Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi
Journal of Agricultural Faculty of Gaziosmanpasa University
http://ziraatdergi.gop.edu.tr/

Araştırma Makalesi/*Research Article*

JAFAG
ISSN: 1300-2910
E-ISSN: 2147-8848
(2017) 34 (3), 18-26
doi: **10.13002/jafag4298**

Forecasting Harvest Area and Production of Strawberry Using Time Series Analyses

Melekşen AKIN[1*] Sadiye PERAL EYDURAN[1]

[1]Igdir University, Agricultural Faculty, Department of Horticulture, Igdir
*e-mail: akinmeleksen@gmail.com

Alındığı tarih (Received): 20.04.2017 Kabul tarihi (Accepted): 16.09.2017
Online Baskı tarihi (Printed Online): 18.10.2017 Yazılı baskı tarihi (Printed): 29.12.2017

Abstract: This study was conducted to model the harvest area and production of strawberry in Turkey using FAOSTAT data from period of 1965 - 2015 to forecast strawberry harvest area and production for 2016-2025 period. Non-stationary time series of strawberry harvest area and production for 1965-2015 period were transformed into stationary time series after taking the first difference of the time series. Three Autoregressive Integrated Moving Average (ARIMA (1,1,0), ARIMA (1,1,1) and ARIMA (0,1,1)) and three Exponential Smoothing (Holt, Brown and Damped) models were used comparatively for time series data sets on strawberry harvest area and production. Holt exponential smoothing model showed the best forecasting and Brown exponential smoothing model was the most appropriate forecasting model for strawberry harvest area and production from the tested six models. We forecasted that the strawberry harvest area is going to be 14 385 ha in 2016 and will increase to 16 591 ha in 2025. The strawberry production forecasted significant increase for the 2016-2025 period, from 396 341 tons to 519 816 tons. Briefly, the present forecasting results might help policy makers to develop macro-level policies for food security and more effective strategies for better planning strawberry production in Turkey.

Keywords: Strawberry, production, harvested area, exponential smoothing, time series

Zaman Serisi Analiz Yöntemlerini Kullanarak Türkiye'deki Çilek Hasat Alanı ve Üretiminin Tahminlenmesi

Öz: Bu çalışma, 1965-2015 yılı FAOSTAT verilerini kullanarak 2016-2025 yılı için Türkiye'deki çilek hasat alanı ve üretimini tahminlemek amacıyla yapılmıştır. 1965-2015 dönemine ait çilek hasat alanı ve üretimi zaman serileri, zaman serilerinin birinci dereceden farkının alınmasıyla durağan hale getirilmiştir. Çilek hasat alanı ve üretimini modellemek için üç bütünleştirilmiş otoregresiv hareketli ortalama (ARIMA (0,1,1), ARIMA (1,1,0) ve ARIMA (1,1,1)) ve üç üstsel düzleştirme (Holt, Brown ve Damped) yöntemi kıyaslanmıştır. Çilek hasat alanı ve üretimini tahminlemede test edilen altı yöntemden Holt üstsel düzleştirme tekniği en iyi projeksiyonu gerçekleştirmiş olmasına rağmen, Brown modeli en uygun yöntem olarak öne çıkmıştır. Bu sonuçlar doğrultusunda, 2016 yılında 14 385 hektar olan çilek hasat alanının 2025 yılında 16 591 hektara yükseleceği; 2016 yılında 396 341 ton olan çilek üretim miktarının ise 2025 yılında 519 816 tona doğru artış göstereceği öngörülmüştür. Kısacası, bu çalışmadan elde edilen sonuçların gıda güvenliğini sağlamak için makro düzeyde politikaların geliştirilmesine ve Türkiye'deki çilek üretiminin daha etkin bir şekilde planlanmasına yardımcı olacağı düşünülmektedir.

Anahtar Kelimeler: Çilek, üretim miktarı, hasat, üstsel düzleştirme, zaman serileri

1. Introduction

Strawberries are very rich source of carotenoids, vitamins, phenols, and flavonoids which play an important role in human nutrition. Strawberries show very high antioxidant activity and have positive health effects. There is a high demand on berries due to their potential for reducing the risk of chronic diseases such as cancer, cardiovascular diseases, and stroke (Wang and Linn, 2000). Therefore, it is important to provide food sustainability for a healthy diet and healthy next generations. Turkey strawberry harvest area was 9 465 ha with a production of 130 000 tons in the 2000 year, but it showed an dramatic increase with 13 423 ha harvest area and 376 070 tons production in the 2014 year

(FAOSTAT, 2017). Increasing strawberry production area could also help develop rural areas by increasing farmer revenues. It is important to establish policies for increasing strawberry production for future sustainability, export incomes, and food safety. Therefore, projection studies are very useful tool for forecasting strawberry harvest area and production, as well as to determine appropriate policies for the future.

There are only a few studies on forecasting crop production for the next years (Masuda and Goldsmith, 2009; Semerci and Ozer, 2011; Suresh et al., 2012; Celik et al., 2013; Borkar, 2016; Celik et al., 2017; Karadas et al., 2017a, b) using ARIMA and exponential smoothing methods time series analysis. To our knowledge, strawberry harvest area and production of Turkey are not forecasted. Therefore, the main aim of this study was to model strawberry harvest area and production of Turkey using time series models for the period of 1965-2015 with the aim to forecast strawberry harvest area and production for the next period of 2016-2025. This study could help establish macro-level policies for food security and sustainability for the future.

2. Material and Methods

Time series data sets on annual harvest area (ha) and production for the 1965-2015 period were downloaded by FAOSTAT database to obtain annual forecasting values for the next 2016-2025 years. Three Autoregressive Integrated Moving Average (ARIMA (1,1,0), ARIMA (1,1,1) and ARIMA (0,1,1)) and three Exponential Smoothing (Holt, Brown and Damped) models were compared (Celik et al., 2017; Karadas et al., 2017a,b`).

To find the best one among the tested six candidate models, we used the following model fit statistics:

Root Mean Square Error,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}}$$

Mean Absolute Percentage Error,

$$MAPE = \frac{\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}}{n}$$

Maximum Absolute Percentage Error,

$$MaxAPE = \max_i\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right)*100, \quad i=1,2,\ldots,N$$

Mean Absolute Error (MAE),

$$MAE = \frac{1}{n}\sum_{i=1}^{M}\left|y_i - \hat{y}_i\right|$$

$$BIC = ln(\hat{\sigma}_{\varepsilon}^{2}) + kln(n)/n$$

Where $\hat{\sigma}_{\varepsilon}^{2}$ is the error variance.

Mean percentage error (MAPE) is defined as a measure of how much a dependent series changes from its model-predictive performance level. The root means square error (RMSE) has been employed as a model fit criterion in order to measure model performance. Maximum absolute percentage error (MaxAPE) represents the largest forecasted error, expressed as a percentage. Statistical analysis of annual harvest area (ha) and production was performed by IBM SPSS program (version 23).

3. Results and Discussion

3.1. Strawberry harvest area

In the present study, time series analysis was used to analyses strawberry harvest area from the 1965-2015 period in order to forecast harvest area for the next 2016-2025 years. Graph of the strawberry harvest area for the 1965-2015 period is given in Figure 1. The graph showed an increasing trend. Time series graphs of autocorrelation (ACF) and partial autocorrelation functions (PACF) are presented in Figure 2 to understand this trend better.

Several terms available in ACF graph exceeded confidence interval, which implies that there was a time series data trend (Fig. 2). To remove the time series from the trend and to make the data stationary, the first-degree difference of the time series was taken. ACF and PACF graphs for the first difference time series are given in Figure 3. The difference series were found stationary (Fig. 3).

Some ARIMA and exponential smoothing models such as stationary R^2, R^2, RMSE, MAPE, MAE, and BIC were evaluated as a goodness of fit criteria.

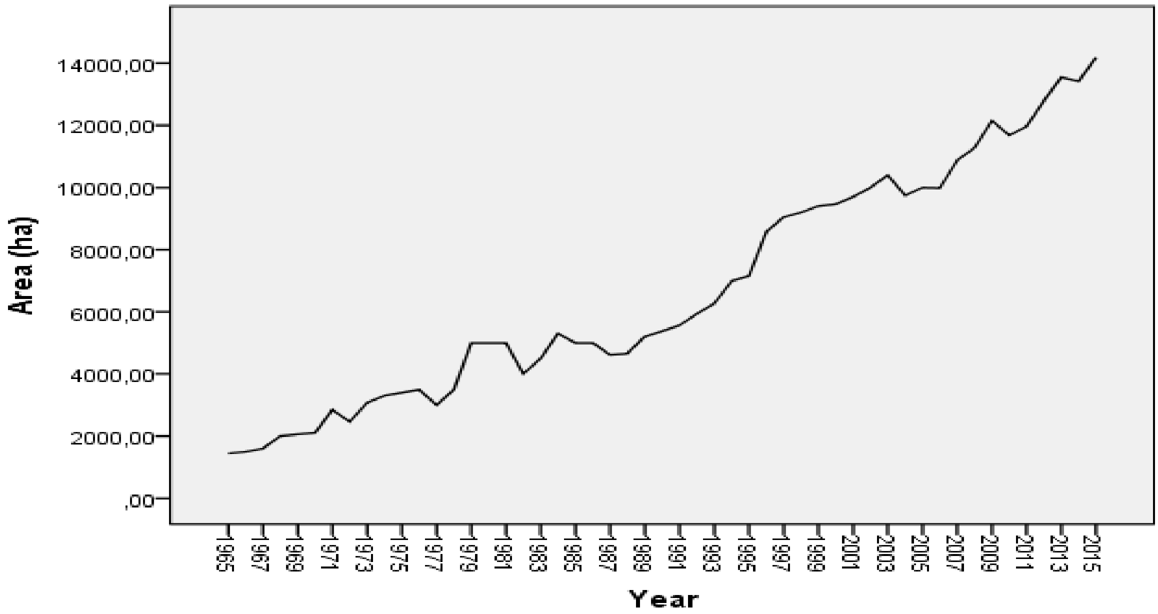

Figure 1. Graph of strawberry harvest area (ha) for the 1965-2015 period
Şekil 1. *1965-2015 yılları arasındaki döneme ait çilek hasat alanı (ha) grafiği*

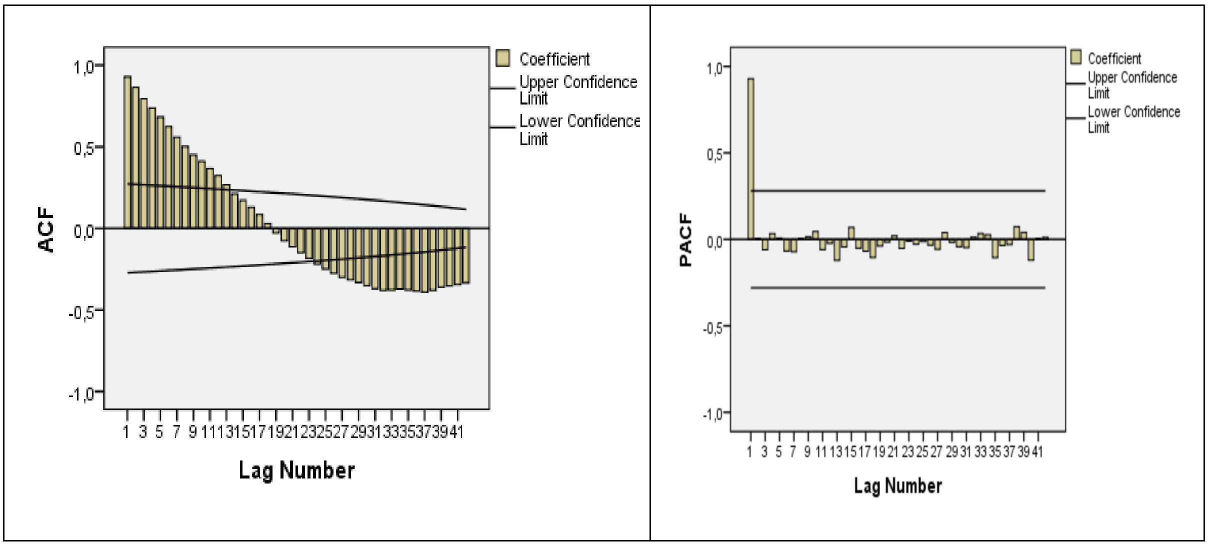

Figure 2. ACF and PACF graphs of strawberry harvest area (ha) for the 1965-2015 period
Şekil 2. *1965-2015 yılları arasındaki döneme ait çilek hasat alanı (ha) ACF ve PACF grafikleri*

The results are summarized in Table 1. Holt was the best model that showed the greatest R^2 and the smallest RMSE, MAPE, MAE, and BIC (Table 1). In addition, Pektas (2013) advised using BIC model as a fit criterion. Holt linear model can be used for Ljung-Box Q=25.411 and p=0.063>0.05.

Model parameters from exponential smoothing coefficients of the Holt model on harvest area are provided in Table 2. Parameter coefficients of Holt linear model were estimated as $\alpha = 0.900$ and $\gamma = 0.000$, and α coefficient was significant (P<0.01).

The relationship degrees of ACF and PACF, according to the Holt model, were within the confidence limits (Fig. 4).

Although 13th lag was found slightly higher than the confidence limit, it was not considered as a problem. Residual terms were white noise (Fig. 4). It was determined that the two series were in agreement with each other (Figure 5).

Forecasts for strawberry harvest area for 2016-2025 period are provided in Table 3. There was an increasing trend for the harvest area of 2016-2025 period, from 14 385 ha to 16 591 ha.

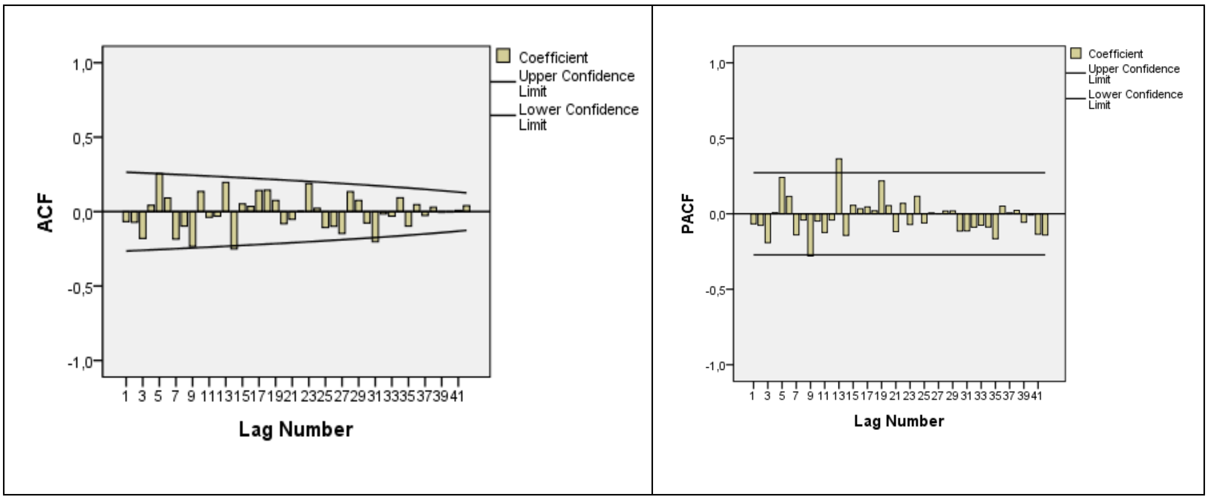

Figure 3. Strawberry harvest area ACF and PACF graphs of the first difference series
Şekil 3. *Çilek hasat alanı birinci derece fark serilerinin ACF ve PACF grafikleri*

Table 1. Strawberry harvest area model goodness of fit statistics
Çizelge 1. *Çilek hasat alanı uyum testi istatistikleri*

Fit Statistic	ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(0,1,1)	**Holt***	Brown	Damped
Stationary R^2	0.008	0.040	0.010	**0.539**	0.454	0.009
R^2	0.983	0.984	0.983	**0.984**	0.981	0.984
RMSE	479.357	476.387	478.785	**474.064**	513.737	479.063
MAPE	6.900	6.932	6.841	**6.702**	7.577	6.715
MAE	350.012	347.304	347.933	**340.823**	392.081	341.189
BIC	12.501	12.567	12.499	**12.477**	12.561	12.575

*Ljung-Box Q=25.411 and p=0.063>0.05

Table 2. Strawberry harvest area exponential smoothing model parameters
Çizelge 2. *Çilek hasat alanı üstsel düzleştirme model parametreleri*

	Estimate	SE	t	Sig.
Alpha (Level)	0.900	0.147	6.140	0.001
Gamma (Trend)	0.000	0.067	0.002	0.998

3.2. Strawberry production

The time series analysis was performed to forecast strawberry production for the 2016-2025 period. Annual production graph for the 1965-2015 period showed a trend (Fig. 6). Graphs of autocorrelation (ACF) and partial autocorrelation functions (PACF) were provided to understand this trend better (Fig. 7). Most of the terms available on the ACF graph exceeded the confidence interval, which implies that there was a time series data trend (Fig. 7). To eliminate the trend and make the data stationary, the first-degree difference of the time series was taken. ACF and PACF graphs for the first difference time series are shown in Figure 8.

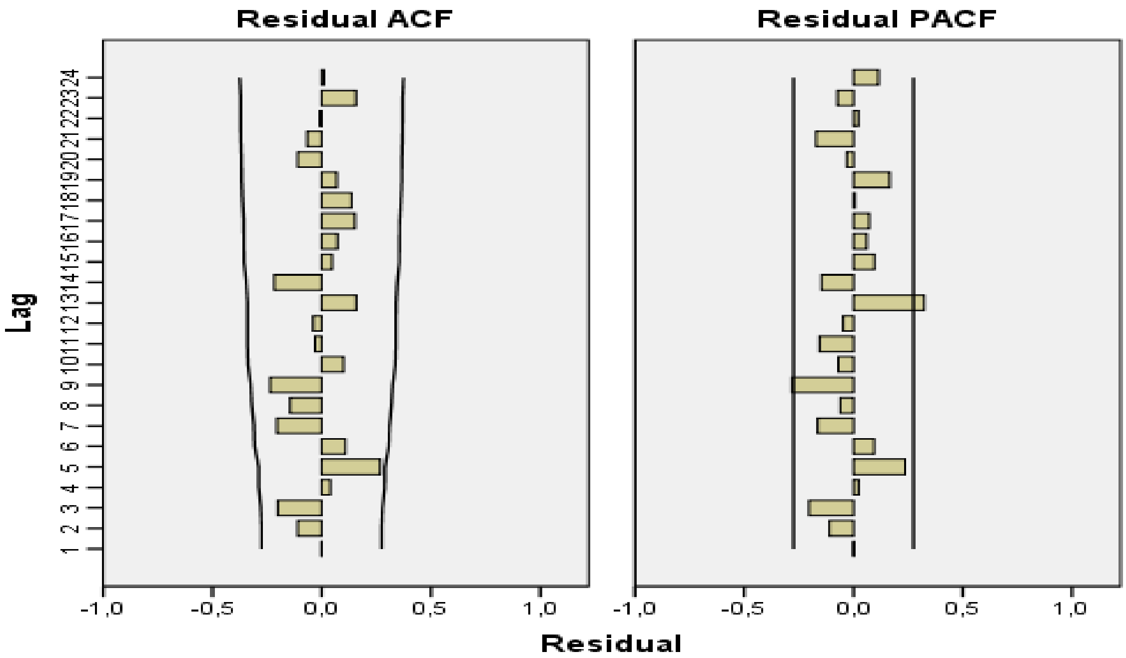

Figure 4. Strawberry harvest area ACF and PACF graphs of residuals
Şekil 4. *Çilek hasat alanı ACF ve PACF hata terimleri grafikleri*

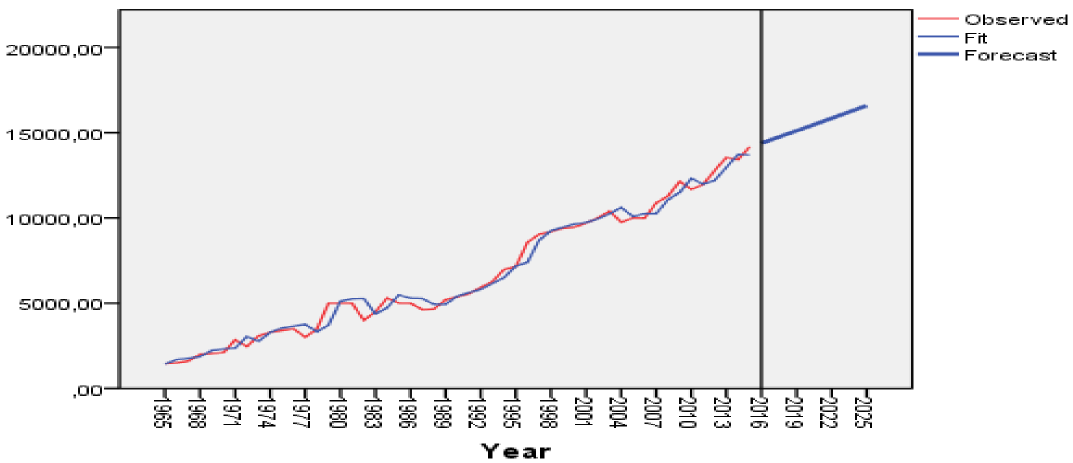

Figure 5. Strawberry harvest area graph of the observed and forecasted (fit) series
Şekil 5. *Çilek hasat alanı gözlenen ve tahminlenen serilerin grafikleri*

Some ARIMA and exponential smoothing models such as stationary R^2, R^2, RMSE, MAPE, MAE, and BIC were evaluated as a goodness of fit criteria. The results are summarized in Table 4. Brown linear exponential smoothing method was determined as the best method that had the greatest R^2 and the smallest BIC (Table 4). Brown linear model can be used for Ljung-Box Q=12.310 and p=0.791>0.05.

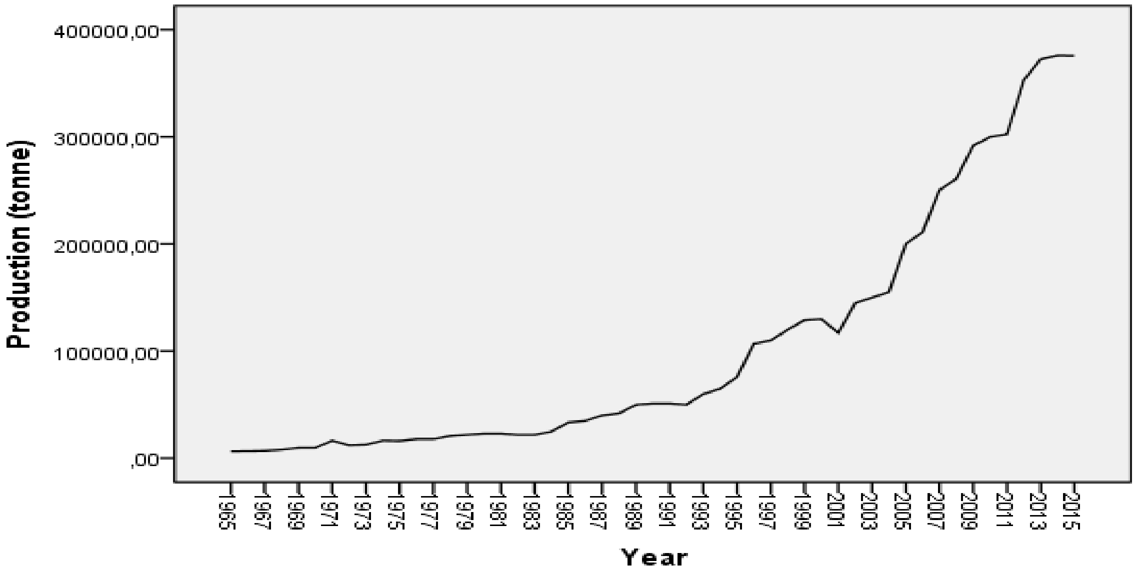

Figure 6. Graph of strawberry production (tons) for the 1965-2015 period
Şekil 6. *1965-2015 yılları arasındaki döneme ait çilek üretimi (ton) grafiği*

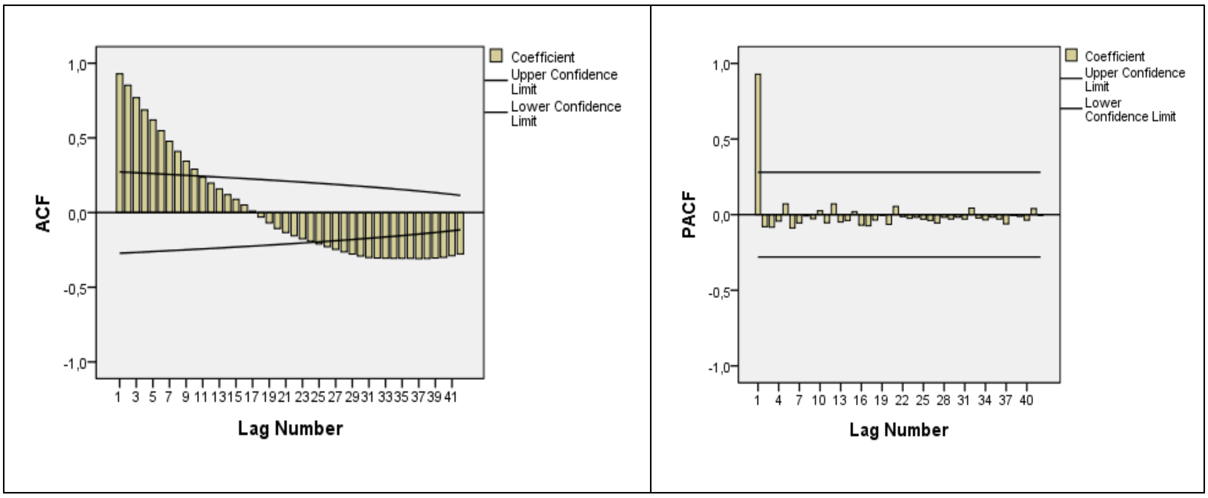

Figure 7. ACF and PACF graphs of strawberry production (tons) for the 1965-2015 period
Şekil 7. 1965-2015 yılları arasındaki döneme ait çilek üretim (ton) ACF ve PACF grafikleri

Table 3. Forecasting strawberry harvested area (ha) values for the 2016-2025 period
Tablo 3. *2016-2025 yılları çilek hasat alanı (ha) tahmini*

Year	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Area (ha)	14 385	14 630	14 875	15 121	15 366	15 611	15 856	16 101	16 346	16 591

Model parameters including smoothing coefficients belonging to the Brown model are presented in Table 5.

Parameter coefficient of the Brown linear model was estimated as α = 0.472 (P<0.01) (Table 5).

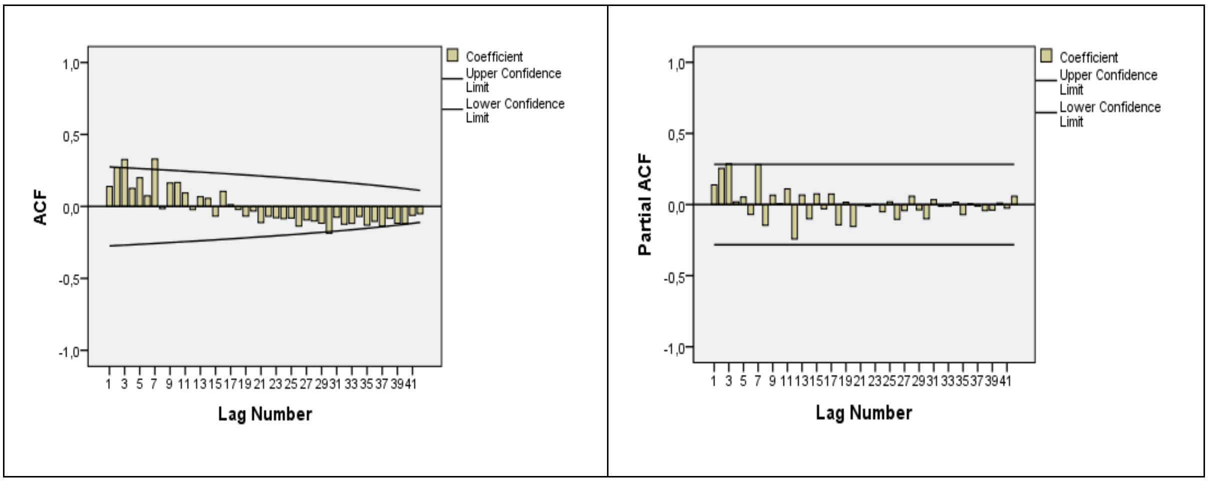

Figure 8. Strawberry production first difference series ACF and PACF graphs
Şekil 8. *Çilek üretimi birinci derece fark serilerinin ACF ve PACF grafikleri*

The relationship degrees of lags in ACF and PACF graphs of residuals according to the Brown model were within the confidence limits; therefore, residuals were white noises (Fig. 9). The observed series was in agreement with the series containing fitted values (Fig. 10).

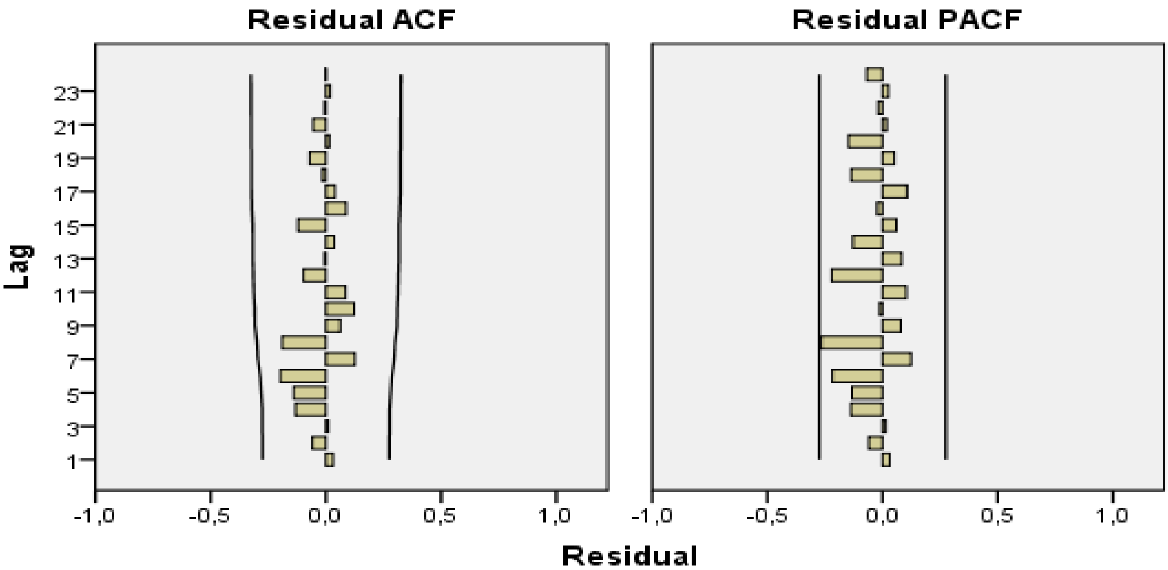

Figure 9. Strawberry production ACF and PACF graphs of residuals
Şekil 9. *Çilek üretimi ACF ve PACF hata terimleri grafikleri*

Table 6 presents forecasting results of strawberry production. It is estimated that the strawberry production from 2016 to 2025 would increase from 396 341 tons to 516 816 tons with the proportion of 130.4 %. To our knowledge, there is no research available on forecasting strawberry harvest area and production in Turkey for planning the security and sustainability of this crop for the future. Besides, there is a limited number of studies on predicting production of other crops (Semerci and Ozer, 2011; Masuda and Goldsmith, 2009; Celik, 2013; Amin et al., 2014; Borkar, 2016).

In the current study, Brown exponential smoothing model revealed that strawberry production increased significantly from 396 341 tons to 519 816 tons for the 2016-2025 period. Sunflower production increasing trend was also reported for the 2010-2013 period (Semerci and Ozer, 2011). The worldwide increasing soybean production trend was reported for the 2020-2030 period using Damped exponential smoothing method (Masuda and Goldsmith, 2009).

Production amounts of pistachios, walnuts, hazelnuts, almond and chestnuts in Turkey were forecasted for the 2012-2020 period and increasing production trends were noted using different ARIMA models (Celik, 2013).

Celik et al. (2017) used ARIMA (0,1,1) model to forecast annual groundnut production for the 2016-2030 period and reported increasing trend for the groundnut production. Holt exponential smoothing method showed increasing trend for sunflower and sesame production in Turkey for the 2016-2025 period. Soybean ranged from 162 878 tons to 179 784 tons, and sunflower ranged from 1 692 269 tons to 1 879 521 tons (Karadas et al. 2017a inpress). Karadas et al. (2017b in press) forecasted 102 310 tons increase in cotton lint production for the 2016-2026 period using Holt exponential smoothing method.

Although there are a few studies on forecasting some crop productions for the next years, prediction studies on economically important crops should be increased for developing better policies for food security and sustainability, as well as better production planning.

Table 4. Strawberry production amount model fit statistics
Tablo 4. *Çilek üretim miktarı uyum testi istatistikleri*

Fit Statistic	ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(0,1,1)	Holt	**Brown***	Damped
Stationary R^2	0.019	0.128	0.013	0.504	**0.494**	0.154
R^2	0.988	0.990	0.988	0.990	**0.990**	0.990
RMSE	12572	11980	12612	11561	**11571**	11681
MAPE	20.731	15.118	21.430	8.366	**9.057**	8.367
MAE	8184.633	7390.375	8213.205	6737.669	**7103.745**	6742.755
BIC	19.035	19.017	19.041	18.865	**18.790**	18.963

*Ljung-Box Q=12.310 and p=0.791>0.05

Table 5. Strawberry production exponential smoothing model parameters
Şekil 5. *Çilek üretimi üstel düzeltme modeli parametreleri*

	Estimate	SE	t	Sig.
Alpha (Level and Trend)	0.472	0.065	7.324	0.001

Table 6. Forecasting strawberry production for the 2016-2025 period
Tablo 6. *2016-2025 yılları çilek üretim (ton) tahmini*

Year	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Forecast (tons)	396 341	410 060	423 780	437 499	451 219	464 938	478 657	492 377	506 096	519 816

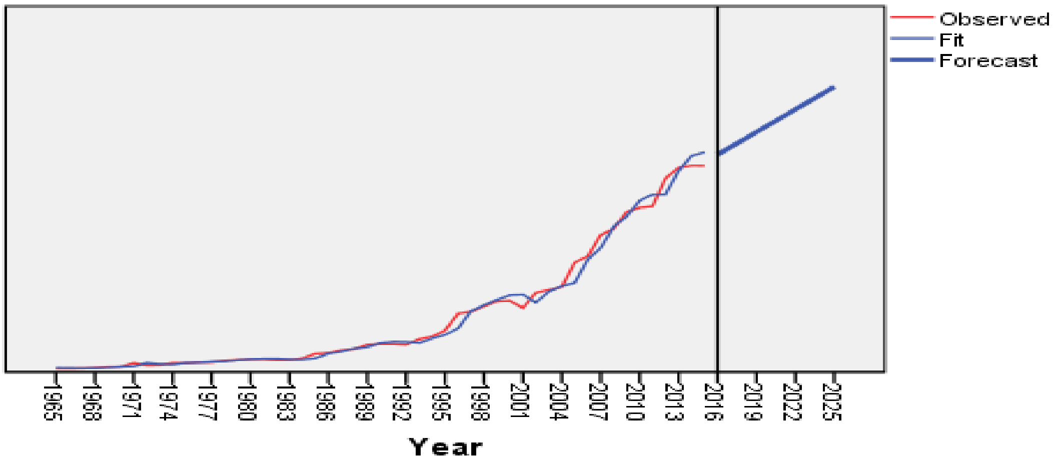

Figure 10. Strawberry production graph of the observed and forecasted (fit) series
Şekil 10. *Gözlenmiş ve tahmin edilmiş serilerin çilek üretim grafikleri*

4. Conclusion

Non-stationary time series of strawberry harvest area and production for the 1965-2015 period were transformed into stationary time series after taking the first difference of the time series. Holt exponential smoothing model was the best forecasting model, among the tested six models, for predicting strawberry harvest area for the next 2016-2025 period. However, Brown exponential smoothing model was described as the most suitable for forecasting strawberry production for the 2016-2025 period. Harvest area prediction ranged from 14 385 ha to 16 591 ha for the 2016-2025 period. Forecasting results of strawberry production for the 2016-2025 period showed a significant increase, from 396 341 tons to 519 816 tons. As a result, the present prediction results might help policymakers to develop macro-level policies for food security and more effective strategies for better planning strawberry production in Turkey.

References

Amin M, Amanullah M, Akbar A (2014). Time series modeling for forecasting wheat production of Pakistan, The Journal of Animal Plant Science, 24(5):1444-1451

Borkar P (2016). Modeling of groundnut production in India using ARIMA model. International Journal of Research in IT & Management. 6(3): 36-43.

Celik S (2013). Modelling of production amount of nuts fruit by using Box-Jenkins technique. Yuzuncu Yil J. Agr. Sci. 23(1):18-30.

Celik S, Karadas K, Eyduran E (2017). Forecasting groundnut production of Turkey via ARIMA models. The Journal of Animal and Plant Science.

FAOSTAT (2017). Statistical database of the food and agriculture organization of the United Nations. http://faostat.fao.org/.

Karadas K, Celik S, Eyduran E, Hopoglu S (2017a). Forecasting production of some oil seed crops in Turkey using exponential smoothing methods. The Journal of Animal and Plant Science (in press).

Karadas K, Celik S, Hopoglu S, Eyduran E, Iqbal F (2017b). A survey of the relationship between production amount, cultivation area and yield of cotton lint in Turkey using time series analysis. The Journal of Animal and Plant Science (in press).

Masuda T, Goldsmith PD (2009). World soybean production: area harvested, yield, and long-term projections. International Food and Agribusiness Management Association. 12(4): 143-161.

Pektas A (2013). SPSS ile veri madenciligi. Dikeyeksen Yayın Dagitim, Yazilim ve Egitim Hizmetleri San. ve Tic. Ltd. Sti.; Istanbul.

Semerci A, Ozer S (2011). Turkiye'de aycicegi ekim alanı, üretim miktarı ve verim değerinde olası değişimler. Journal of Tekirdag Agricultural Faculty, 8(3): 46-52.

Suresh K, Kiran R, Giridhar K, Sampath K (2012). Modelling and forecasting livestock feed resources in India using climate variables. Asian-Australasian Journal of Animal Science, 25(4): 462-470.

Wang SY, Lin HS (2000). Antioxidant activity in fruits and leaves of blackberry, raspberry and strawberry varies with cultivar and developmental stage. J. Agr. Food.