BEU Journal of Science
10 (3), 841-848, 2021

Araştırma Makalesi / Research Article

Assessment of tourist arrival from Russian to Antalya using the univariate time series methods

Hatice ONCEL CEKİM[1*], Ahmet KOYUNCU[2]

[1]*Department of Statistics, Hacettepe University, Beytepe 06800 Ankara, Turkey*
[2]*NETAŞ, Üniversiteler Mah. İhsan Doğramacı Blv. ODTÜ/Teknokent 06531 Çankaya, Ankara, Turkey*
(ORCID: 0000-0001-8538-6296) (ORCID: 0000-0002-1492-2191)

Abstract
With the continuous growth of Antalya tourism, the need for more accurate tourism forecasts emerge and the forecast performance is evaluated according to time series methods. Seasonal fluctuations are the most important feature of the tourism series and this feature makes it a suitable environment for comparing the forecast productivity of different models. In this study, the data of tourists coming from Russia to Antalya from 2007 to 2018 are used. The parametric and nonparametric univariate time series techniques, ARIMA, ETS, Combination (or Hybrid) and SSA, are compared in forecasting tourism demand. As a result of this paper, it is understood that the nonparametric SSA technique is more accomplished with respect to the accuracy of the obtained forecasts.

Keywords: Tourism, forecast, time series methods, univariate models.

Tek değişkenli zaman serileri yöntemlerini kullanarak Rusya'dan Antalya'ya gelen turistlerin değerlendirilmesi

Öz
Antalya turizminin sürekli büyümesiyle birlikte, daha doğru turizm öngörülerine duyulan ihtiyaç ortaya çıkmakta ve öngörü performansı zaman serisi yöntemlerine göre değerlendirilmektir. Mevsimsel dalgalanmalar turizm serilerinin en önemli özelliğidir ve bu özelliği onu farklı modellerin öngörü performanslarını karşılaştırmak için uygun bir ortam haline getirmektedir. Bu çalışmada, 2007-2018 yılları arasında Rusya'dan Antalya'ya gelen turistlerin verileri kullanılmaktadır. Turizm talebinin öngörüsünde parametrik ve parametrik olmayan tek değişkenli zaman serisi teknikleri, ARIMA, ETS, Kombinasyon (veya Hibrit) ve SSA, karşılaştırılmaktadır. Bu çalışma sonucunda, elde edilen tahminlerin doğruluğu açısından parametrik olmayan SSA yönteminin daha başarılı olduğu anlaşılmaktadır.

Anahtar kelimeler: Turizm, öngörü, zaman serisi yöntemleri, tek değişkenli modeller.

1. Introduction

Turkey is located in tourist arrivals in the first ten in the world [1]. This situation is reflected significantly to the economy in Turkey. While its total contribution to GDP is 11.6%, it is noteworthy that one out of every 10 employees works in the tourism sector [2]. The sea tourism in Turkey is surrounded by seas on three sides is of great importance. Antalya is the first city with the highest number of tourist visits in sea tourism. Russia has been the country that has sent the most visitors to Antalya for many years. However, the increase of Russian tourists, which has continued with an increasing trend for years, has started a sudden decline in 2016 due to Russia's banning of charter flights to Antalya. Therefore, Antalya tourism has lost its most valuable tourist resource until 2015. In the light of the information obtained, it is easily seen that Russian tourists have a very important role in Antalya's economy.

Antalya is an attractive region for international tourists with its many features such as the elevated temperature, sunny days, golf resorts, blue flag beaches, waterfalls and caves. Russian tourists

*Corresponding author: oncelhatice@hacettepe.edu.tr
Received: 28.04.2021, Accepted: 21.06.2021

generate the most important part of foreigner tourists in Antalya. Therefore, making the forecast of Russian tourists coming to Antalya will be beneficial in many ways. There are many different time series methods for making time-dependent future predictions. The ARIMA is one of the well-known models used in these cases. Therefore, the researches done with this model in tourism are also quite a number. In this context, Kim et al. [3] provide the interval of the tourist arrivals via the ARIMA model. Moreover, ARIMA methods are used by the benchmark model in the literature. Similarly, the study of Diaz and Nadal [4] compares tourism on the accuracy of the forecast by using the benchmark models as ARIMA and autoregressive neural network. Error, Trend and Seasonal (ETS), which is a member of the ARIMA, is one of the prominent univariate methods. According to the papers, the method can make a consistent forecast [5-7]. Lim and McAleer [8] find the best model concerning the forecast accuracy between the varied level differences used in ETS methods. Furthermore, a new approach and technique in these methods are proposed by Hyndman et al. [7] and Bergmeir et al. [9]. Singular spectrum analysis (SSA) analyzes and predicts non-stationary and complicated time series without an assumption, contrary to other classical methods [8]. Many studies indicated that SSA is a powerful forecast technique. Therefore, researches of the different issues such as genetic [9], hydrological [10], geophysical [11], tourism [12], business and finance [8] are preferred in time series. The combination method, which is another frequently preferred method, brings a lot of independents models together. The key point in this method is to find out how to combine the models in order to produce the best output model. Thus, every study establishes its own combination. Firstly, Bates and Granger [15] associate with the Brown's exponential smoothing and ARIMA models to forecast. Shen et al. [16] investigate the international demands for tourism using the six combination models. Besides, those and other studies in the literature show that the weight choices used are substantial as well as the selection of the combined models [13], [14, 15, 16].

In this study, the tourist arrivals to Antalya from Russia are obtained through the TURKSTAT [19]. The data is monthly and comprises the years between 2007 and 2018. The data divide Feb-2007 and Dec-2017 as training data, twelve months of 2018 as testing data to utilize the forecast trueness. The use of the training data is made to predict and testing data to survey the testification of the model prediction in this paper. Also, the tourist visits of bilateral relations between Turkey and Russia to examine the impact on tourism is taken into account until 2018. With this aim, the time series models are implemented for the future prediction of tourist visits in Antalya. The paper structure continues as follows: Section 2 contains considered methods, Section 3 extends the analysis results, and the final section highlights implications, limitations, and future study suggestions.

2. Material and Methods

2.1. ARIMA Method

To briefly describe the method in general terms, ARIMA models predict the future values using its past values. The ARIMA model is shown as ARIMA(p, d, q)(P, D, Q)$_s$. Here, p and P symbolize the seasonal and non-seasonal notation of the autoregressive model, q and Q represent the seasonal and non-seasonal notation of the moving average model, d and D mean the seasonal and non-seasonal notation of differencing, respectively and s is the period number. The general model equation is defined as follows:

$$\Upsilon_s^D \Upsilon^d \Lambda_P(B^s)\lambda_p(B)Y_t = \Pi_P(B^s)\pi_p(B)\xi_t \tag{1}$$

where $\Pi_P(B^s)$ and $\Lambda_P(B^s)$ are the moving average and seasonal autoregressive equations, $\pi_p(B)$ and $\lambda_p(B)$ are the non-seasonal moving average and autoregressive equations, $\Upsilon_s^D = (1 - B^s)^D$ and $\Upsilon^d = (1 - B)^d$ are the seasonal and non-seasonal differencing parameters, ξ_t and Y_t are the error and time series, respectively. B is the backward transaction $(BY_t = Y_{t-1})$, where Y_t is the time series. The application process of this method are succeeded, as summarized by authors in [20]. Detailed information about the ARIMA method, which is frequently used in analyzing time series, is included in many studies such as Preez and Witt [21], Prideaux [22], Kim et al. [3], Beneki and Silva [23], Oliveira and Oliveira [24] and Thamanukornsri and Tiensuwan [25].

2.2. ETS (Error, Trend and Seasonal) Method

The method was improved by the authors in [26] as an original method to exponential smoothing, formed from a different perspective for the ARIMA method. It has different models for error, trend, and seasonal equations. For the model equations, one can see [27]. It is one of the highly preferred models within the time-series studies in which univariate models are compared with Lim and McAleer [28], Hyndman and Khandakar [20], Hassani et al. [28], Naim and Mahara [29]. Therefore, the authors in [27] proposes a state space framework with the help of the ETS technique.

2.3. Singular Spectrum Method (SSA)

This is a method that predicts the future values by decomposing the time series into components by the embedding and singular value decomposition steps at first, then reconstructing these components with the grouping and diagonal averaging steps. This method attributes by integrating the mathematical analysis and visualisation tools [17]. SSA is a nonparametric method without satisfying any assumptions. For this reason, the process is not affected when the time series is stationary or non-stationary. Briefly, this method aims at producing the predicted series with minor error and without any assumptions. More information on the mathematical analysis is explained with a detailed description in the study of Hassani [30].

2.4. Combination (Hybrid) Method

The primary purpose of the combination method (which can be defined as the hybrid model) is to obtain a low mean square and increase the forecast accuracy. Based on this, the best model is chosen among the developed combinations. There are six model combinations of the ARIMA, ETS, Theta, NNETAR (Neural Network), STLM (Seasonal and Trend decomposition using Loess) and TBATS (Box-Cox transform, ARMA errors, Trend, and Seasonal components) used by the Hybrid package. It is significant that the weights (a_j) of these models determine the averages. Let $Y = [Y_1, Y_2, \dots, Y_N]^T$ be the vector of the time series. Assume that $\hat{Y}^j = \left[Y_1^j, Y_2^j, \dots, Y_N^j\right]^T$ be the forecast series when forecasting these series with s varied models $(j = 1,2, \dots, s)$. The simple linear model is described by

$$\hat{Y}_k^j = \sum_{i=1}^{s} a_i \hat{Y}_k^i. \tag{2}$$

The weights are added to the model to diminish their weakness and to improve forecast accuracy. This aim, some model combination techniques, which are widely preferred, can be mentioned as the simple average [18], [19], the trimmed average [14], the variance-based model [31], the Winsorized average [32], the outperformance method [33], the error-based combining [14], the median-based combining [34].

2.5. Methods of evaluation criteria

There are some measurement values that we can evaluate the methods in themselves to choose the method that best forecasts the series of Russian tourists visiting Antalya. For this reason, we calculate the root mean square error (RMSE) values of the considered models in this study to find the best forecasting method. The RMSE formula is defined as follows [29]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \left(y_t - \hat{y}_{t|t-1}\right)^2}, \tag{3}$$

where y_t and $\hat{y}_{t|t-1}$ symbolize, respectively, t^{th}observation and its forecasted unit depending previous unit $(t = 1, \dots, N)$.

3. Results

In this study, we use the R program for the estimating the tourist visits with "forecastHybrid", "Rssa", "forecast" and "vars" package. The arrival of Russian tourists, which has continued with increases and decreases at certain periods for years, can be seen very clearly in Figure 1, the great decrease experienced in 2016. This sudden drop in the series is expected to significantly affect the estimates of the models due to the different procedures. The logarithm of the series is used to get rid of this deterministic trend.

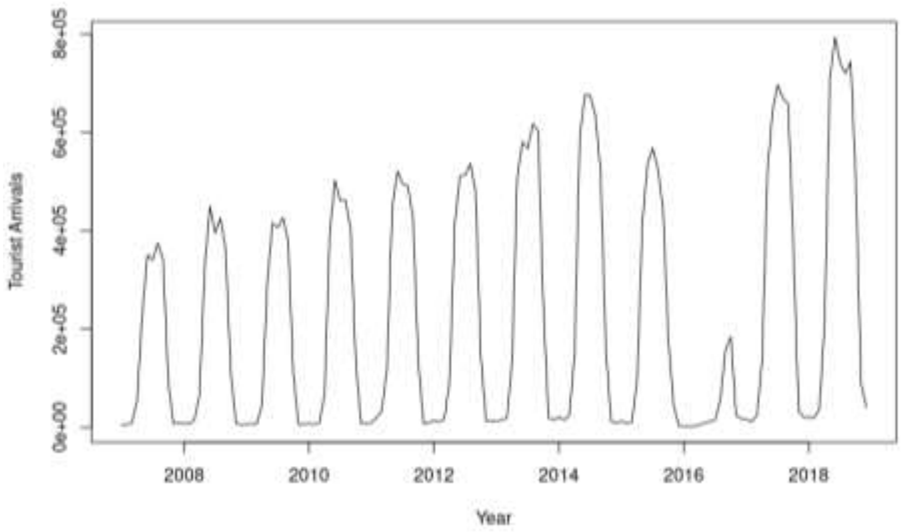

Figure 1. Time series graphs of series

The accuracy performance of the methods is investigated in this research compared to the RMSE values. According to RMSE in Figure 2, the SSA method has the smallest RMSE value. The errors of three models (the ARIMA, ETS and Hybrid) are normally distributed and the white noise. The SSA method requires no assumption checks due to it is a nonparametric test.

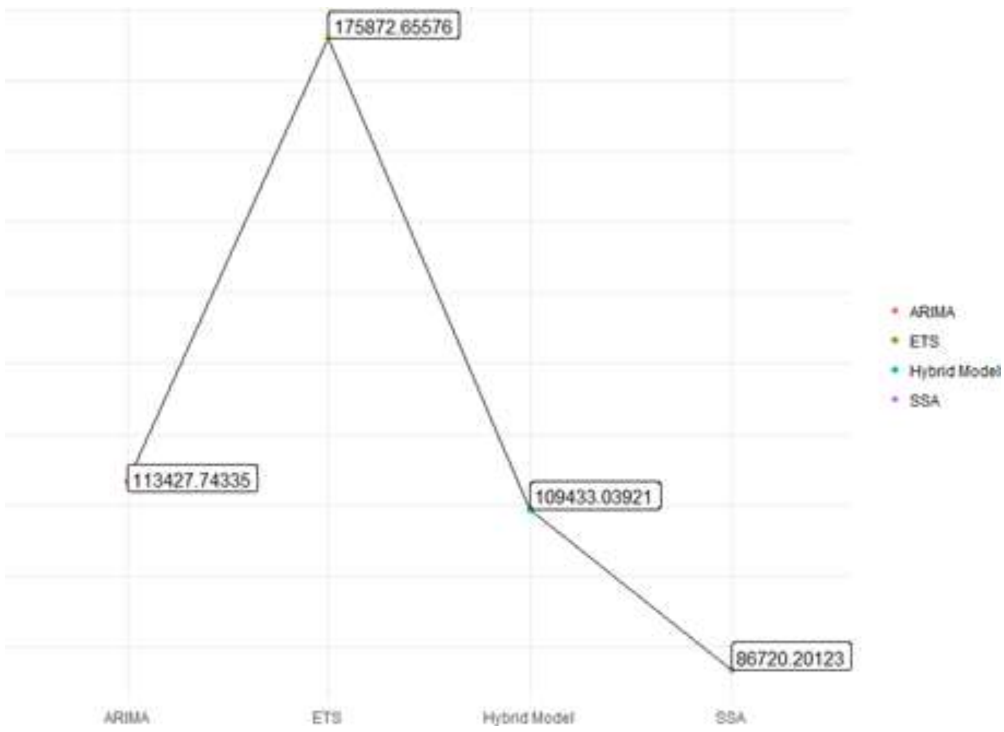

Figure 2. RMSE score of time series models

Table 1. Real tourist numbers in 2018 vs predictions of tourist arrivals through ARIMA, ETS, Hybrid, SSA methods

Months	Real	ARIMA	ETS	Hybrid	SSA
1	20452	22877	24095	23522	50966
2	20134	18651	21853	20282	38555
3	37446	29501	32840	31061	46262
4	215360	144679	175338	159175	238410
5	708263	507735	733316	602894	899958
6	793401	633691	959864	759128	715032
7	739544	679255	969457	789153	842357
8	719907	866132	1116476	951856	850501
9	743661	982569	1094314	1004028	666728
10	495893	491886	424064	442772	398752
11	88278	51937	36391	42879	42795
12	40568	27371	24390	25090	19888

Table 1 shows the forecasts and real values for 2018 obtained from the study's interested time series methods. Based on these results in Table 1, the prediction values have reached the best values by the different methods for each month in 12 months. This section explores the best time series methods to explain the series that does not have a fixed trend.

Table 2. RMSE values of each method for different horizons

Models	ARIMA	ETS	Hybrid	SSA
h=1	2424.97	3643.34	3089.19	30513.96
h=3	4871.52	3532.86	4114.82	21198.72
h=6	108616.90	70683.33	51238.26	86379.00
h=12	113427.70	175872.70	109433.00	86720.20

For different horizons, RMSE values are calculated in all methods in Table 2. When the results in Table 2 are examined, it is seen that the best estimation in the short term (h=1) forecast is obtained by ARIMA method. In medium term (h=3 and h=6) forecasting, the Hybrid model provides better estimates than other methods. The SSA method is stand out in 12-month long-term forecasts with the smallest RMSE value.

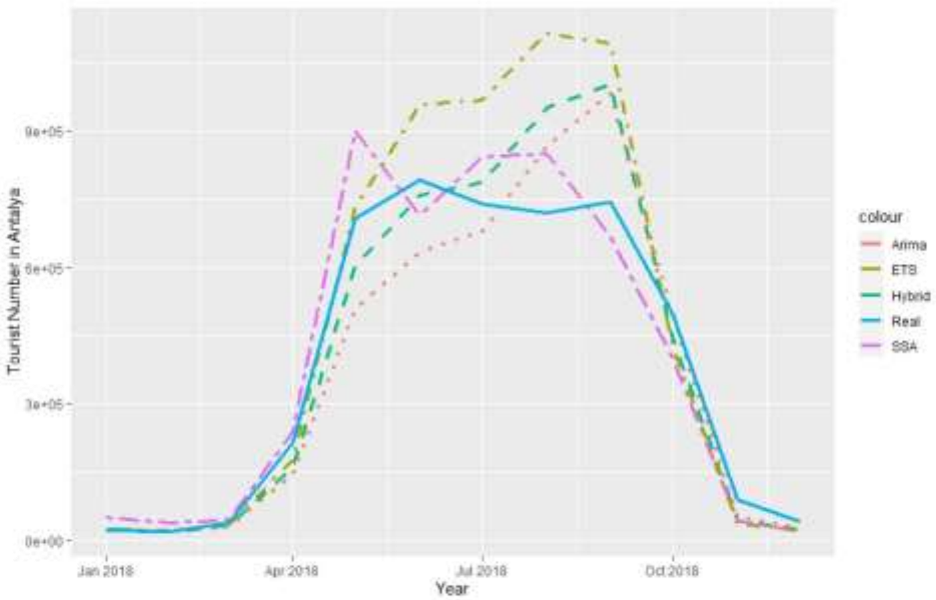

Figure 3. Graph of the actual and forecast values for 2018

Using the values given in Table 1, the forecast values obtained ARIMA, ETS, Hybrid and SSA models are described in Figure 3 to visually see the compatibility with the original series more clearly. When Figure 3 is examined in detail, we can see that the closest future prediction to the series is by the SSA method.

4. Conclusion

The number of Russian tourists coming to Antalya, which is the favorite holiday destination of tourists in many respects, continued to increase until 2015. While in Turkey for many years and cooperation in bilateral relation with Russia, they were faced with a serious crisis hardship. However, the process was restored in the second half of 2016 in line with the mutual request of the two publics. This situation has significantly affected Antalya tourism and we can observe this very clearly in Figure 1. This study is conducted to see which time series method gives better results in series with sudden ups and downs.

ARIMA and ETS methods are successful parametric methods frequently applied in time series analysis. Besides, the Hybrid model is a mixed-method that includes the advantages of 6 different time series models. Furthermore, the nonparametric SSA method can be easily adapted to different time series structures. The number of Russian tourists coming to Antalya is estimated with these powerful time series methods taken into consideration in the study.

Compared with the considered models according to the RMSE scores in Figure 2, the SSA method outperformed others models. As expected from previous studies of Hassani et al. [10] [28], Claveria and Torra [35], the nonparametric SSA method provides more successful results than the other parametric tests. When the methods are evaluated in terms of horizons in Table 2, it is seen that ARIMA, Hybrid and SSA methods are the best methods in small, medium and large horizon forecasts, respectively.

The tourist demand series has similar fluctuations in this study after the covid-19 disease with restrictions and prohibitions in Turkey in March 2020. For the future studies of this series, it is planned to make analyzes using methods similar to the SSA method. However, there are two options to make these forecasts more reliable: waiting for the end of the disease and returning to the normal procedure or thinking that it will continue for many years.

Authors' Contributions

All authors participated equally to contribute to the research.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The author(s) declares that this study complies with Research and Publication Ethics

References

[1] World Travel & Tourism Council (WTTC), 2017. City travel & tourism. https://www.wttc.org/-/media/files/reports/special-and-periodic-reports. (access date: 25.09.2020).

[2] World Travel & Tourism Council (WTTC), 2018. Travel & Tourism. https://dossierturismo.files.wordpress.com/2018/03/wttc-global-economic-impact-and-issues-2018-eng.pdf. (access date: 15.10.2020).

[3] Kim J.H., Wong K., Athanasopoulos G., Liu S. 2011. Beyond point forecasting: evaluation of alternative prediction intervals for tourist arrivals. International Journal of Forecasting, 27 (3): 887-901.

[4] Alvarez-Diaz M., Rossello-Nadal J. 2010. Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. Tourism Economics, 16 (1): 153-168.

[5] De Livera A.M., Hyndman R.J., Snyder R.D. 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106 (496): 1513-1527.
[6] Dong Z., Yang D., Reindl T., Walsh W.M. 2013. Short-term solar irradiance forecasting using exponential smoothing state space model. Energy, 55: 1104-1113.
[7] Hyndman R.J., Koehler A.B., Snyder R.D., Grose S. 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18 (3), 439-454.
[8] Lim C., McAleer M. 2001. Forecasting tourist arrivals. Annals of Tourism Research, 28 (4): 965-977.
[9] Bergmeir C., Hyndman R.J., Benitez J.M. 2016. Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. International journal of forecasting, 32 (2): 303-312.
[10] Hassani H., Heravi S., Zhigljavsky A. 2009. Forecasting European industrial production with singular spectrum analysis. International journal of forecasting, 25 (1): 103-118.
[11] Hassani H., Ghodsi Z. 2015. A glance at the applications of singular spectrum analysis in gene expression data. Biomolecular detection and quantification, 4: 17-21.
[12] Khan M.A., Poskitt D.S. 2017. Forecasting stochastic processes using singular spectrum analysis: Aspects of the theory and application. International Journal of Forecasting, 33 (1): 199-213.
[13] Rekapalli R., Tiwari R.K. 2014. Windowed SSA (Singular Spectral Analysis) for Geophysical Time Series Analysis. Journal of Geological Resource and Engineering, 3: 167-173.
[14] Beneki C., Eeckels B., Leon C. 2012. Signal extraction and forecasting of the UK tourism income time series: A singular spectrum analysis approach. Journal of Forecasting, 31 (5): 391-400.
[15] Bates J.M., Granger C.W. 1969. The combination of forecasts. Journal of the Operational Research Society, 20 (4): 451-468.
[16] Shen S., Li G., Song H. 2011. Combination forecasts of international tourism demand. Annals of Tourism Research, 38 (1): 72-89.
[17] Lemke C., Gabrys B. 2010. Meta-learning for time series forecasting and forecast combination. Neurocomputing, 73 (10-12): 2006-2016.
[18] Wong K.K., Song H., Witt S.F., Wu D.C. 2007. Tourism forecasting: To combine or not to combine? Tourism management, 28 (4): 1068-1078.
[19] TURKSTAT, 2019. Tourism Statistics, TURKSTAT, https://data.tuik.gov.tr/Kategori/GetKategori?p=Egitim,-Kultur,-Spor-ve-Turizm-105. (access date: 05.09.2020).
[20] Hyndman R.J., Khandakar Y. 2008. Automatic time series forecasting: The forecast Package for R. Journal of Statistical Software, 27 (3): 1-22.
[21] Preez J., Witt S.F. 2003. Univariate versus multivariate time series forecasting: an application to international tourism demand. International Journal of Forecasting, 19: 435-451.
[22] Prideaux B. 2000. The role of the transport system in destination development. Tourism management, 21 (1): 53-63.
[23] Beneki C., Silva E.S. 2013. Analysing and forecasting european union energy data. International Journal of Energy and Statistics, 1 (2): 127-141.
[24] Oliveira E.M., Oliveira F.L. 2018. Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. Energy, 144: 776-788.
[25] Thamanukornsri C., Tiensuwan M. 2018. Applications of Box-Jenkins (Seasonal ARIMA) and GARCH models to dengue incidence in Thailand. Model Assisted Statistics and Applications, 13: 95-105.
[26] Makridakis S., Wheelwright S.C., Hyndman R.J. 2008. Forecasting methods and applications. John Wiley & Sons, New York, 1-656.
[27] Hyndman R.J., Athanasopoulos A.G. 2018. Forecasting: principles and practice. OTexts, Monash University, Australia, 1-384.
[28] Hassani H., Silva E.S., Antonakakis N., Filis G., Gupta R. 2017. Forecasting accuracy evaluation of tourist arrivals. Annals of Tourism Research, 63: 112-127.

[29] Naim I., Mahara T. 2018. Comparative Analysis of Univariate Forecasting Techniques for Industrial Natural Gas Consumption. International Journal of Image, Graphics & Signal Processing, 10 (5): 33-44.
[30] Hassani H. 2007. Singular spectrum analysis: methodology and comparison. Journal of Data Science, 5: 239-257.
[31] Newbold P., Granger C.W. 1974. Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society, 131-165.
[32] Jose V.R., Winkler R.L. 2008. Simple robust averages of forecasts: Some empirical results. International Journal of Forecasting, 24 (1): 163-169.
[33] Bunn D.W. 1975. A Bayesian approach to the linear combination of forecasts. Journal of the Operational Research Society, 26 (2): 325-329.
[34] Agnew C.E. 1985. Bayesian consensus forecasts of macroeconomic variables. Journal of Forecasting, 4 (4): 363-376.
[35] Claveria O., Torra S. 2014. Forecasting tourism demand to Catalonia: Neural networks vs. time series models. Economic Modelling, 36: 220-228.