7th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Time series analysis of Turkish National Sea Level Monitoring System (TUDES) level data for Amasra Station

Ahsen Çelen*[1], Yasemin Şişman [1]

[1]Ondokuz Mayıs University, Faculty of Engineering, Department of Geomatics Engineering, Samsun, Türkiye

Keywords
Time series analysis
Sea level
TUDES
Minitab

Abstract

The observation and prediction of sea level are crucial for various reasons including the vertical datum determination, crustal movement forecasting, oceanographic modeling, and coastal infrastructure planning. In Turkey, a sea level monitoring system has been established by the General Directorate of Mapping and aims to measure sea level. Through the Turkish National Sea Level Monitoring System (TUDES), sea level is monitored using data collected at 20 tide gauge stations at 15-minute intervals. Time series analysis is considered a highly suitable modeling and forecasting method for data that is periodically measured. In this study, time series analysis models including ARIMA, SARIMA, and Holt-Winter's methods were applied using data from the Amasra tide gauge station within the TUDES for the year 2019. Additionally, a prediction for January 2020 at the same station was performed. The results were compared with the measured tide gauge data to assess the performance of the models. Evaluation criteria included the Mean Absolute Percentage Error (MAPE) for the Holt-Winter's method and the corrected Akaike Information Criteria (AICc) for the ARIMA and SARIMA models. The SARIMA (3,0,0) (0,2,2) model with an AICc value of -1307.83, indicating a seasonality of 12, was observed to be the best-performing model.

1. Introduction

The main objectives of geodesy are to define the shape and size of the earth and to obtain data on the spatial information of points (Vanícek and Krakiwsky, 2015). Due to the inherent impracticality of directly performing mathematical calculations for the Earth's shape, various reference surfaces are employed to acquire positional information. Reference surfaces define the parameters necessary for the mathematical representation of geometric and physical quantities (Drewes, 2009). Depending on the scope and purpose of the study, different reference surfaces such as the sphere, ellipsoid, and geoid can be selected (Jekeli, 2016).

The geoid is an assumed equipotential water surface that extends beneath the continents (Sansò & Sideris, 2013). This equipotential surface used for vertical referencing can be determined through the long-term measurements of the average sea level. The average sea level is defined as the vertical datum (Altamimi et al., 2010). Therefore, the significance and analysis requirement of sea level measurements emerge.

The sensitivity of satellite data over the past few decades, is approximately +3mm per year (Cazenave et al., 2014). For purpose of tracking that change, there is global cooperation to measure sea level. The Intergovernmental Oceanographic Commission (IOC), a subsidiary of UNESCO, addresses this issue on a global scale. The need for long-term monitoring of sea level changes with globally distributed tide gauge stations led IOC to establish the Global Sea Level Observing System (GLOSS). The contact organization of the system in Turkey is the General Directorate of Mapping (https://tudes.harita.gov.tr/).

To extract reliable information from data sets requiring long-term observations of sea level, statistical analysis is necessary. Time series analysis, a type of statistical analysis, is a powerful option for examining sea level data. Time series analyses allow for understanding the stochastic mechanisms of the measured data and gaining insights into future predictions based on past data (Cryer and Chan, 2008).

In this study, the time series analysis in sea level data for the year 2019 at the Amasra tide gauge station was examined using ARIMA, SARIMA, and Holt-Winter's time series analysis methods, and forecasting were made for January 2020. The obtained forecasted values were compared with the actual data, and the best model was observed to be SARIMA (3,0,0) (0,2,2) with a seasonality of 12.

* **Corresponding Author**

*(e-mail) ORCID ID xxxx – xxxx – xxxx – xxxx
(e-mail) ORCID ID xxxx – xxxx – xxxx – xxxx

Cite this study

Çelen, A., & Şişman, Y. (2023). Temporal Analysis of Turkish National Sea Level Monitoring System (TUDES) Level Data (Amasra Example). Intercontinental Geoinformation Days (IGD), 7, 90-93, Peshawar, Pakistan

2. Material and Methods

2.1. Material

The task of monitoring sea level in Turkey is conducted under the General Directorate of Mapping through the Turkish National Sea Level Monitoring System (TUDES) system, there are 20 GNSS-integrated radar sensor tide gauge stations distributed along the coasts of Turkey and the Turkish Republic of Northern Cyprus, adhering to GLOSS standards. These stations record measurements at 15-minute intervals, capturing not only sea level but also meteorological parameters affecting sea level changes, such as atmospheric pressure, wind speed, humidity, and temperature (https://tudes.harita.gov.tr/).

For the purposes of this study, TUDES data was provided by the General Directorate of Mapping, and the sea level data for the Amasra tide gauge station was accessed through the website https://tudes.harita.gov.tr/.

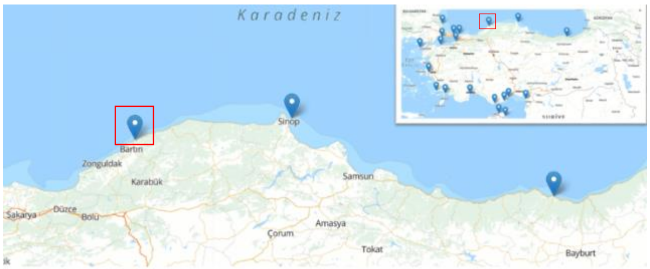

Figure 1. Study area

In the year 2019, a total of 34,921 observation units were obtained for the Amasra tide gauge station. To organize and these data, the code snippet was written using the Python programming language, which calculates daily averages for each day: The organized data was examined for general statistical information using the Minitab program, and tests for normality and outliers were conducted.

2.2. Method

Time series analysis examines the statistical distributions of periodic data within a specific time interval and consists of Autoregressive (AR) and Moving Average (MA) models. In AR models, the dependent variable is considered as a function of its past values. In the AR(p) model, the Zt value is represented as a linear function of the weighted sum of the series' past p values and error terms, as shown in the Equation 1.

$$Zt = \mu + \phi 1\ Zt\text{-}1 + \phi 2\ Zt\text{-}2 + \ldots + \phi p\ Zt\text{-}p + Zt \quad (1)$$

In this equation, Zt-1, Zt-2, .., Zt-p represent past observed values, μ represents the mean, Zt represents the error term, and φ1, φ2, ..., φp represent the coefficients of past observations. The goal in the model is to obtain the model order that makes the sum of squared errors zero and determine the unknown coefficients (Kara, 2009).

In the MA method, the aim is to reduce the effects of momentary, erroneous, and outlier data on the overall data. There are various types of moving average (MA) methods, such as Simple, Cumulative, Weighted, and Exponential. The equation 2 for the MA method is represented as:

$$Zt = \mu + \alpha t - \theta 1\ \alpha\ t\text{-}1 - \ldots - \theta q\ \alpha t\text{-}q \quad (2)$$

Here, θ1, ..., θq represents the coefficients of error terms, and αt, α t-1, ..., αt-q represent the error terms. The right side of the equation is expressed in terms of a meaningful q number of errors. The error term in the equation has a mean of zero and a constant variance (Kara, 2009).

2.3.1. ARIMA

It is a method used for performing univariate time series analysis and forecasting, also known as Box-Jenkins models. It represents an integrated model that incorporates operations such as MA, autocorrelation, and differencing. In the model expressed as ARIMA (p, d, q), p denotes the degree of the autoregressive (AR) model, d represents the differencing operation, and q indicates the degree of the MA model (Cryer, 1986). The model is represented as shown in Equation 3.

$$y_t = \alpha_0 + \sum_{t=1}^{p} \alpha_t (y_{t-1} - \mu) + \varepsilon_t \quad (3)$$

Here, α_0 and α_t represent autoregressive parameters to be estimated, and ε_t represents the random errors with zero mean and finite variances.

2.3.2. SARIMA

For time series data that exhibit seasonality and are non-stationary, ARIMA models often do not yield satisfactory results. Therefore, SARIMA models, which account for seasonality, are employed. In SARIMA models, denoted as SARIMA (p, d, q) (P, D, Q), in addition to the parameters used in ARIMA (p, d, q), there are additional parameters P, D, and Q that represent the seasonal AR order, differencing operation, and seasonal moving average order. These models take into consideration both the non-seasonal and seasonal components, offering a more comprehensive approach to time series modeling (Shumway and Stoffer, 2017).

2.3.3. Holt-Winter's

The Holt-Winter's method is one of the exponential smoothing techniques that involves a three-equation structure, accounting for level, trend, and seasonality. The seasonal equation can be formulated in two ways: multiplicative when trend and seasonality move together, and additive when they do not. (Hafid and Al-maamary, 2011). The model is represented as shown in Equation 4.

Level: $L_t = \alpha\frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + m_{t-1})$;
Trend: $m_t = \beta(L_t - L_{t-1}) + (1-\beta)m_{t-1}$
Seasonal: $S_t(t) = \gamma\frac{Y_t}{L_t} + (1-\gamma)\, S_{t-s}(t)$ (4)
Forecast: $F_{t+\tau} = (L_t + m_t q)\, S_{t-s}(t)$

Here; α, β and γ are smoothing constants, t is the time period, Y_t is the actual observed values, s is the length of seasonality, L_t is the level component, m_t is the trend component, S_t is the seasonal component and $F_{t+\tau}$ is the forecast for τ periods ahead.

3. Application and Results

For all modeling, the 2021 version of the Minitab program was employed. The model evaluation criterion is based on AICc. AICc is essentially AIC with an extra penalty term for the number of parameters. The smaller AIC is, the better the model fits the data (Minitab, 2021). The AIC is an information-theoretic indicator rooted in Kullback-Leibler Divergence, primarily assessing the information loss incurred by a given model. Consequently, the AIC criterion operates on the premise that the less information a model forfeits, the higher its quality (Kasali and Adeyemi, 2022). On the other hand, the BIC criteria are founded on Bayesian theory, with the goal of maximizing a model's posterior probability given the available data. The Bayesian information criterion (BIC) serves as a pivotal tool in the realm of statistics for model selection from a finite set of options. It maintains a close relationship with the Akaike information criteria and is partly reliant on the likelihood function (AIC) (Kasali and Adeyemi, 2022). Here are the AICc and BIC formulas (Minitab, 2021):

AIC = $2[(\rho + 1) - L_c]$; $L_c(y_i\, \mu_i\, \Phi) = \sum_{i=1}^{n} l_i$
$l_i = \ln(f(y_i, \widehat{\mu_i}, \Phi))$; $y_i \ln(\widehat{\mu_i}) + (m_i -)\ln(1 - \widehat{\mu_i})$

Here; p: the regression degrees of freedom; L_c: the log-likelihood of the current model; y_i: the number of events for the i^{th} row; m_i: the number of trials for the i^{th} row; Φ: 1, for binomial models; $\widehat{\mu_i}$: the estimated mean response of the ith row

AICc = $-2\ln(\text{Likelihood}) + 2p + \frac{2p(p+1)}{n-p-1}$
AICC is not calculated when $n - p - 1 \leq 0$
BIC = $-2\ln(\text{Likelihood}) + p\ln(n)$

The ARIMA model that does not account for seasonality was tested. The optimal parameters for the model were calculated with the assistance of the program, resulting in ARIMA (2,0,2) (Figure 3).

The SARIMA model takes seasonality into account, all combinations of the following values were tested: "3, 4, 12" for seasonality; "0, 1, 2" for differencing; "0, 1, 2" for seasonal differencing. According to the AICc criterion, the models that provided the best results for SARIMA (0,1,0) (1,2,3), SARIMA (0,0,2) (3,2,0), SARIMA (0,1,0) (1,2,3), SARIMA (3,0,0) (0,2,2) and SARIMA (1,2,2) (3,1,0) were given Figure 4a and 4b and Figure 4c according to seasonality " 3 4 and 12"

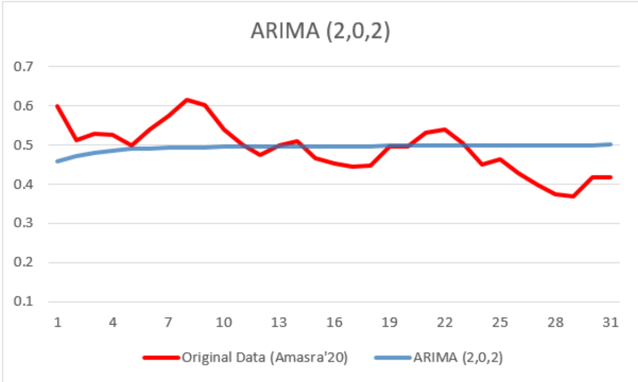

Figure 3. ARIMA (2,0,2) model

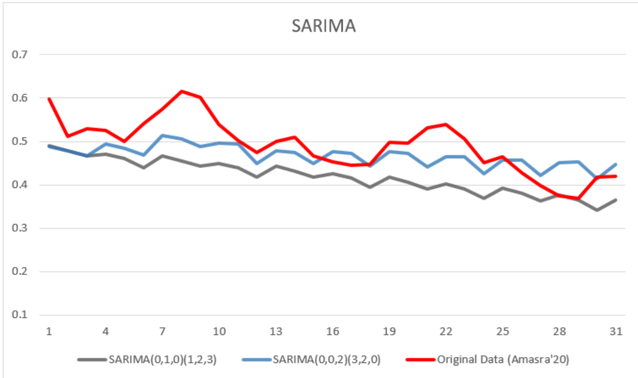

Figure 4a. SARIMA (0,1,0) (1,2,3) and SARIMA (0,0,2) (3,2,0) models

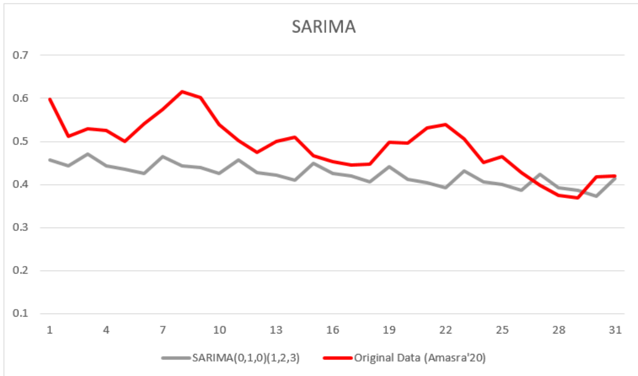

Figure 4b. SARIMA (0,1,0)(1,2,3) model

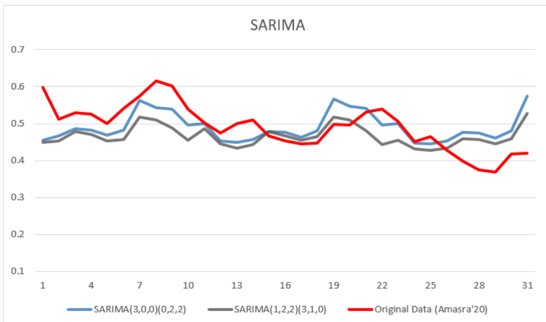

Figure 4b. SARIMA (3,0,0) (0,2,2) and SARIMA (1,2,2) (3,1,0) models

The model performance summaries of ARIMA and SARIMA models were made according to Mean Square Error (MSD), AICc and BIC values, (Table 1).

Table 1. Model Summaries

Model	MSD (Mean Sq. Dev.)	AICc (-)	BIC (-)
ARIMA(2,0,2)	0.0007982	1553.29	1530.13
SARIMA(0,1,0)(1,2,3)$_3$	0.0010915	1389.50	1370.26
SARIMA(0,0,2)(3,2,0)$_3$	0.0010598	1407.44	1384.37
SARIMA(0,1,0)(1,2,3)$_4$	0.0009654	1420.31	1401.10
SARIMA(3,0,0)(0,2,2)$_{12}$	0.0010689	1307.83	1285.09
SARIMA(1,2,2)(3,1,0)$_{12}$	0.0010199	1365.45	1338.75

Here what the abbreviations represent:
MSD: Mean Square Deviation

Finally, the Holt-Winter's method was applied to the data. Sequentially, combinations of α, β, and γ parameters ranging from "0.1 to 0.9" were tested for seasonality values of "3, 4, and 12". The best result was obtained with a seasonality of "4" and α, β, γ parameters set to "0.4", which was adopted in the additive model.

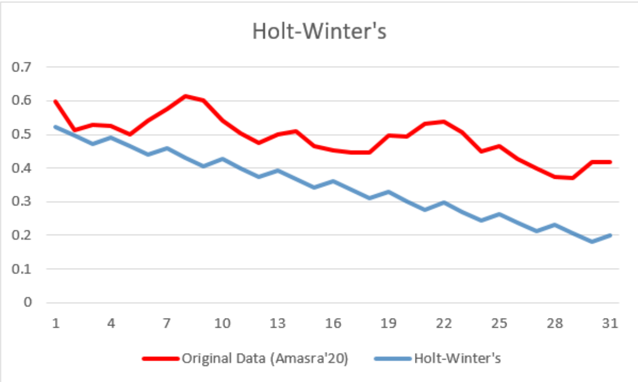

Figure 5. Holt-Winter's model

The obtained outputs to evaluate the model are as follows:

Table 1. Holt-Winter's model accuracy measures

Measures	MAPE	MAD	MSD
Values	6.29203	0.03136	0.00161

MAPE: Mean Abs. Per. Error; MAD: Mean Abs. Dev.

4. Conclusion

In the scope of this study, time series analysis models, including ARIMA, SARIMA, and Holt-Winter's methods, were applied using the 2019 data from the Amasra tide gauge station within the TUDES system. Furthermore, forecasting was made for the same station for the month of January 2020. The obtained results were compared with the measured tide gauge data, and the model's performance was assessed. Evaluation criteria included the MSD for the Holt-Winter's method and the AICc for the ARIMA and SARIMA models. The best model observed was the SARIMA (3,0,0) (0,2,2) model with an AICc value of "-1307.83", indicating a seasonality of "12". And finally, the MSD value of SARIMA (3,0,0) (0,2,2)$_{12}$ method was compared with the MSD value of the Holt Winter's method, revealing that the SARIMA model with the value of "0.0010689" outperformed the Holt-Winter's method with the value of "0.00161".

At the light of these explanation and applications it is said that the SARIMA (3,0,0) (0,2,2)$_{12}$ model is more suitable for these sea level data.

Acknowledgement

The authors thank the General Directorate of Mapping for providing the sea level data.

References

Al-Hafid, Majed, S. (2012). Short term electrical load forecasting using holt-winters method. Al-Rafidain Engineering, 20. 15-22. https://doi.org/10.33899/rengj.2012.63377.

Blewitt, G., Altamimi, Z., Davis, J., Gross, R., Kuo, C. Y., Lemoine, F. G., ... & Zerbini, S. (2010). Geodetic observations and global reference frame contributions to understanding sea-level rise and variability. Understanding sea-level rise and variability, 256-284.

Cryer, J. D. (1986). Time Series Analysis.

Cryer, J. D., & Chan, K. S. (2008). Time Series Analysis with Applications in R (2nd ed.). New York: Springer. http://dx.doi.org/10.1007/978-0-387-75959-3

Drewes, H. (2009). Reference Systems, Reference Frames, and the Geodetic Datum - Basic Considerations. 3-9.

Jekeli, C. (2016). Geometric Reference Systems in Geodesy. 2016 Edition. Ohio State University. USA.

Kara, T. (2009). Sabit GPS istasyonlarında zaman serileri analizi. Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya.

Minitab, LLC. (2021). Minitab. Retrieved from https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical- modeling/doe/how-to/factorial/analyze-binary-response/methods-and-formulas/model-summary/

Sansò, F., & Sideris, M. (2013). Geoid Determination: Theory and Methods. 10.1007/978-3-540-74700-0.

Shumway, R. H. & Stoffer, D. S. (2017) Time Series Analysis and Its Applications: With R Examples. Fourth Edition, Springer Texts in Statistics, Springer, Cham.

Torge, W., Müller, J., & Pail, R. (2023). Geodesy.

Türkiye Ulusal Deniz Seviyesi İzleme Sistemi. (t.y.). Deniz seviyesi gözlemleri. Harita Genel Müdürlüğü. https://tudes.harita.gov.tr/Portal /Index/32?lang=tr/Deniz%20Seviyesi%20G%C3% Gözlemleri.

Vanícek, P., Krakiwsky, E.J. (2015). Geodesy: The concepts (Revised 2. Edition). North-Holland, Amsterdam. Elsevier.