Research Article/Araştırma Makalesi

Iğdır Üniversitesi Tarım Bilimleri Dergisi

Agro-Science Journal of Iğdır University

List of contents accessible via [DergiPark](https://dergipark.org.tr)

e-ISSN: 3023-7017 **Volume:** 2 **Issue:** 1 **Year:** 2024

Forecasting Soybean Production in Türkiye: A Comparative Analysis of Automated and Traditional Methods

Hakan DUMAN[1*]

[1] Iğdır University, Iğdır Vocational School, Department of Transportation Services, 76000, Iğdır, Türkiye.

Keywords:

Soybean production
Türkiye
Time series forecasting
ARIMA algorithm
NNAR
Auto-ARIMA

ABSTRACT

Türkiye's climate and soil are well-suited for the cultivation of oilseed crops, which are of vital importance to various industries and human and animal diets. Among oilseeds, soybeans, a legume, possess a distinctive nutritional profile. While existing research covers soybean production in Türkiye, this study aims to: a) evaluate production levels using different forecasting algorithms to identify the most accurate model, and b) based on the chosen model, forecast future production and assess the current and future entrepreneurial potential of the soybean industry in Türkiye.

Soybean production data (1990-2022) from TURKSTAT was divided into training (n=25) and test (n=8) sets for cross-validation. By applying univariate time series methods, including ARIMA, SES, NNAR, MN, and Naive to the training dataset, it was found that ARIMA (1,1,1) performed best according to test set RMSE values. The performance ranking (in terms of RMSE) was as follows: ARIMA (13019) < SES (13888) < Naive (14240) < NNAR (58393) < MN (80418). Notably, for this dataset, the performance of automated processes was relatively worse than that of manual methods, suggesting that relying solely on automated methods may lead to suboptimal forecasting results. These findings underscore the importance of human oversight in the use of automated algorithms for time series forecasting and highlight the need for caution when employing automated methods.

The ARIMA (1,1,1) model predicts a flat trend in production from 2023 to 2032, with an initial production volume of 154 516 tonnes and a slight decline to 153 607 tonnes. This predicted stagnation implies that, in the context of economic and population growth, soybean production will fall further behind domestic demand, leading to increased import reliance. These findings are of serious importance to farmers and policymakers alike, as they can assist in the formulation of informed decisions pertaining to resource allocation, crop planning, and market strategies. Local producers may potentially benefit from increased production efficiency, improved competitiveness, and potential revenue growth by catering to both domestic and export markets. Furthermore, an understanding of these trade dynamics can assist stakeholders in identifying potential avenues for collaboration or investment within the Turkish soybean industry. Further analysis of these results is ongoing in order to gain deeper insights into the factors influencing soybean production trends in Türkiye.

**Corresponding Author/Sorumlu Yazar:* Hakan DUMAN (hakan.duman.acad@gmail.com)

Received: 02/06/2024 **Accepted:** 18/06/2024 **Published:** 30/06/2024

How to cite:

Duman, H., (2024). **Forecasting Soybean Production in Türkiye: A Comparative Analysis of Automated and Traditional Methods**. *Agro-Science Journal of Iğdır University*, 2(1), 90-102.

INTRODUCTION

Fats, known as triglycerides of fatty acids, are an important source of energy in the human diet and serve as crucial industrial raw materials. Due to the high cost and insufficiency of animal oil production, a large proportion (91.7%) of the oils needed for human nutrition is supplied by vegetable oils. Many crops contain oil in their seeds, with some of the most important oil-producing crops being annual plants such as soybean, sunflower, rapeseed, peanut, sesame, and safflower. Additionally, perennial crops like olives, dates, and coconuts also play a significant role in crude oil production (Arıoğlu, 2016).

Soybean, scientifically known as *Glycine max* L. Merrill, is a legume that holds significant importance due to its dual role as a source of protein and oil (Pagano and Miransari, 2016). It is the world's leading oil crop, with soybean oil valued not only for human consumption but also for various industrial applications beyond food preparations and animal nutrition (Pratap *et al.*, 2012; Tiwari, 2017). Soybean oil accounts for 53% of global oilseed production, making it an essential component in the agricultural systems of major countries such as the USA, China, Brazil, Argentina, and India (Pratap *et al.*, 2012). The majority of soybeans undergo industrial processing to create value-added products like soymeal for animal feed and edible oil. The strong connection between soybean farming and industries, particularly the food and feed industries, makes it a highly desirable global trading commodity (Tiwari, 2017).

Soybean cultivation in Türkiye began in the 1930s, primarily in the Black Sea Region. The introduction of the second crop project subsequently facilitated the expansion of soybean farming into the Mediterranean Region. Today, soybean ranks as the second most important crop in the Mediterranean Region in terms of both cultivated area and production volume. Notably, the Çukurova sub-region exhibits a particularly strong emphasis on soybean cultivation. Irrigation has significantly contributed to the successful establishment of soybeans in this region due to its ease of cultivation under irrigated conditions (Özcan, 2023). However, Türkiye's soybean trade pattern is characterized by a heavy reliance on imports. This is primarily due to insufficient domestic production to meet the high domestic consumption demands and market requirements for soybeans and soybean meal. Consequently, Türkiye is a net importer of soybeans, which is the country's most imported oilseed crop (Tüfekçi, 2019).

Considering the significance of soybean production within a national economy, numerous studies have been initiated with the primary focus on estimating soybean production levels, consumption patterns, and market trends. The outcomes of these analyses are expected to provide essential market intelligence for agricultural decision makers. This valuable information can help them anticipate potential opportunities and threats within the soybean market. Uçum (2016), for example, employed the ARIMA (1,1,1) model to analyze and forecast both current and future soybean production trends in Türkiye. Güler *et al.* (2017), compared the performance of ARIMA and Artificial Neural Networks (ANNs) in forecasting import quantities of various oilseed crops, including soybean, chickpea, sunflower, and rapeseed.

Turkey's climate and soil are well-suited for cultivating oilseed crops, vital for various industries and human and animal diets. Among oilseeds, soybeans, a legume, have a distinctive nutritional profile. While existing research explores soybean production in Turkey, this study aims to: (a) evaluate production levels using different forecasting algorithms to identify the most accurate model and (b) forecast future production based on the chosen model to assess the current and future entrepreneurial potential of Turkey's soybean industry. This study will evaluate and compare the performance of various soybean production projection models to identify the most accurate one. Utilizing the outputs of the most successful model, the study will then forecast soybean production for the next decade. This information will be a valuable resource for agricultural decision-makers, enabling them to anticipate potential market opportunities and threats.

MATERIALS AND METHODS

This study utilized datasets from FAOSTAT to determine the current status of soybean production (Anonymous, 2022a) and trade (Anonymous, 2022b; Anonymous, 2022c) worldwide in 2022. Additionally, a 33-year soybean production dataset (in tonnes) compiled by TurkStat (Anonymous, 2022d) for the period 1990-2022 was used. The TurkStat data was divided into training and testing sets, with the first 25 years used for model training and the remaining 8 years for model evaluation based on goodness-of-fit criteria. This process facilitated the identification of the most effective model.

Within the domain of time series forecasting, two prominent methodologies dominate: Exponential Smoothing (ES) and Autoregressive Integrated Moving Average (ARIMA) models. ES approaches target the trend and seasonality inherent in the data, while ARIMA models prioritize capturing the autocorrelations within the series. These distinct techniques offer complementary strengths, enabling effective forecasting solutions (Hyndman, 2021).

The seminal work by Box and Jenkins (1970) in their publication “Time Series Analysis: Forecasting and Control” marked a pivotal turning point in time series forecasting. Known as the Box–Jenkins (BJ) methodology or ARIMA methods, these techniques represent a shift from constructing single-equation or simultaneous-equation models towards analyzing the stochastic properties of time series data independently. In contrast to regression models where the dependent variable Y_t is explained by independent variables $X_1, X_2, X_3, \dots, X_k$ ARIMA models enable Y_t to be influenced by past (lagged) values of itself and random error terms (Gujarati and Porter, 2009).

The Box–Jenkins method for ARIMA model building is a widely-used technique but poses challenges when implemented on a large scale, requiring both expertise and substantial time investment (Mélard and Pasteels, 2000). In business environments where over a thousand product lines necessitate monthly forecasts and trained personnel are in short supply, automatic forecasting algorithms prove indispensable (Hyndman and Khandakar, 2008). These methods must identify suitable models, estimate parameters, generate predictions, accommodate unusual patterns, and function efficiently without user intervention for vast numbers of series. To make time series analysis accessible to individuals lacking the necessary expertise, software vendors have introduced automated time series forecasting methods as alternatives (Mélard and Pasteels, 2000). For instance, R’s “auto.arima” function in the “forecast” package determines all ARIMA model parameters automatically (Hyndman and Khandakar, 2008). This study employed the ‘auto.arima’ algorithm to identify the optimal ARIMA model based on the minimum Bayesian Information Criterion (BIC) score. The performance of the selected model was then compared to manually chosen ARIMA models.

ES techniques, including popular variants such as the Holt-Winters methods, trace their origins to the work of Robert G. Brown for the US Navy around 1944. Brown developed these methods to address trend and seasonality in discrete time series. Concurrently, Charles Holt was developing a different version of ES at the US Office of Naval Research (ONR). The work of these pioneers was further refined by the researchers, leading to the development of various statistical models for forecasting using exponential smoothing. The success of these methods in both forecasting and inventory control has inspired extensive research aimed at deriving equivalent point forecasts from other models. Many of these models are state space models that yield minimum mean squared error forecasts identical to those of simple ES (Hyndman *et al.*, 2008). In this study, we compare the performance of Simple Exponential Smoothing (SES), Holt’s Linear Trend Method (HLT) (Holt, 2004), and Damped Holt’s Trend Method (DHLT) (Gardner and Mckenzie, 1985).

This study utilized artificial neural networks (ANNs) for time series forecasting. ANN models, such as Neural Network Autoregression (NNAR), are employed for their ability to handle complex nonlinear relationships between variables. These networks use lagged time series values as inputs and have a single hidden layer with an automatic determination of nodes based on optimal performance. The

R package fits NNAR(p,k) models where p represents the number of lagged inputs and k is the number of nodes in the hidden layer. For non-seasonal data, the default is the optimal number of lags for a linear AR model. ANNs are iteratively applied to generate forecasts one step at a time using historical inputs as well as previous forecasts until all required forecasts have been computed (Hyndman *et al.,* 2008). In this study, while the ‘p’ value is determined by an automatic algorithm, the ‘k’ value is set to 5, 10, 25, and 50.

To establish a baseline for comparison with more sophisticated models, we included two simple forecasting methods: the mean method and the naive method. The mean method forecasts future values by setting them equal to the average of the historical data. Conversely, the naive method forecasts future values by setting them equal to the most recent observation in the time series.

Stationarity is a crucial concept in statistical analysis and modeling, particularly when working with time series data or performing econometric analyses. Its importance lies in the assumptions made by various analytical methods and models, which rely on stationary data to ensure accurate results. Stationarity refers to a property of a time series where its statistical characteristics, such as mean, variance, and covariance, remain constant over time. When dealing with non-stationary data, these properties may change over the observation period, leading to biased or misleading conclusions when applying certain analytical techniques (Gujarati and Porter, 2009).

In this study, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is employed to examine the stationarity of the soybean production time series. Kwiatkowski et al. (1992) introduced this approach to assess whether an observable time series follows a deterministic trend while being stationary around it. The proposed model includes deterministic trend, random walk, and stationary error components. The null hypothesis assumes the random walk has zero variance, which is tested using the LM test statistic. Asymptotic distributions under both the null and alternative hypotheses (difference-stationarity) are derived.

Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are commonly employed metrics to evaluate the accuracy of time series forecasting models. RMSE measures the difference between observed and predicted values, while MAE calculates the absolute differences between them. MAPE represents the average percentage error across all observations. Equations for these goodness-of-fit criteria (Hyndman *et al*., 2008; Akın *et al*., 2021):

- Forecast Error (FE): $\epsilon_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$ where ϵ_t is the forecast horizon, y is the observation and $\hat{y}$ is the forecast. The training data is given by $\{y_1, y_2, \dots, y_T\}$ and the test data is given by $\{y_{t+1}, y_{T+2}, \cdots, y_{T+H}\}$.
- Root Mean Square Error (RMSE): $\frac{1}{n}\sqrt{\sum(\epsilon_t{}^2)}$ where ϵ_t is the forecast error.
- Mean Absolute Error (MAE): $\frac{1}{n}\sum|\epsilon_t|$ where ϵ_t is the forecast error.
- Mean Absolute Percentage Error (MAPE): $\frac{1}{n}\sum|P_t|$ where P_t is the percentage error for the t-th prediction ($P_t = 100 \times \frac{\epsilon_t}{y_t}$)

In this study, RMSE was used as the primary goodness-of-fit criterion to assess model performance; however, MAE and MAPE were also reported for a comprehensive evaluation of the forecasting models.

This study used the R statistical environment, version 4.2.2, developed by R Core Team (2022). The tidyverse meta-package, version 2.0.0, created by Wickham *et al.* (2019), was employed for data manipulation and cleaning. For time series data extension, the tsibble package (version 1.1.3), developed by Wang *et al.* (2020), was utilized. To build forecasting models, the fable package (version 0.3.3) created by O’Hara-Wild *et al.* (2023a) was employed. For feature extraction and statistical analysis, the feasts package (version 0.3.1), developed by O’Hara-Wild *et al.* (2023b), was utilized. To create world maps, rnaturalearth version 0.3.4 by Massicotte and South (2023), rnaturalearthdata version 0.1.0 by South (2017), sf package version 1.0.14 by Pebesma (2018), and sp package version 2.1.2 with contributions from Pebesma and Bivand (2005) and Bivand *et al.* (2013), were employed.

RESULTS AND DISCUSSION

This study begins by examining the production and trade dynamics of soybeans. Soybeans are a vital component in various industries, including food processing, animal feed production, and industrial applications. Understanding worldwide trends in soybean production and trade can provide valuable insights into market conditions, price fluctuations, potential investment opportunities, threats, and strategic partnerships.

According to data from the year 2022 (Anonymous, 2022a), Brazil is the leading producer of soybeans, with an impressive volume of approximately 120.7 million tons. The United States of America follows closely behind, with an estimated production of approximately 116.4 million tonnes during the same year. Argentina ranks third with a reported production of approximately 43.9 million tonnes, while China contributes approximately 20.3 million tonnes to the global market. India is the fifth-largest producer of soybeans, with an estimated annual production of approximately 12.6 million tons. Conversely, Türkiye's reported production volume was approximately 155,000 tonnes, placing it in the lower part of the ranking at position 32 (see Figure 1).

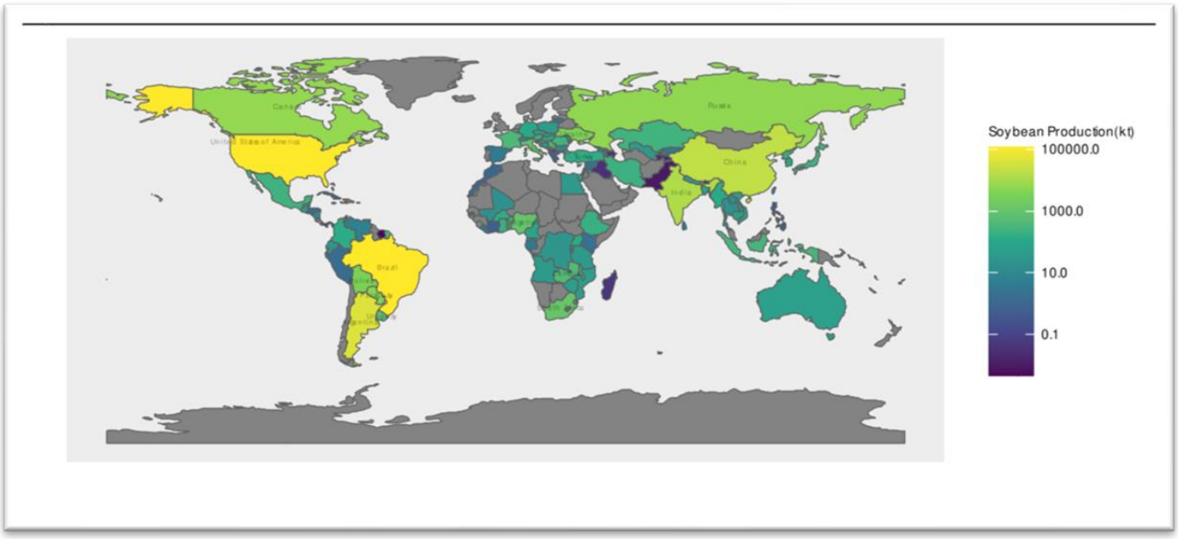

Figure 1. Soybeans Production by Countries (2022) (Data source: Anonymous (2022a)).

As indicated by the data (Anonymous, 2022b), China occupies the leading position as the largest importer of soybeans, with approximately 91.1 million tonnes imported during that period. The Netherlands follows with approximately 4.0 million tonnes imported, while Mexico ranks third with a reported import volume of approximately 3.9 million tonnes. Japan is in fourth place with estimated annual soybean imports of approximately 3.5 million tonnes, while Germany holds the fifth position with approximately 3.4 million tonnes imported. Türkiye's reported soybean imports during this period were approximately 3 million tonnes, placing it ninth in this comparison (Figure 2).

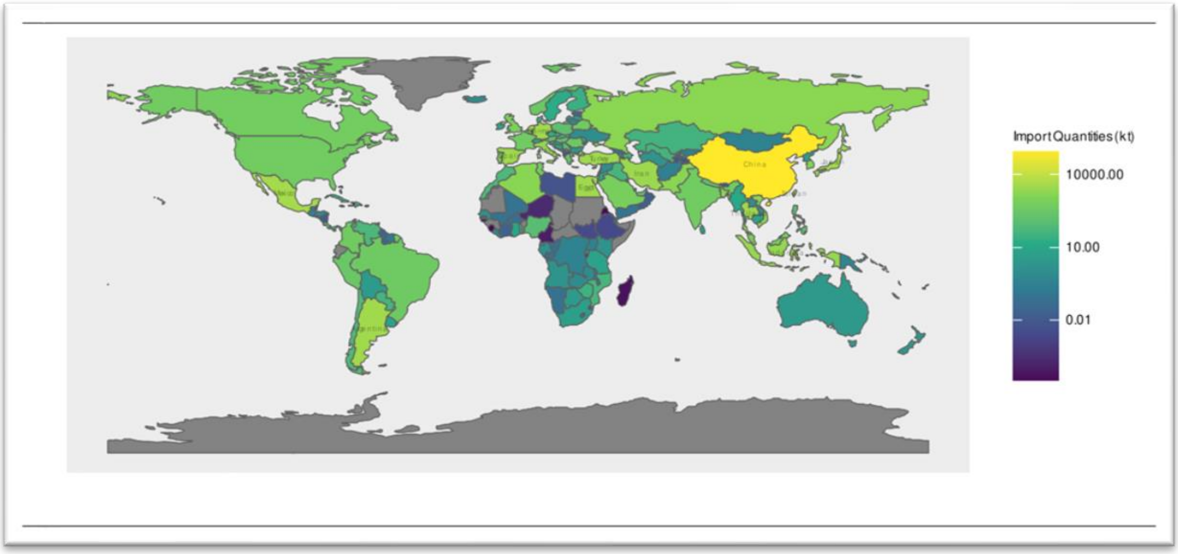

Figure 2. Soybeans Import Quantities by Countries (2022) (Data source: Anonymous (2022b)).

The data, presented by Anonymous (2022b), indicates that Brazil holds the leading position as the largest exporter of soybeans. During that period, approximately 78.9 million tonnes were exported from Brazil. The United States of America follows closely behind, exporting approximately 57.3 million tonnes, while Argentina ranks third with a reported export volume of approximately 5.2 million tonnes. Canada is in fourth place with estimated annual soybean exports of approximately 4.3 million tonnes, while Uruguay holds the fifth position with approximately 3.1 million tonnes exported. Türkiye's reported soybean exports during this period were approximately 106,907 tonnes, placing the country in nineteenth position in this comparison (Figure 3).

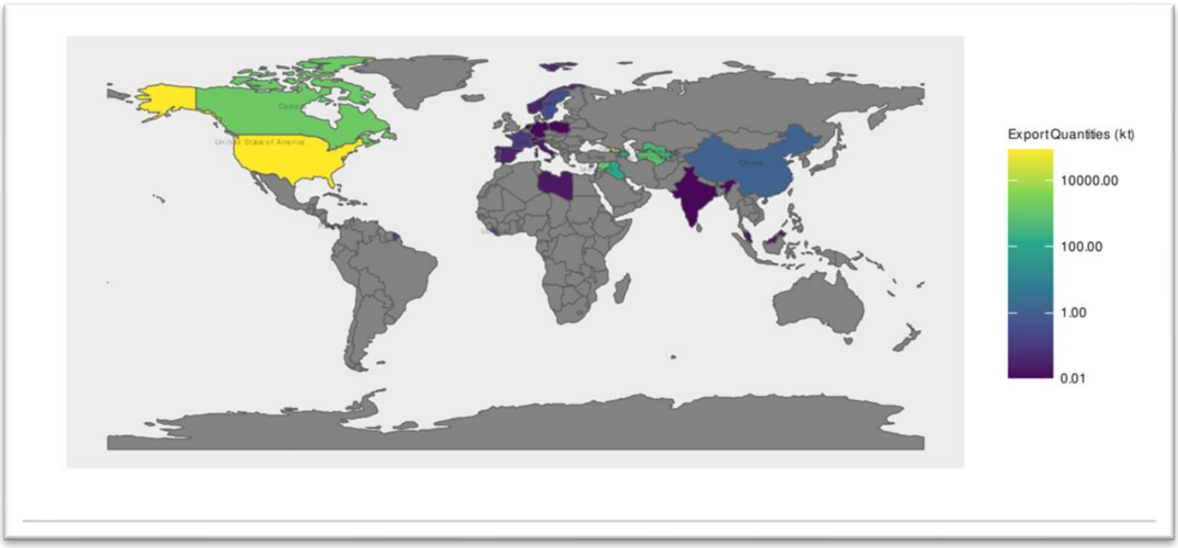

Figure 3. Soybeans Export Quantities by Countries (2022) (Data source: Anonymous (2022b)).

This study compares the quantities of soybeans imported into Türkiye from five major supplying countries as of recent data (Anonymous, 2022c). Brazil holds the first position with a substantial volume of 2,015,828 tonnes exported to Türkiye, followed closely by Ukraine in second place with an export quantity of 699,865 tonnes. The United States of America ranks third with soybean exports totalling 229,817 tonnes, while Argentina occupies the fourth position with a relatively smaller volume of 39,387 tonnes exported. Uruguay completes the list in fifth place with an export quantity of 14,194 tonnes. These findings highlight the significant role these countries play in Türkiye's soybean imports (Figure 4).

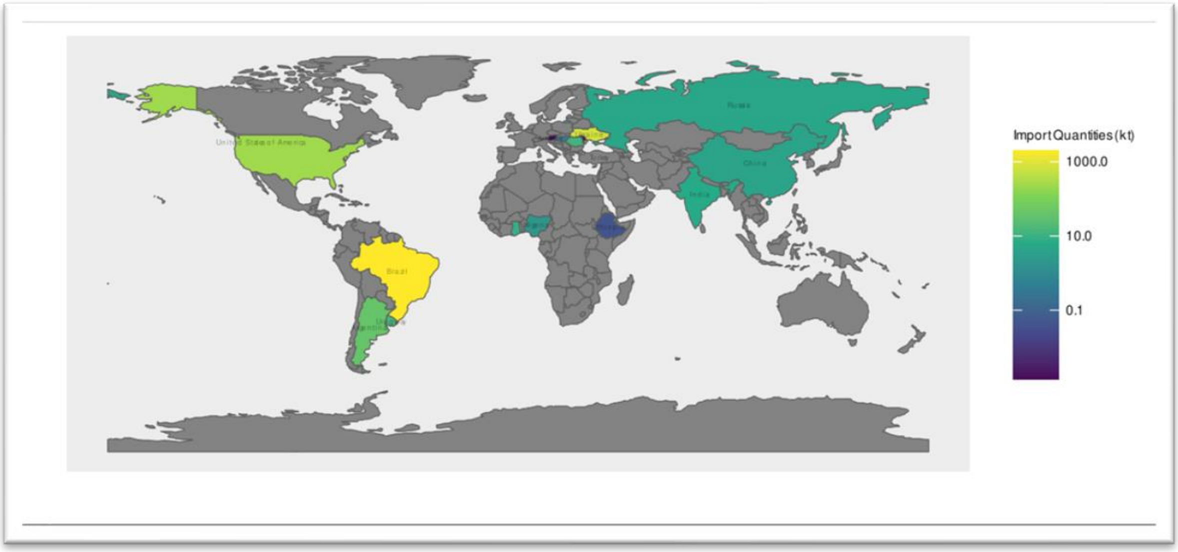

Figure 4. Total Soybean Imports into Türkiye in 2022 (kilo tonne) (Data source: Anonymous (2022c)).

The data (Anonymous, 2022c) on the export quantities of soybeans from Türkiye to various countries indicates that the United States of America is the leading importer of Turkish soybeans, with a significant volume of 88,376 tonnes. Georgia occupies the second position in this ranking with an import quantity of 12,102 tonnes, followed by Northern Cyprus (Cyprus), with 2,350 tonnes. Canada completes the top four list with a relatively smaller import volume of 1,541 tonnes (Figure 5).

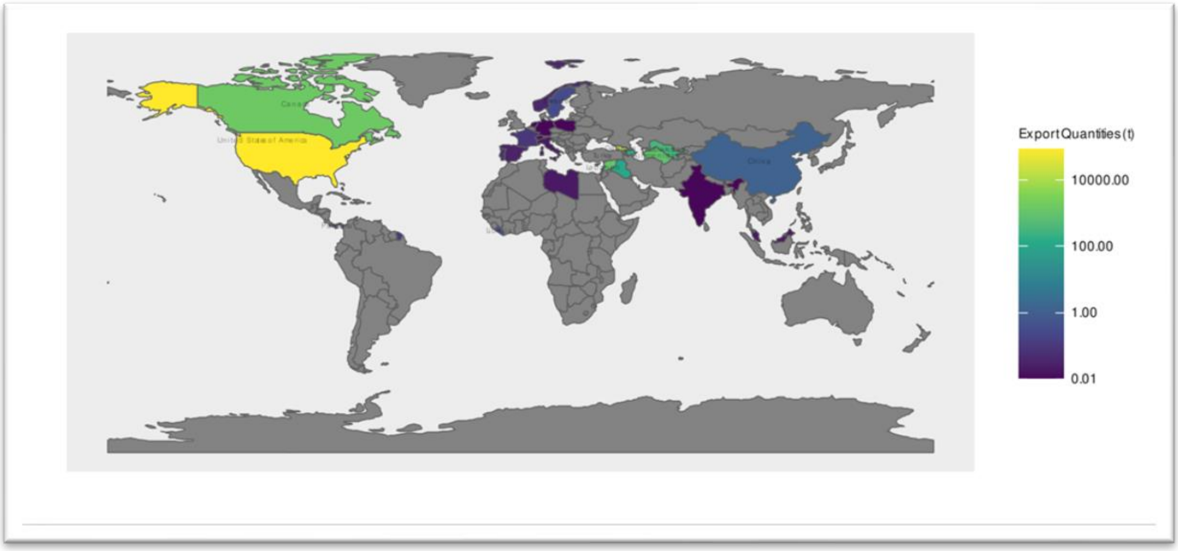

Figure 5. Total Soybean Exports from Türkiye in 2022 (tonnes) (Data source: Anonymous (2022c)).

Figure 6 presents the annual quantities of soybean production in Türkiye from 1990 to 2022. The KPSS test for stationarity yielded a test statistic of 0.48, which did not reject the null hypothesis of stationarity at the 5% significance level (p = 0.046). Nevertheless, the first differences of the series were found to be stationary (p = 0.10).

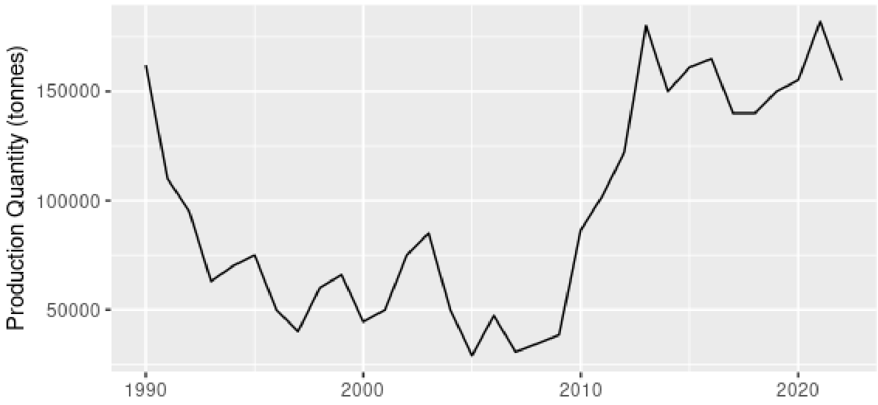

Figure 6. Annual Soybean Production in Türkiye (tonnes, 1990-2022) (Data source: Anonymous (2022d)).

The dataset utilized in this analysis is divided into two subsets. The first subset of data, designated as the training set, encompasses the period from 1990 to 2014. This portion of the data is employed for the development and calibration of statistical models and machine-learning algorithms. The second subset of data, designated as the test set, comprises observations from the years 2015 to 2022. The primary objective is to assess the performance and accuracy of the developed model using data that was not included during its development or training phase. This assessment provides insights into the model's ability to generalize new information and make reliable predictions for future time points.

By separating the dataset into a training set and a test set, researchers ensure an objective evaluation of their models while minimizing potential overfitting issues that may arise when using all available data for both development and testing purposes (Kuhn and Johnson, 2013).

Table 1. ARIMA Models Test Dataset Results Comparison.

Models	RMSE	MAE	MAPE
ARIMA (1,1,1)	13 019	10 363	6.6
ARIMA (0,1,0)	14 240	11 028	6.8
ARIMA (1,1,0)	16 367	13 033	7.9
ARIMA (0,1,1)	17 230	13 724	8.4
AUTOARIMA	49 096	42 202	28.0

ARIMA models are popular statistical tools used in time series forecasting. Among the various ARIMA models tested, the ARIMA (1,1,1) model demonstrated the highest performance based on three commonly used goodness-of-fit metrics: RMSE, MAE, and MAPE (Table 1). Specifically, the RMSE value for this model was 13 019, MAE was 10363, and MAPE was a relatively low 6.6%. On the other hand, the application of AutoARIMA, an automatic algorithm used to select optimal ARIMA parameters, did not yield successful results in this particular study. This observation underscores the importance of careful consideration and validation when using automated methods for model selection.

ES techniques are another popular class of time series forecasting models that adaptively revise previous predictions based on new data points. In the context of this study, 3 ES models were evaluated using an internal test dataset to assess their goodness-of-fit performance. Among these methods, SES emerged as the top performer. The SES model achieved impressive results in terms of RMSE, which was 13 888; MAE, with a value of 10 806; and MAPE at a relatively low rate of 6.7% (Table 2).

Table 2. ES Models Test Dataset Results Comparison.

Models	RMSE	MAE	MAPE
SES	13 888	10 806	6.7
DHLT	14 033	10 896	6.7
HLT	25 351	19 958	12.0

Four neural network (NNAR) models were trained with different numbers of nodes in their hidden layers: 5, 10, 25, and 50. As shown in Table 3, among these models, the one with only 5 nodes in the hidden layer achieved the best goodness of fit values (RMSE = 58 393, MAE = 56 135, MAPE = 45). However, the model's performance was found to be significantly inferior compared to the ARIMA and ES methods. As with the results from the auto-ARIMA process, a potential reason for the inferior performance could be that the automated processes used to determine the AR(p) value in the NNAR model might have been less accurate or optimal than manual methods. In some cases, manually selecting an appropriate order (p) can lead to better results and improved overall performance of the NNAR model. Furthermore, the suboptimal performance of the NNAR models may be attributed to the distinctive characteristics of the data. In the event that the data exhibits non-linearity, seasonality, or intricate patterns that are inadequately captured by the automated AR(p) selection process, this could result in suboptimal model performance.

Table 3. NNAR Models Test Dataset Results Comparison.

Models	RMSE	MAE	MAPE
NNAR_5	58 393	56 135	35
NNAR_10	87 658	86 643	55
NNAR_50	107 494	105 733	67
NNAR_25	108 472	107 395	69

In order to establish a baseline for comparison against more advanced forecasting models, two fundamental time series prediction techniques were included: the mean method and the naive method. The mean method postulates that future values can be approximated by the average of historical data points. In contrast, the naive method assumes that future observations will be consistent with the most recent value. The performance metrics were as follows: The mean method exhibited a RMSE of 80 418, a MAE of 79 377, and a MAPE of 50.5%. In contrast, the naive method exhibited a RMSE of 14 240, a MAE of 11 028, and a MAPE of 6.8%. These straightforward techniques serve as a foundation for evaluating more intricate models, such as NNAR. Notably, the significantly superior performance of the naive method indicates that the dataset may not be optimally suited for automated selection algorithms utilized in NNAR or Auto-ARIMA. This suggests a necessity for manual tuning to achieve optimal results. The forecasted values (each model represented in different colors) and the observed values (shown as a black line) for the test set are depicted in Figure 7.

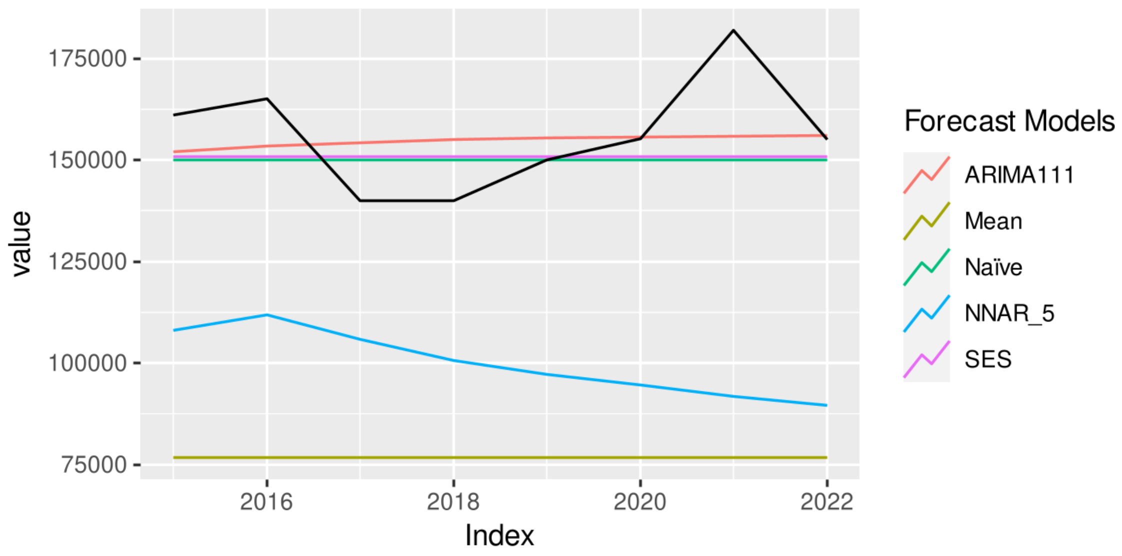

Figure 7. Visual Comparison of Forecast Models in the Test Set.

'The analysis of time series data revealed that the ARIMA (1,1,1) model provided the optimal fit to the historical soybean production data. Although neither the Autoregressive (AR) coefficient nor the Moving Average (MA) coefficient were statistically significant as shown in Table 4, the ARIMA (1,1,1) model demonstrated superior performance during cross-validation. These estimates enabled the forecasting of soybean production over the next ten years, as illustrated in Figure 8.

Table 4. Coefficients and Standard Errors for ARIMA (1,1,1) Model.

Parameter	Value	Standard Error	P Values
ar[1]	0.66	0.85	0.44^{ns}
ma[1]	-0.60	0.88	0.50^{ns}

^{ns} not significant

A reliable forecasting method is expected to exhibit residual errors that adhere to the following characteristics: the residual errors should be uncorrelated and their mean value should be zero. If the mean value deviates from zero, the forecasts are biased. In addition to these fundamental characteristics, it is beneficial, though not essential, for the residuals to exhibit constant variance (homoscedasticity) and to follow a normal distribution (Hyndman *et al.*, 2008).
In evaluating the quality of our forecasts, it is essential to examine the characteristics of the residual errors. The mean value of the residual errors was found to be -189.41. The applied Ljung-Box test indicated no autocorrelation problem among the residual errors, with a Q statistic of 4.84, degrees of freedom (df) equal to 10, and a p-value of 0.90. This result suggests that there is no evidence of correlation between successive residuals, which is an essential characteristic for unbiased forecasts. The Shapiro-Wilk test suggested that the residual errors follow a normal distribution with a W statistic value of 0.97 and a p-value of 0.49. This result indicates that there is no significant deviation from normality in the residuals, which is beneficial but not essential for accurate forecasts. These findings provide strong evidence that the ARIMA(1,1,1) model produces unbiased and robust forecasts for soybean production in Türkiye over the next decade (Table 5).

Table 5. Soybean Production Forecast in Türkiye (Tonnes, 2023-2032)[*].

Year	Mean	CI % 80[l]	CI % 80[u]	CI % 95[l]	CI % 95[u]
2023	154 490	125 616	179 988	102 870	209 806
2024	154 536	110 309	194 480	89 919	221 663
2025	154 396	100 011	207 434	71 617	236 583
2026	154 507	91 682	218 100	55 951	251 647
2027	154 231	83 321	226 654	44 841	264 794
2028	153 965	75 264	234 520	34 537	275 546
2029	153 443	67 461	240 257	24 237	284 888
2030	153 801	61 545	244 421	14 334	296 458
2031	154 157	54 988	252 841	6 828	306 411
2032	153 968	51 074	258 312	-5 327	313 024

[l] lower bound, [u] upper bound, CI = Confidence Interval, [*] the results obtained using the ARIMA(1,1,1) model

The obtained forecasts indicate that soybean production in Türkiye will continue to follow a horizontal trend, with only slight fluctuations observed over the next decade. These findings are crucial for farmers and policymakers as they can help inform decisions related to resource allocation, crop planning, and market strategies. Further analysis of these results is ongoing to gain deeper insights into the factors influencing soybean production trends in Türkiye (Figure 8).

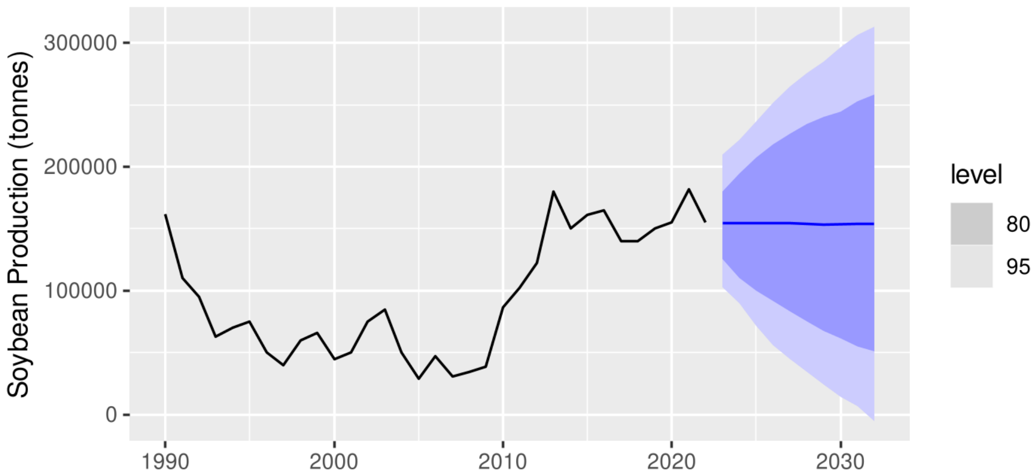

Figure 8. Forecasted Soybean Production in Türkiye (Thousand Tonnes, Next Decade).

CONCLUSION

For this time series dataset, the ARIMA (1,1,1) model was found to be the most effective. The SES model followed closely but performed slightly worse, though the difference was not significant. It is important to note that the performance of automated processes was relatively worse than that of the manual methods, suggesting that relying solely on automated methods may lead to suboptimal forecasting results. These findings underscore the importance of human oversight in the use of automated algorithms for time series forecasting and the need for caution when employing automated methods.

Brazil and the United States collectively dominate both the production and export markets for soybeans, while China plays a significant role in the import market. Türkiye occupies a relatively minor position in terms of production (32nd) and exports (19th). Brazil is the largest supplier of soybean imports into Türkiye, with a substantial volume of 2 015 828 tonnes. Ukraine follows closely in second place, while the United States holds the third position as both a significant importer and exporter. The findings underscore the significant role that Brazil, Ukraine, and the United States play in both Turkish soybean imports and exports.

Local producers in Türkiye have a significant opportunity to thrive in the market. This opportunity arises from Türkiye's reliance on imports due to insufficient domestic production capacity. While ARIMA (1,1,1) forecasts indicate stagnant production growth in Türkiye over the next decade, despite an anticipated increase in demand, this situation presents potential benefits for local producers. However, local producers need to compete with international competitors by supplying both domestic and export markets. In this context, policymakers have a role to play in supporting local producers.

This study underscores the pivotal role of soybeans in a multitude of industries, including food processing, animal feed production, and industrial applications. A comprehensive understanding of these trade dynamics can also assist stakeholders in identifying potential avenues for collaboration or investment within the Turkish soybean industry. An understanding of the global trends in soybean production and trade provides valuable insights into several key areas, including market conditions, price fluctuations, potential investments, threats, and strategic partnerships. Further analysis is currently being conducted in order to gain a more profound understanding of the factors that are influencing the trends in soybean production in Türkiye.

ACKNOWLEDGEMENTS

Statement : This research was orally presented at the Silk Road 2023 International Scientific Research Conference on September 26-27, 2023, at Iğdır University, and its abstract was published in the conference proceedings.

Statement : During the preparation of this work the author used ChatGPT, Gemini and DeepL in order to improve readability and language. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Statement : The GitHub repository, which includes the source code, data and code references is available at https://github.com/hakan-duman-acad/article-soybeans-forecasting

REFERENCES

Akın, M., Eyduran, S.P., Çelik, Ş., Aliyev, P., Aykol, S., Eyduran, E., (2021). Modeling and forecasting cherry production in Turkey. *Journal of Animal & Plant Sciences* 31(3): 773–781. https://doi.org/10.36899/JAPS.2021.3.0267

Anonymous., (2022d). Yağlı Tohumlar. https://data.tuik.gov.tr/Bulten/DownloadIstatistikselTablo?p=4VOhKR8BrNY4kV5kQy40A4Ds6MrlLde0frCYb2fEpnEWC9xHHViRbOji7MtzOuUp, (Date of access: September 2, 2023).

Anonymous., (2022a). FAOSTAT database: Production - crops and livestock products. https://www.fao.org/faostat/en/#data/QCL, (Date of access: May 20, 2024).

Anonymous., (2022b). FAOSTAT database: Trade - crops and livestock products. https://www.fao.org/faostat/en/#data/TCL, Date of access: May 20, 2024).

Anonymous., (2022c). FAOSTAT database: Detailed trade matrix. https://www.fao.org/faostat/en/#data/TM, (Date of access: May 20, 2024).

Arıoğlu, H., (2016). Türkiye'de yağlı tohum ve ham yağ üretimi, sorunlar ve çözüm önerileri. *Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi* 25(ÖZEL SAYI-2): 357–368.

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., (2013). *Applied spatial data analysis with R, second edition*. Springer, NY.

Box, G.E.P., Jenkins, G.M., (1970). *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, USA.

Gardner, E.S., Mckenzie, Ed., (1985). Forecasting Trends in Time Series. *Management Science* 31(10): 1237–1246. https://doi.org/10.1287/mnsc.31.10.1237

Gujarati, D.N., Porter, D.C., (2009). *Basic Econometrics*. McGraw-Hill, Boston, USA.

Güler, D., Saner, G., Naseri, Z., (2017). Yağlı tohumlu bitkiler ithalat miktarlarının arıma ve yapay sinir ağları yöntemleriyle tahmini. *Balkan ve Yakın Doğu Sosyal Bilimler Dergisi* 3(1): 60–70.

Holt, C.C., (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20(1): 5–10. https://doi.org/https://doi.org/10.1016/j.ijforecast.2003.09.015

Hyndman, R.J., (2021). *Forecasting: Principles and practice*. OTexts, Melbourne, Australia.

Hyndman, R.J., Khandakar, Y., (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software* 27(3): 1–22. https://doi.org/10.18637/jss.v027.i03

Hyndman, R.J., Koehler, A., Ord, K., Snyder, R., (2008). *Forecasting with exponential smoothing: The state space approach.* Springer, Berlin.

Kuhn, M., Johnson, K., (2013). *Applied predictive modeling*. Springer, New York.

Massicotte, P., South, A., (2023). *Rnaturalearth: World map data from natural earth*.

Mélard, G., Pasteels, J-M., (2000). Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting* 16(4): 497–508. https://doi.org/https://doi.org/10.1016/S0169-2070(00)00067-4

O'Hara-Wild, M., Hyndman, R., Wang, E., (2023b). *Feasts: Feature extraction and statistics for time series*.

O'Hara-Wild, M., Hyndman, R., Wang, E., (2023a). *Fable: Forecasting models for tidy time series*.

Özcan, M., (2023). Ürün raporu SOYA 2023. Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü, Ankara.

Pagano, M.C., Miransari, M., (2016). The importance of soybean production worldwide. In Miransari M (ed.) *Abiotic and biotic stresses in soybean production*. Academic Press, San Diego, 1–26.

Pebesma, E., (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10(1): 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E.J., Bivand, R., (2005). Classes and methods for spatial data in R. *R News* 5(2): 9–13.

Pratap, A., Gupta, S.K., Kumar, J., Solanki, R.K., (2012). Soybean. In Gupta SK (ed.) *Technological innovations in major world oil crops, volume 1: breeding*. Springer New York, New York, NY, 293–321.

R Core Team., (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

South, A., (2017). *Rnaturalearthdata: World vector map data from natural earth used in 'rnaturalearth'*.

Tiwari, S.P., (2017). Emerging trends in soybean industry. *Soybean Research* 15(1): 1–17.

Tüfekçi, Ş., (2019). 2019 ar-ge soya fasülyesi raporu. report. Ereğli Ticaret Borsası, Konya.

Uçum, İ., (2016). ARIMA modeli ile Türkiye soya üretim ve ithalat projeksiyonu. *Tarım Ekonomisi Araştırmaları Dergisi* 2(1): 24–31.

Wang, E., Cook, D., Hyndman, R.J., (2020). A new tidy data structure to support exploration and modeling of temporal data. *Journal of Computational and Graphical Statistics* 29(3): 466–478. https://doi.org/10.1080/10618600.2019.1695624

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software* 4(43): 1686. https://doi.org/10.21105/joss.01686