecasting the TEC value because the parameters are significant and other assumptions are met. So to find the best model, ARIMA (2,1,1) and ARIMA (1,1,1) were selected using the mean square error (MSE) value.

Table 3. MSE value

| ARIMA Model | MSE |
|-------------|-----|
| (2,1,1)     | 2.480 |
| (1,1,1)     | 2.453 |

ARIMA model is chosen based on the smallest MSE value. So for forecasting the average value of the total electron content in March 2019 using the ARIMA (1,1,1).

3.5. Forecasting

The best ARIMA model is the ARIMA (1,1,1) model so it can be used to determine the average value of the total electron content in March 2019. Using Minitab 16 software, the forecast results are shown in Table 4.
Table 4. Forecast total electron content (TEC) during March 2019

| Period | Forecast | Period | Forecast |
|--------|----------|--------|----------|
| 59     | 17.3321  | 73     | 17.9972  |
| 60     | 17.4582  | 74     | 18.0358  |
| 61     | 17.5228  | 75     | 18.0744  |
| 62     | 17.5692  | 76     | 18.1130  |
| 63     | 17.6101  | 77     | 18.1517  |
| 64     | 17.6494  | 78     | 18.1903  |
| 65     | 17.6883  | 79     | 18.2289  |
| 66     | 17.7269  | 80     | 18.2675  |
| 67     | 17.7656  | 81     | 18.3061  |
| 68     | 17.8042  | 82     | 18.3447  |
| 69     | 17.8428  | 83     | 18.3833  |
| 70     | 17.8814  | 84     | 18.4219  |
| 71     | 17.9200  | 85     | 18.4605  |
| 72     | 17.9586  | 86     | 18.4991  |

The results of the forecast for the daily average value of TEC during March 2019 can be seen in the calculation results above which are located in the forecast column, starting from period 59 (day 59) to period 86 (day 86) daily average TEC value has increased every day in March 2019. The maximum daily average value of TEC occurs on the 86th day, which is 18.4991 TECU (1 TECU = 10^{16} \text{electrons/m}^2).

4. Conclusions

Based on the results of the research and discussion, it can be concluded that the best ARIMA model to predict the daily average value of total electron content (TEC) is the ARIMA model (1,1,1) because it is based on the ARIMA model diagnosis (1,1,1) fulfilling the three tests, the parameter significance test, the residual normality test (normal distribution), and the independence test (white noise). Also, this model has a smaller mean square error (MSE) value than the other models so that the ARIMA model (1,1,1) is feasible to use in predicting TEC values. The prediction results of the daily average total electron content (TEC) value in the Ionosphere layer in March 2019 have increased every day. Starting at the 59th period (59 day) the TEC value was 17.3321 TECU and the maximum TEC value occurred in the 86th period (86th day) which was 18.4991 TECU (1 TECU = 10^{16} \text{electrons/m}^2).

References

[1] Martiningrum D R 2012 Ketidakteraturan Lapisan Ionosfer dan Kaitannya dengan Proses-proses Kopling Atmosfer Ionosfer Indonesia Jurnal Sains Dirgantara 9(2) 147-148.

[2] Martiningrum D R, Purwono A, Nuraeni F and Muhamad J 2012 Fenomena Cuaca Antariks (Bandung: Pusat Sains Antariksa).

[3] Cahyadi M N, Rahadyan A N and Muslim Budan 2019 Analisis Nilai Total Electron Content (TEC) Near Real Time Menggunakan Data GPS Dua Frekuensi (Study Kasus Surabaya). Journal of Geodesy and Geomatics 14(2) 56-60.

[4] Hanke J E and Wichern D W 2005 Business Forecasting Eight Edition (New Jersey).

[5] Nurjanah I S, Ruhiat D, Andiani D 2018 Implementasi Model Autoregresive Moving Average (ARIMA) untuk Peramalan Jumlah Penumpang Kereta Api di Pulau Sumatra Jurnal Teorema: Teori dan Riset Matematika. 3(2)145-156
[6] Lilipaly G S, Hatidja D and Kekenusa J S 2014 Prediksi Harga Saham PT BRI Tbk Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average) Jurnal Sains Ilmiah 14(2) 60-66.

[7] Makridakis 1999 Metode dan Aplikasi Peramalan Jilid 1 Edisi ke 2 Translated Untung Sus Andriyanto and Abdul Baasit Jakarta: Erlangga.

[8] Ruhiat D and Effendi A 2018 Pengaruh Faktor Musiman pada Pemodelan Deret Waktu untuk Peramalan Debit Sungai dengan Metode SARIMA Jurnal Teorema: Teori dan Riset Matematika 2(2) 117-128.

[9] Wei W W S 2006 Time Series Analysis Univariate and Multivariate Methods Second Edition (New York: Pearson Education).