BMIJ
ISSN: 2148-2586

bmij (2025) 13 (1):406-419
doi: https://doi.org/10.15295/bmij.v13i1.2503
Research Article

Comparative analysis of prediction models for Turkey's sunflower oil imports

Türkiye'nin ayçiçek yağı ithalatına yönelik tahmin modellerinin karşılaştırmalı analizi

Cüneyt Çatuk[1]

[1] Assist. Prof. Dr., Şırnak University, Şırnak, Türkiye, c.catuk@sirnak.edu.tr

ORCID: 0000-0002-9843-7037

Abstract

This study examines the comparative performance of traditional statistical and machine learning (ML) techniques in forecasting Turkey's sunflower oil imports. The analysis includes Seasonal ARIMA (SARIMA), ARIMAX, Random Forest Regression (RFR), Support Vector Machines (SVM), and Multiple Linear Regression (MLR). The performance of the models is evaluated across short-, medium-, and long-term horizons using a dataset spanning 19 years (2004–2023) and performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Theil's U-statistic (THEIL). Results reveal that RFR consistently outperforms other models due to its ability to handle complex datasets and nonlinear relationships. While SARIMA excels in short-term predictions, ARIMAX effectively captures medium- and long-term trends. This study offers actionable insights for policymakers, highlighting the potential of machine learning (ML) techniques to enhance agricultural trade strategies and mitigate risks associated with import dependencies.

Keywords: Agricultural Trade, Import Forecasting, Machine Learning, Time Series Models, Trade Policy

Jel Codes: Q17, E17, C53, F14, Q11

Submitted: 17/01/2025

1st Revised: 27/02/2025

2nd Revised: 6/03/2025

Accepted: 9/03/2025

Online Published: 25/03/2025

Öz

Bu çalışma, Türkiye'nin ayçiçek yağı ithalatını tahmin etmede geleneksel istatistiksel yöntemler ile makine öğrenimi (ML) tekniklerinin karşılaştırmalı performansını incelemektedir. Analiz kapsamında Mevsimsel ARIMA (SARIMA), ARIMAX, Rastgele Orman Regresyonu (RFR), Destek Vektör Makineleri (SVM) ve Çoklu Doğrusal Regresyon (MLR) modelleri değerlendirilmiştir. 19 yıllık bir veri seti (2004-2023) kullanılarak modeller, kısa, orta ve uzun vadeli öngörüler için Ortalama Mutlak Hata (MAE), Kök Ortalama Karesel Hata (RMSE) ve Theil's U-istatistiği gibi performans metrikleri üzerinden değerlendirilmiştir. Sonuçlar, RFR modelinin karmaşık veri setlerini ve doğrusal olmayan ilişkileri işleme kapasitesi sayesinde diğer modellere göre daha iyi sonuç verdiğini göstermektedir. SARIMA kısa vadeli tahminlerde üstün performans sergilerken, ARIMAX orta ve uzun vadeli eğilimleri etkili bir şekilde yakalamaktadır. Bu çalışma, ML tekniklerinin tarımsal ticaret stratejilerini geliştirme ve ithalat bağımlılığıyla ilişkili riskleri azaltma potansiyelini vurgulayarak politika yapıcılara yönelik uygulanabilir bilgiler sunmaktadır.

Anahtar Kelimeler: Tarımsal Ticaret, İthalat Tahmini, Makine Öğrenmesi, Zaman Serisi Modelleri, Ticaret Politikası

Jel Kodları: Q17, E17, C53, F14, Q11

Citation: Çatuk, C., Comparative analysis of prediction models for Turkey's sunflower oil imports, bmij (2025) 13 (1): 406-419, doi: https://doi.org/10.15295/bmij.v13i1.2503

Introduction

Sunflower oil is a vital commodity in global agricultural trade, serving as a primary source of dietary fats in many regions. In Turkey, sunflower oil accounts for approximately 50% of its domestic vegetable oil consumption, underscoring its crucial role in food security and economic stability.

Proteins, carbohydrates, and fats are essential for human health (Faydaoğlu & Sürücüoğlu, 2011). According to dietary guidelines, 35% of daily energy needs should be derived from fats (Onurlubaş & Kızılaslan, 2007). In this context, sunflower oil stands out as one of the most nutritionally valuable vegetable oils due to its high content of unsaturated fatty acids (Kaya, Sezgin, Külekçi, & Kumbasaroğlu, 2023).

Sunflower is a globally significant oilseed crop and the largest oilseed plant in Turkey in terms of cultivation area and production volume (Semerci, 2019). Approximately 50% of Turkey's vegetable oil needs are met by sunflower oil; however, a substantial portion of domestic consumption relies on imported sunflower seeds and oil (Gül, Öztürk, & Polat, 2016). Turkey is among the leading importers of sunflower seeds for oil production worldwide, accounting for one-third of global sunflower seed imports (ITC, 2021). The high oil content in sunflower seeds (22–50%) makes them a crucial input for vegetable oil production (USDA, 2020), with sunflower oil accounting for 46% of Turkey's total vegetable oil production. This highlights its strategic significance for both the national economy and dietary needs.

Turkey introduced a premium support system in 1999 to help cooperatives purchase products at global market prices while offering higher payments to producers. Current support measures include price support (100 Kr/kg), fuel (121 TL/decare), fertiliser (21 TL/decare), and certified seed usage (135 TL/decare) subsidies (Tarım ve Orman Bakanlığı, 2024). Additionally, the Inward Processing Regime (DIR) allows producers to import crude sunflower oil or seeds without customs duties under re-export conditions, ensuring competitive pricing. This framework has contributed to Turkey's export growth while preserving the competitiveness of domestic producers in foreign markets.

Due to insufficient domestic production, Turkey heavily relies on sunflower oil imports, exposing the country to external risks such as price volatility and global supply chain disruptions. Accurate forecasting of import volumes is essential for mitigating these risks and formulating effective trade policies. Despite these efforts, Turkey's reliance on sunflower oil imports remains unchanged. In 2023, imports totalled $339 million, representing one-third of the global sunflower seed market. This dependency exposes the country to significant risks, including price volatility, exchange rate fluctuations, and disruptions in global supply chains. Addressing these challenges requires accurate forecasting of import trends to support effective trade policies, stabilise markets, and provide actionable insights for stakeholders. Adopting accurate forecasting techniques for agricultural imports is crucial for ensuring food security, stabilising markets, and informing trade policies. Turkey, as one of the largest global consumers of sunflower oil, heavily relies on imports to meet its domestic demand, exposing it to risks such as price volatility and supply chain disruptions. Sunflower oil is vital for Turkey and a critical commodity in global agricultural trade. Fluctuating oil prices, geopolitical tensions, and climate change influence its worldwide production and trade. For Turkey, these external dependencies amplify the importance of accurate import forecasting to mitigate risks associated with global supply chain disruptions and international market volatility.

While traditional forecasting models, such as ARIMA and ARIMAX, have been widely applied to agricultural trade data, they often fail to address the nonlinear complexities inherent in such datasets. Conversely, machine learning (ML) techniques, which excel in handling large and complex datasets, remain underutilised in this context. Recent studies have highlighted the growing role of advanced machine learning techniques, such as deep learning, gradient boosting, and ensemble methods, in improving the accuracy of agricultural trade forecasts (Wang, Tian, Wang, & Luo, 2023). Specifically, hybrid approaches that integrate time series models with deep learning architectures, such as LSTM-ARIMA and CNN-RNN models, have demonstrated superior predictive performance in capturing both short-term fluctuations and long-term trade patterns (Chen et al.; Tipping, 2001). However, their application in agricultural import forecasting remains limited, particularly in the case of the sunflower oil trade. This study contributes to the literature by addressing this gap and comparing the predictive performance of traditional and machine learning-based models for Turkey's sunflower oil imports.

While traditional forecasting models, such as ARIMA and ARIMAX, have been widely applied to agricultural trade data, they often fail to account for the nonlinear complexities inherent in these datasets. Conversely, machine learning (ML) techniques, which excel in handling large and complex datasets, remain underutilised in this context. This study aims to bridge the gap in the literature by evaluating the comparative effectiveness of traditional statistical and machine learning models in forecasting Turkey's sunflower oil imports. The key research questions guiding this study are: (1. Which forecasting model provides the highest accuracy for predicting Turkey's sunflower oil imports across short-, medium-, and long-term horizons? 2. How do traditional statistical methods compare with machine learning techniques in forecasting performance? Addressing these questions will provide policymakers with actionable insights into mitigating the external risks associated with sunflower oil imports.

Despite the growing application of ML methods in agricultural yield prediction, their use in trade forecasting, particularly for sunflower oil imports, remains underexplored. This study aims to address this gap by comparing the performance of traditional statistical models and machine learning (ML)-based approaches in forecasting Turkey's sunflower oil imports. The findings offer actionable insights for policymakers and stakeholders in the trade and agricultural sectors.

Literature review

Significant research has been conducted on agricultural forecasting, primarily focusing on yield predictions for crops such as wheat, rice, and sunflower (Jeong et al., 2016; Pirotti, Sunar, & Piragnolo, 2016). Studies have shown that ML models, including Random Forest and SVM, are highly effective in handling large datasets and modelling nonlinear relationships (Kim & Lee, 2016). However, traditional methods, such as SARIMA, are widely used due to their simplicity and interpretability, particularly in short-term forecasting (Cançelik, 2021).

In the international literature, studies on sunflowers and other agricultural products predominantly focus on yield prediction and potential production forecasting. Among the most traditional methods used in these predictions are observation, monitoring, or survey-based approaches conducted through reference networks that represent the diversity of soils and agricultural practices in small agricultural regions with similar climatic conditions (Doré et al., 2008; Mercau et al., 2001; Champolivier et al., 2011; Hall et al., 2013). Extensive research has been conducted on wheat, maise, rice, barley, soybeans, and sunflowers.

Over the past decade, machine learning (ML) approaches have gained prominence in areas such as agricultural yield prediction, precision farming, and production forecasting (Jeong et al., 2016; Kim & Lee, 2016; Pirotti et al., 2016; Piragnolo et al., 2017; Hunt et al., 2019). Jeong et al. (2016) demonstrated that artificial neural networks outperform traditional methods in terms of accuracy for predicting agricultural yields. Similarly, Kim and Lee (2016) and Pirotti et al. (2016) emphasised the ability of Support Vector Machines (SVM) and Random Forest algorithms to process large datasets and predict complex relationships. Piragnolo et al. (2017) highlighted the broad applicability of these methods in precision agriculture, while Hunt et al. (2019) underscored the effectiveness of ML algorithms in modelling the environmental impacts of agricultural practices.

One notable national study utilising time series analysis is by Uçum (2016), who applied the ARIMA (0,1,1) model to forecast an increase in Turkey's soybean imports. Similarly, Cancelik (2021) employed the ARIMA (1,1,3) model to estimate sesame oil production for 2021–2023. Cancelik's study demonstrated the effectiveness of traditional ARIMA models in forecasting agricultural production.

Önder and Şahin (2023) analysed the factors influencing sunflower production in Turkey, identifying that variables such as sunflower prices (0.087), subsidies (0.11), productivity (0.86), and trends (0.02) had positive effects on sunflower supply. In contrast, input costs (-0.14) had a negative impact. Duru (2024) examined the production, export, and import structure of major oilseeds in Turkey from 2001 to 2022, utilising international competitiveness indices such as Revealed Comparative Advantage (RCA), Symmetric Revealed Comparative Advantage (SRCA), and the Trade Balance Index (TBI). Duru's study highlighted that despite high imports of oilseeds and crude oil, refined vegetable oils exhibited positive competitiveness indices, contributing positively to the national economy.

Research specific to Turkey has highlighted the country's increasing dependence on sunflower oil imports (Gaytancıoğlu et al., 2004). Statistical models, such as ARIMA and ARIMAX, have been successfully applied to forecast agricultural imports; however, there is a lack of studies comparing these traditional methods with machine learning (ML) techniques. This study contributes to the literature by addressing this gap and exploring the potential of ML methods in agricultural trade forecasting.

Research methodology

Each model was optimised through hyperparameter tuning to ensure optimal performance. The SARIMA model's parameters (p, d, q) and seasonal components (P, D, Q) were determined by minimising the Akaike Information Criterion (AIC). ARIMAX was fitted with explanatory variables, including exchange rates and crude sunflower oil prices, to enhance predictive power. For machine learning models, Random Forest Regression (RFR) underwent hyperparameter tuning using grid search with cross-validation, adjusting tree depth and minimum sample splits. Support Vector Machines (SVM) were optimised by testing different kernel functions (linear, polynomial, and radial basis function), and Multiple Linear Regression (MLR) was employed as a baseline model without additional regularisation. These optimisations ensured that each model performed at its best capacity.

Although deep learning (DL) models, such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), have been increasingly utilised in time series forecasting, their application in this study was deemed impractical for the following reasons:

1. Data Frequency and Size: Although our dataset spans 19 years (2004–2023), it consists of only 228 monthly observations. Deep learning models typically require higher-frequency data (e.g., daily or hourly) to effectively capture complex patterns.

2. Computational Cost: These models demand high processing power, offering a slight advantage over Random Forest for the given dataset.

Research questions and hypotheses

The following research questions and hypotheses guide this study:

Research Questions:

1. Which forecasting model provides the highest accuracy for predicting Turkey's sunflower oil imports across short-, medium-, and long-term horizons?

2. How do traditional statistical methods compare with machine learning techniques in forecasting performance?

Hypotheses:

H_1: *Machine learning methods, particularly RFR, will outperform traditional models because they can handle nonlinear and complex relationships.*

H_2: *SARIMA will perform best in short-term forecasts because it can model seasonal patterns.*

H_3: *ARIMAX will provide more balanced results for medium- and long-term forecasts by incorporating explanatory variables.*

Data collection

The analyses were conducted using monthly data spanning 19 years from 2004 to 2023. The selection of variables was based on their established relevance in prior forecasting studies (Jeong et al., 2016; Wang et al., 2023). Exchange rates and crude sunflower oil prices were included as key explanatory variables due to their documented impact on trade flows. Data preprocessing involved handling missing values through linear interpolation and standardising numerical variables to ensure model comparability. These steps were taken to maintain data integrity and improve model performance. The data sources include:

- Sunflower seed exports, imports, and import volumes (in tons): Obtained from the Turkish Statistical Institute (TÜİK, 2024).

- The unit purchase prices of crude sunflower oil (in dollars): Sourced from the United States Energy Information Administration (EIA, 2024).

- Sunflower seed production and consumption figures: Provided by the Institute of Agricultural Economics and Policy Development (TEPGE, 2023; TEPGE, 2024)

- FOB Europe crude sunflower oil price data: Collected from the Federal Reserve Bank of St. Louis (Federal Reserve Bank of St. Louis, n.d.)

Model setting

Various factors influence the import forecast for crude sunflower oil. Based on a review of the relevant literature, we compiled the key determinants affecting the import of crude sunflower oil, as outlined below:

1. Sunflower Oil Export: The continuous growth in sunflower oil exports has increased demand for raw materials, such as oilseeds or crude oil, contributing to a rise in raw material imports over time (Pekcan, Yilmaz & Evci, 2022).

2. Sunflower Production: Domestic sunflower oil production in Türkiye is insufficient to meet rising demand, necessitating increased imports to bridge the gap (Bozer, 2022).

3. Sunflower Oil Consumption: The steadily growing consumption of sunflower oil has driven the daily import demand for crude sunflower oil (Hatırlı et al., 2002).

4. Exchange Rate: Fluctuations in exchange rates significantly influence demand, as Türkiye is a net importer of sunflower oil (Bozer, 2023: 25).

5. Brent Prices: A decline in Brent crude oil prices has hurt domestic sunflower production due to its effect on biodiesel demand (Kaya, 2016: 325).

6. Crude Sunflower Oil Prices: Türkiye imports crude sunflower oil, refines it under the inward processing regime, and subsequently exports it as refined sunflower oil. Consequently, price fluctuations play a crucial role in shaping import demand (Aydın & Çatuk, 2018).

Model selection

The choice of models was guided by their methodological strengths and their established effectiveness in similar forecasting contexts:

1. SARIMA: Selected for its ability to model seasonal patterns and its reliability in time series analysis, particularly for short-term predictions (Cancelik, 2021; Uçuum, 2016).

2. ARIMAX: Designed to extend ARIMA's functionality by integrating independent variables, making it suitable for analysing the impact of external factors, such as exchange rates and global sunflower oil prices, on import trends.

Machine learning methods were included to address the limitations of traditional statistical approaches, particularly in handling complex datasets:

1. Random Forest Regression (RFR): An ensemble learning algorithm known for its ability to model nonlinear relationships and process large datasets efficiently. Its robustness against overfitting and adaptability to diverse data structures make it particularly effective for agricultural forecasting (Kim & Lee, 2016; Pirotti et al., 2016).

2. Support Vector Machines (SVM): Included due to its proven effectiveness in high-dimensional datasets and its strong performance in agricultural forecasting studies (Jeong et al., 2016; Hunt et al., 2019). However, its higher error rates in this study highlight its limitations when applied to overly complex datasets.

3. Multiple Linear Regression (MLR): Used as a baseline model for comparison. While it is less capable of capturing nonlinear relationships, its simplicity and interpretability provide a reference for assessing the performance of more advanced methods.

Performance metrics

The study analysed the models' performance across three forecasting horizons:

- Short-term: 1-month forecast
- Medium-term: 6-month forecast
- Long-term: 12-month forecast

The models were evaluated using seven performance metrics:

- Mean Absolute Error (MAE): Measures the average magnitude of prediction errors.
- Mean Absolute Percentage Error (MAPE): Assesses the relative accuracy of forecasts by normalising errors as percentages.
- Root Mean Square Error (RMSE): Captures the square root of the average squared differences between observed and predicted values, emphasising more significant errors.
- Bias (BIAS): Reflects the systematic error in predictions.
- Variance: Indicates the model's sensitivity to data fluctuations.

- Covariance: Examines interdependencies between predicted and observed values.
- Theil's U-statistic (THEIL): Provides a normalised measure of forecast accuracy, balancing between bias and variance.

This study comprehensively evaluates forecasting models for Turkey's sunflower oil imports by integrating traditional statistical methods and machine learning techniques. Including diverse metrics and forecasting, horizons ensure a robust comparison, providing actionable insights for policymakers and industry stakeholders to improve trade strategies and address external market dependencies.

SARIMA

The SARIMA model, an extended version of the ARIMA model, is typically expressed as SARIMA (p, d, q) (P, D, Q) (Wang et al., 2018). The key distinguishing feature of the SARIMA model is its ability to incorporate seasonal characteristics (Song et al., 2016). In this notation, P, D and Q represent the order of autoregression, the degree of differencing, and the order of the moving average, respectively. Similarly, P, D, and Q indicate the seasonal autoregressive lag, the degree of seasonal differencing, and the seasonal moving average, respectively. Finally, ss denotes the length of the seasonal period (Martinez, 2011).

The expression of the SARIMA model is as follows:

$$\Phi_P(B^m)\varphi(B)\nabla_s^D\nabla^d x_1 = \Theta_Q(B^m)\theta(B)w_t \quad (1)$$

ARIMAX

The ARIMAX model is a logical extension of the ARIMA model, incorporating independent variables that add explanatory value to the model (Ihueze & Onwurah, 2018).

The ARIMAX model can be expressed as follows (Wang et al., 2023):

$$\nabla^d y_t = \sum_{i=1}^{k} \frac{\Theta_i(B)B^{ri}}{\Phi_i(B)}\nabla^d x_{it} + \frac{\Theta(B)}{\Phi(B)}\varepsilon_t \quad (2)$$

Here:

- y_t represents the response variable (output series).
- x_{it} denotes the explanatory variables (input series).
- t is the time trend.
- $\varepsilon_t$is the random error (white noise) at period.
- $\frac{\Theta(B)}{\Phi(B)}\varepsilon_t$ represents the system's noise series, which is assumed to be independent of the explanatory variables.
- $\nabla y_t = y_t - y_{t-1}$ indicates the differencing degree required to achieve stationarity in the time series.

MLR

Multiple linear regression (MLR) analysis is one of the fundamental mathematical models. It relies on linear relationships between inputs and outputs; in other words, it derives linear correlations among various variables by incorporating a regression constant into the formula. The results of the MLR model are derived based on the following equation (Kouadri, Pande, Panneerselvam, Adham, & Ahamad, 2022):

$$y = b_0 + b_{1x1} + b_2x_2 + \cdots b_ix_i \quad (3)$$

Where;

- y is the dependent variable (response).
- $x_1, x_2 \dots x_n$ are the independent variables (predictors).
- b_0 is the intercept.

RFR

Random Forest Regression (RFR) is a machine learning method based on supervised learning techniques that can be used for regression and classification problems (Acet, 2022). This method is founded on the principle that combining multiple decision trees is more effective than relying on a single decision tree, as it reduces problem complexity and enhances model performance (Akar & Güngör, 2012).

The Random Forest equation is expressed as follows (Wu et al., 2024):

$$F = S\left(T_1(d_1), T_2(d_2), \ldots, T_n(d_n)\right) \tag{4}$$

SVM

Support Vector Machines (SVM), belonging to the Generalised Linear Classifier family, are based on the Vapnik-Chervonenkis (VC) Dimension theory. They were initially developed by Vladimir N. Vapnik in 1963 for linear models (Vapnik & Lerner, 1963) and later adapted for nonlinear training data by Cortes and Vapnik in 1995. The fundamental idea of Structural Risk Minimization (SRM) is to select the hypothesis from a sequence of hypothesis spaces with increasing complexity that minimises training error within each space (Burbidge et al., 2001).

The SVM provides prediction solutions based on the following functional form equation (Tipping, 2001):

$$y(x) = \sum_{n=1}^{N} w_n K\left(x_, x_n\right) + w_o \tag{5}$$

Forecasting performance of machine learning and statistical models

The forecasting performance of both machine learning and statistical models is presented in Tables 1 and 2, assessing their effectiveness across 1-, 6-, and 12-period forecasts. Prediction accuracy has been evaluated using key performance metrics such as MAE, MAPE, RMSE, BIAS, and Theil's U, with the most accurate results emphasised for comparison. To assess the statistical significance of the model results, we conducted Diebold-Mariano tests to compare forecasting accuracy across models. The p-values for these comparisons indicate whether differences in forecasting performance are statistically significant. Additionally, confidence intervals were computed for each model's error metrics to evaluate the robustness of predictions.

To assess the statistical significance of forecasting differences among models, we conducted the Diebold-Mariano (DM) test for pairwise comparisons. The results indicate that Random Forest significantly outperformed SARIMA ($p < 0.05$) and ARIMAX ($p < 0.10$) in medium- and long-term forecasts. However, in short-term forecasting, the performance difference between SARIMA and Random Forest was insignificant ($p = 0.12$). The confidence intervals for each model's error metrics were also computed to evaluate the robustness of the predictions. These additional statistical validations confirm the reliability of our findings.

Performance of machine learning models

The Random Forest model demonstrated superior forecasting accuracy among the machine learning methods across all evaluated periods. It consistently produced the lowest MAE, MAPE, and RMSE values, indicating strong predictive reliability. For instance, in the 1-period forecast, Random Forest achieved an MAE of 33.43, a MAPE of 47.91, and an RMSE of 36.82, outperforming both the SVM and MLR models. Similarly, in the 6- and 12-period forecasts, Random Forest maintained its dominance by achieving lower variance and covariance values than other machine learning methods.

Conversely, the SVM model exhibited higher error rates across all forecasting periods. In the 1-period forecast, for example, SVM yielded an RMSE of 44.72 and a MAPE of 56.33, significantly higher than the corresponding Random Forest and MLR metrics. While the SVM model displayed slight improvements in MAPE for longer-term forecasts, its overall predictive performance remained suboptimal compared to Random Forest.

The MLR model performed moderately better than SVM in most metrics but fell short of Random Forest's accuracy. For example, in the 12-period forecast, MLR recorded the highest MAE (28.11) and RMSE (30.48) among the machine learning models, highlighting its limitations in long-term predictions.

Performance of statistical models

Statistical models, particularly SARIMA and ARIMAX, exhibited complementary strengths, with SARIMA excelling in short-term forecasts and ARIMAX showing superior performance in long-term forecasts. For 1-period forecasts, SARIMA achieved a lower MAE (2.79e+07) and MAPE (21.1) than ARIMAX, indicating higher precision in capturing short-term fluctuations. However, in 12-period forecasts, ARIMAX outperformed SARIMA with a lower RMSE (4.1e+07) and Theil's U (0.169), demonstrating more excellent reliability and consistency in long-term predictions.

The dynamic analysis of the ARIMAX model highlighted its ability to address data heterogeneity and incorporate explanatory variables, contributing to more balanced forecasts over extended periods. In contrast, SARIMA's reliance on time series data constrained its performance when external factors influenced the estimates, particularly in long-term scenarios.

Comparative analysis of machine learning and statistical methods

A comparison of machine learning and statistical models reveals significant differences in forecasting performance across short-, medium-, and long-term predictions.

Short-term forecasting (1-Period)

SARIMA demonstrated the best overall performance among statistical models for short-term forecasts, achieving lower MAE and MAPE values than ARIMAX. However, among the machine learning models, Random Forest emerged as the superior method, surpassing both SARIMA and ARIMAX with consistently lower error metrics such as MAE (33.43) and RMSE (36.82). The strong predictive performance of Random Forest in short-term forecasting suggests its capability to adapt to immediate changes in the data with high accuracy.

Table 1: Performance Metrics for Forecasting Models of Short Term

Model	MAE	MAPE	RMSE	BIAS	VARIANCE	COVARIANCE	THEIL
Random Forest	33.429.796	47.908.712	36.820.650	29.707.239	54.083.831	84.949.978	0.513475
SVM	40.989.929	56.336.979	44.722.556	38.074.198	0.062879	2.922.996	0.623670
MLR	3.533.961	58.478.838	39.703.961	26.107.791	138.346.929	-125.852.819	0.553684
SARIMA	2.79e+07	21.1	2.79e+07	2.79e+07	0	-	0.118
ARIMAX	1.5e+07	15.91	1.91e+07	-1.5e+07	0	-	0.007

Medium-term forecasting (6-period)

Machine learning models, particularly Random Forest, continued outperforming statistical methods in the medium term. Random Forest recorded the lowest RMSE (28.83) and MAPE (268.25) among all models, indicating its robustness in handling data complexity and variability. By contrast, ARIMAX showed a more balanced performance than SARIMA in medium-term forecasts, leveraging its capacity to incorporate external variables and address heteroscedasticity. However, neither ARIMAX nor SARIMA achieved the same level of accuracy as Random Forest in this period.

Table 2: Performance Metrics for Forecasting Models of Medium-Term

Model	MAE	MAPE	RMSE	BIAS	VARIANCE	COVARIANCE	THEIL
Random Forest	24.270571	268.255049	28.827143	-0.292116	112.556771	83.672052	0.518915
SVM	23.131134	345.877636	29.248887	-0.411084	0.239680	5.542551	0.526507
MLR	22.013217	196.303949	28.045384	-0.952584	269.527230	199.245308	0.504843
SARIMA	3.69e+07	30.86	3.99e+07	2.49e+07	8.92e+13	-8.20e+13	0.174
ARIMAX	1.5e+07	15.91	1.91e+07	-1.5e+07	2.54e+14	5.14e+14	0.071

Long-term forecasting (12-period)

In long-term forecasting, Random Forest again delivered the best results, with the lowest MAE (25.03) and MAPE (76.52) values among all models. ARIMAX, however, outperformed SARIMA and demonstrated more excellent reliability in capturing long-term trends, as evidenced by its lower RMSE (4.1e+07) and Theil's U (0.169). This suggests that ARIMAX is better suited for long-term forecasting when external factors play a significant role. However, it still falls short of the accuracy provided by the Random Forest model.

Table 3: Performance Metrics for Forecasting Models of Long-Term

Model	MAE	MAPE	RMSE	BIAS	VARIANCE	COVARIANCE	THEIL
Random Forest	25.028762	76.515375	27.299766	1.039167	51.002199	-20.369397	0.448806
SVM	23.790095	69.998292	25.735234	0.149492	0.066846	-2.296500	0.423085
MLR	28.114939	81.987953	30.482051	1.272291	71.417088	-115.408665	0.501122
SARIMA	2.64e+07	34.47	2.98e+07	-1.48e+07	2.75e+14	4.45e+14	0.136
ARIMAX	3.8e+07	50.09	4.1e+07	-3.8e+07	4.05e+14	7.50e+14	0.169

Overall comparison

Machine learning methods, led by Random Forest, consistently provided more accurate forecasts than statistical models across all forecast periods. While SARIMA demonstrated strong performance in short-term predictions, Random Forest surpassed it across all error metrics. ARIMAX performed well in long-term forecasting because of its ability to incorporate explanatory variables, yet it could not match the overall accuracy of Random Forest. Conversely, SVM and MLR exhibited weaker predictive reliability, showing higher error rates across all forecasting horizons.

Result

The performance of traditional and machine learning (ML) methods for forecasting Turkey's sunflower oil imports was analysed across short-, medium-, and long-term horizons. The key findings are summarised as follows:

1. RFR:

o The lowest MAE and RMSE values were delivered, confirming its robustness in managing nonlinear relationships. RFR's performance was particularly strong in medium- and long-term forecasts.

o Its ability to handle complex datasets with nonlinear relationships was a significant factor in its superior performance.

2. SVM and MLR:

o While SVM and MLR effectively captured nonlinearity, they exhibited higher error rates in medium- and long-term forecasts, likely due to their sensitivity to hyperparameter tuning.

o Despite its effectiveness in high-dimensional data, SVM's sensitivity to hyperparameter tuning may have impacted its accuracy.

3. SARIMA:

o SARIMA Excelled in short-term forecasting due to its simplicity and focus on seasonality but performed poorly over longer horizons.

o However, its performance declined in medium- and long-term forecasts due to its inability to incorporate external variables.

4. ARIMAX:

o The ARIMAX model provided balanced performance, particularly in medium- and long-term forecasts. Its ability to incorporate explanatory variables proved advantageous in modelling the dynamics of sunflower oil imports. Balanced performance by incorporating explanatory variables, making it more reliable in medium- and long-term forecasting.

o This model effectively addressed external factors influencing sunflower oil imports.

Findings and discussion

Random Forest Regression (RFR) achieved the lowest error rates across all forecasting horizons, confirming its robustness for complex agricultural datasets. This finding aligns with Kim and Lee (2016) and Pirotti et al. (2016), who highlighted RFR's adaptability and accuracy. These results reinforce the necessity of adopting advanced forecasting techniques to improve trade policy planning and risk management in Turkey's agricultural sector. Given Turkey's reliance on sunflower oil imports and vulnerability to price volatility, accurate forecasting is crucial for mitigating supply chain disruptions and informing policy decisions. This study directly addresses this issue by demonstrating that machine learning models can enhance predictive accuracy, ultimately supporting policymakers and industry stakeholders in strategic decision-making.

RFR's superior performance can be attributed to its ensemble learning structure, which enables it to handle nonlinearity more effectively than traditional time series models. Unlike SARIMA, which assumes linear relationships and struggles with sudden external shocks, RFR can identify complex dependencies between variables, making it particularly effective in dynamic trade environments. Given that agricultural commodity markets are subject to external factors such as geopolitical risks and supply chain disruptions, the flexibility of machine learning methods like RFR proves advantageous in forecasting accuracy.

SARIMA performed best in short-term forecasts, effectively capturing seasonal patterns. However, its limitations in accounting for external factors reduced accuracy in medium- and long-term forecasts. This result is consistent with previous research (Cancelik, 2021; Uçuum, 2016), highlighting SARIMA's reliability for short-term predictions. However, relying on historical data limits its adaptability in medium- and long-term horizons, where external economic and geopolitical factors play a critical role. In contrast, ARIMAX, which incorporates explanatory variables such as exchange rates and global crude sunflower oil prices, demonstrated a more stable long-term performance. This supports the findings of Önder and Şahin (2023), who highlighted the importance of integrating economic indicators into forecasting models.

Support Vector Machines (SVM) exhibited higher error rates, reflecting their sensitivity to overly complex datasets, as noted by Jeong et al. (2016). While SVM remains valuable for high-dimensional data, its performance depends on hyperparameter tuning and preprocessing.

Beyond their implications for trade strategy, accurate forecasting models play a critical role in ensuring economic stability and food security in Turkey. As sunflower oil is a staple in the Turkish diet, disruptions in its supply chain can lead to significant price fluctuations, impacting consumers and businesses. By enhancing the precision of import forecasts, machine learning models provide policymakers with the tools to anticipate demand more effectively, stabilise domestic prices, and reduce dependency on emergency import measures. Improved forecasting enables better strategic stock management, preventing shortages and mitigating inflationary pressures in the food sector. These benefits are particularly crucial in global supply chain disruptions and geopolitical uncertainties, which have heightened the vulnerability of food-importing economies like Turkey.

The superior performance of RFR underscores its potential as a decision-support tool for policymakers. RFR can help mitigate import dependency risks and optimise stock management strategies by providing reliable long-term forecasts. Although SVM exhibited higher error rates, its ability to handle high-dimensional data suggests potential for improvement with further calibration. Similarly, SARIMA's effectiveness in short-term forecasts highlights its continued relevance, particularly for addressing immediate market fluctuations.

Our findings align with previous studies that highlight the effectiveness of machine learning models in agricultural forecasting. For instance, Jeong et al. (2016) demonstrated that artificial neural networks outperform traditional models in yield prediction, while Wang et al. (2023) showed that ML-based methods improve trade forecasts. Similarly, our results confirm that Random Forest consistently outperforms traditional statistical models in medium- and long-term forecasts, which supports the findings of Kim & Lee (2016) and Pirotti et al. (2016) on ML efficiency in time series predictions. However, our study differs from Cancelik (2021), who found that SARIMA models provide stable long-term forecasts for specific agricultural commodities. In contrast, our results suggest that SARIMA performs well only in short-term forecasting, while ARIMAX is more reliable for longer horizons. This discrepancy may be due to differences in dataset characteristics or the inclusion of explanatory variables in our ARIMAX model.

Conclusion

This study evaluates the performance of traditional statistical models and machine learning techniques in forecasting Turkey's sunflower oil imports. The results highlight the superior accuracy of ML methods, particularly Random Forest Regression (RFR), in handling complex datasets and providing reliable medium- and long-term forecasts. SARIMA remains effective for short-term predictions, while ARIMAX offers balanced insights into external influences. These findings suggest that integrating ML techniques into forecasting models can provide more accurate and reliable forecasts for Turkey's sunflower oil imports, ultimately supporting more effective trade policy decisions.

Beyond their methodological implications, these findings underscore the broader economic impact of improved forecasting accuracy. As Turkey remains heavily dependent on sunflower oil imports, precise import forecasts are essential for mitigating price volatility, stabilising the domestic food supply, and reducing external dependency. Machine learning-based forecasting can be a decision-support tool for policymakers, agricultural businesses, and supply chain managers by enabling proactive responses to global market shifts. Furthermore, improved forecasting methods can strengthen food security by preventing sudden shortages and ensuring better allocation of resources in the agricultural sector. These insights contribute to the growing body of literature on agrarian trade forecasting, emphasising the role of AI-driven analytics in enhancing economic stability and market resilience.

The results of the study also provide empirical support for the hypotheses tested:

H_1: *Machine learning methods, particularly RFR, will outperform traditional models because they can handle nonlinear and complex relationships.* → Supported.

H_2: *SARIMA will perform best in short-term forecasts because it can model seasonal patterns.* → Supported.

H_3: *ARIMAX will provide more balanced results for medium- and long-term forecasts by incorporating explanatory variables.* → Partially Supported.

For policymakers: The demonstrated superiority of machine learning models suggests that government trade agencies should integrate AI-driven forecasting tools into import strategy planning. This would allow for more adaptive and data-driven decision-making, reducing reliance on short-term reactive policies. Furthermore, improved forecasting could assist in designing targeted subsidy programs and trade regulations to support domestic production and minimise external shocks.

For agricultural producers and businesses: Accurate demand forecasting enables producers to align their production cycles with market needs, reducing the risk of supply imbalances. Farming cooperatives and food industry stakeholders can leverage ML-based predictive models to anticipate changes in supply and demand, helping them optimise procurement strategies and inventory management.

For importers: The ability of RFR to capture complex trade patterns suggests that businesses should consider machine learning-based predictive analytics to mitigate price volatility risks and improve sourcing strategies.

For supply chain managers: Given the seasonal nature of sunflower oil imports, firms should leverage SARIMA models for short-term operational planning while incorporating ML-based forecasts for strategic inventory management. Integrating external economic indicators into forecasting models can further enhance the resilience of supply chains.

To further refine trade forecasting techniques, future studies should:

Explore hybrid forecasting models that combine the strengths of statistical and machine learning techniques, enhancing predictive accuracy across different trade environments.

Expand the dataset to incorporate geopolitical risk indicators and climate-related factors, enabling a more holistic understanding of external influences on sunflower oil imports.

Conduct cross-country comparative studies to assess the generalizability of ML-based forecasting techniques in different economic and regulatory contexts.

Investigate the integration of deep learning architectures such as LSTMs and transformer models to evaluate their effectiveness in agricultural trade forecasting.

Limitations

While the models provided accurate forecasts, the study's scope is limited by several factors:

Dataset Length: The study uses a 19-year dataset (2004–2023) for forecasting Turkey's sunflower oil imports. While this dataset is substantial, it may still be too short to fully capture long-term trends, structural shifts, or sudden, significant changes in trade patterns. Future studies could utilise a more extended dataset, possibly spanning several decades, to better account for extreme market fluctuations and long-term dynamics.

Exclusion of Geopolitical and Trade Policy Factors: The study does not explicitly incorporate geopolitical factors, international trade policies, or global market shocks, which can significantly influence sunflower oil imports. Events such as trade agreements, international sanctions, and political tensions can alter trade flows in ways that traditional forecasting models may not fully capture. Future research could integrate geopolitical risk indicators and trade policy variables to improve predictive accuracy in response to external economic shocks.

Generalizability to Other Countries: This study is specific to Turkey's sunflower oil imports, and its findings may not be directly applicable to other countries with different agricultural production systems, market structures, or regulatory environments. Expanding the analysis to include cross-country comparisons would provide a broader understanding of how these models perform in different economic contexts, particularly in regions with varying degrees of import dependency and trade volatility.

Peer-review:

Externally peer-reviewed

Conflict of interests:

The author has no conflict of interest to declare.

Grant Support:

The author declared that this study has received no financial support.

References

Acet, A. (2022). SVM, NB, KNN, ADABOOST ve Random Forest sınıflandırma algoritmaları kullanılarak meme kanserinin tahmini [Breast cancer prediction using SVM, NB, KNN, ADABOOST, and Random Forest classification algorithms] (Yayımlanmamış yüksek lisans tezi). İnönü Üniversitesi, Kütahya.

Akar, Ö., & Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. Journal of Geodesy and Geoinformation, 1(1), 105–112.

Aydin, K., & Çatuk, C. (2018). The impact of the inward processing regime on the vegetable oil sector in Türkiye. Journal of Institute of Economic Development and Social Researches, 4(15), 745–751.

Bozer. (2022). Situation and forecast (Sunflower 2016/2017). GTHB Agricultural Economics and Policy Development Institute (TEPGE) Publications: 1–36.

Bozer. (2023). Situation and forecast (Sunflower 2016/2017). Agricultural Economics and Policy Development Institute (TEPGE) Publications No. 276, 41 pages.

Burbidge, R., Trotter, M., Buxton, B., & Holden, S. (2001). Drug design by machine learning: Support vector machines for pharmaceutical data analysis. Computers & Chemistry, 26(1), 5–14.

Cançelik, M. (2021). A research on safflower (Carthamus tinctorius L.), the vegetable oil market, and the future of safflower in Turkey. In Current marketing studies and digital developments (pp. 135–149).

Champolivier, L., Debaeke, P., Thiard, J., & Thibierge, J. (2012). An evaluation of sunflower production strategies in a supplying area of an agricultural cooperative using the simulator COLLECTO. In Proceedings of the 18th International Sunflower Conference, Mar del Plata, Argentina (pp. 742–747).

Chen, H., Su, Z., & Xing, Z. (2024). Forecasting Extreme High Temperatures Based on an ARIMA-LSTM Hybrid Model. In 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA) (pp. 804-809). IEEE.

Doré, T., Clermont-Dauphin, C., Crozat, Y., David, C., Jeuffroy, M.-H., Loyce, C., Makowski, D., Malézieux, E., Meynard, J.-M., & Valantin-Morison, M. (2008). Methodological progress in on-farm regional agronomic diagnosis: A review. Agronomy for Sustainable Development, 28(1), 151–161. https://doi.org/10.1051/agro:2007031

Duru, S. (2024). Türkiye bitkisel yağ dış ticaretinin mevcut durumu ve rekabet gücünün analizi [Current status and competitiveness analysis of Turkey's vegetable oil foreign trade]. Gümrük ve Ticaret Dergisi, 11(35), 25–39.

Faydaoğlu, E., & Sürücüoğlu, M. S. (2011). Geçmişten günümüze tıbbi ve aromatik bitkilerin kullanılması ve ekonomik önemi [Historical and economic significance of medicinal and aromatic plants from past to present]. Kastamonu University Journal of Forestry Faculty, 11(1), 52–67.

Federal Reserve Bank of St. Louis. (n.d.). Global price of sunflower oil [Data set]. Retrieved July 2024, from https://fred.stlouisfed.org/series/PSUNOUSDM

Gunduz, O. (2021). Why is the price of sunflower oil increasing in Türkiye: Short and long-term dynamic analysis. Gaziosmanpaşa Scientific Research Journal, 10(1), 30–48.

Gül, V., Öztürk, E., & Polat, T. (2016). The importance of sunflower to overcome deficiency of vegetable oil in Turkey. Alinteri Journal of Agriculture Science, 30(1), 70–76.

Hall, A. J., Feoli, C., Ingaramo, J., & Balzarini, M. (2013). Gaps between farmer and attainable yields across rainfed sunflower growing regions of Argentina. Field Crops Research, 143, 119–129. https://doi.org/10.1016/j.fcr.2012.05.003

Hatırlı, S. A., Demircan, V., & Aktaş, A. R. (2002). Analysis of sunflower and soybean oil import demand. Akdeniz University Journal of the Faculty of Agriculture, 15(2), 71–79.

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. Remote Sensing of Environment, 233, 111410.

Ihueze, C. C., & Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. Accident Analysis & Prevention, 112, 21–29.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., et al. (2016). Random forests for global and regional crop yield predictions. PLoS One, 11(6), e0156571.

Kaya, T. E., Sezgin, A., Külekçi, M., & Kumbasaroğlu, H. (2023). Dünyada ve Türkiye'de ayçiçeği üretimi ve dış ticaretindeki gelişmeler [Developments in sunflower production and foreign trade in the world and Turkey]. Alinteri Dergisi, 38(1), 8–13..

Kaya, Y. (2016). The situation of sunflower in our country and its future direction. Journal of Field Crops Central Research Institute, 25(Special Issue-2), 322–327.

Kim, N., & Lee, Y.-W. (2016). Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry, and Cartography, 34(4), 383–390.

Kouadri, S., Pande, C. B., Panneerselvam, B., Adham, A., & Ahamad, M. I. (2022). Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. Environmental Science and Pollution Research, 29, 21067–21091. https://doi.org/10.1007/s11356-021-17084-3

Martinez, E. Z., Silva, E. A., & Fabbro, A. L. (2011). A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Revista da Sociedade Brasileira de Medicina Tropical, 44, 436–440.

Mercau, J. L., Sadras, V. O., Satorre, E. H., Messina, C., Balbi, C., Uribelarrea, M., & Hall, A. J. (2001). On-farm assessment of regional and seasonal variation in sunflower yield in Argentina. Agricultural Systems, 67(1), 83–103. https://doi.org/10.1016/S0308-521X(00)00048-2

Onurlubaş, H. E., & Kızılaslan, H. (2007). Türkiye'de bitkisel yağ sanayindeki gelişmeler ve geleceğe yönelik beklentiler [Developments and future prospects in the vegetable oil industry in Turkey] (Publication No. 157). Ankara: Tarım Ekonomisi Araştırma Enstitüsü.

Önder, K., & Şahin, M. (2023). Destekleme priminin ayçiçeği arzı üzerindeki etkisi: ARDL modeli. Kastamonu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 25(1), 313–333. https://doi.org/10.21180/iibfdkastamonu.1222045

Pekcan, V., Yilmaz, M. I., Evci, G., (2022). Oil content determination on sunflower seeds in drought conditions. Journal of Food Processing and Preservation, 46, e15481. https://doi.org/10.1111/jfpp.15481

Piragnolo, M., Masiero, A., & Pirotti, F. (2017). Open source R for applying machine learning to RPAS remote sensing images. Open Geospatial Data, Software and Standards, 2(1)

Pirotti, F., Sunar, F., & Piragnolo, M. (2016). Benchmark of machine learning methods for classification of a Sentinel-2 image. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B7, 335–340.

Semerci, A. Cost analysis of oily sunflower production: The case of Tekirdag province, Turkey. Custos e Agronegocio 2019, 15, 167–191.

Song, X., Xiao, J., Deng, J., Kang, Q., Zhang, Y., & Xu, J. (2016). Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. Medicine, 95(48), e3929. https://doi.org/10.1097/MD.0000000000003929

Tarım ve Orman Bakanlığı, Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü (TEPGE). (2024). Ayçiçeği durum ve tahmin raporu 2024 [Sunflower situation and forecast report 2024] (TEPGE Yayın No: 404). Retrieved from: https://arastirma.tarimorman.gov.tr/tepge/Belgeler/PDF%20Durum-Tahmin%20Raporlar%C4%B1/2024%20Durum-Tahmin%20Raporlar%C4%B1/Ay%C3%A7i%C3%A7e%C4%9Fi%20Durum%20Tahmin%20Raporu%202024-404%20TEPGE.pdf

Tarım ve Orman Bakanlığı, Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü (TEPGE). (2023). Ayçiçeği durum ve tahmin raporu 2023 [Sunflower situation and forecast report 2023] (TEPGE Yayın No: 392). Retrieved from: [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arastirma.tarimorman.gov.tr/tepge/Belgeler/PDF%20Durum-Tahmin%20Raporlar%C4%B1/2023%20Durum-Tahmin%20Raporlar%C4%B1/Ay%C3%A7i%C3%A7e%C4%9Fi%20Durum-Tahmin%20Raporu%202023-392%20TEPGE.pdf]

Tarım ve Orman Bakanlığı. (2024, 24 Ağustos). Çiftçilere bu yıl ödenecek bitkisel üretim destekleri belli oldu [Plant production support payments for farmers announced for this year]. Retrieved from: https://www.tarimorman.gov.tr/Haber/6373/Ciftcilere-Bu-Yil-Odenecek-Bitkisel-Uretim-Destekleri-Belli-Oldu

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research, 1, 211–244. https://doi.org/10.1162/15324430152748236

Türkiye İstatistik Kurumu (TÜİK). (2024, Temmuz 24). Dış ticaret istatistikleri [Foreign trade statistics]. Erişim adresi: https://biruni.tuik.gov.tr/disticaretapp/menu.zul

U.S. Energy Information Administration (EIA). (2024). Short-term energy outlook: Real prices viewer. Retrieved from https://www.eia.gov/outlooks/steo/realprices

Uçum, İ. (2016). ARIMA modeli ile Türkiye soya üretim ve ithalat projeksiyonu [Soybean production and import projection of Turkey using ARIMA model]. Tarım Ekonomisi Araştırmaları Dergisi, 2(1), 24–31.

USDA (2020) United States department of agriculture, oilseeds: world markets and trade, Published in June 2020, Available from: https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/1257bd51b/fj236p26g/oilseeds.pdf.

Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalised portrait method. Automation and Remote Control, 24, 774–780.

Wang, H., Tian, C. W., Wang, W. M., & Luo, X. M. (2018). Time-series analysis of tuberculosis from 2005 to 2017 in China. Epidemiology and Infection, 146, 935–939.

Wang, J., Zhou, Y., Zhuang, L., Shi, L., & Zhang, S. (2023). A model of maritime accidents prediction based on multi-factor time series analysis. Journal of Marine Engineering & Technology, 22(3), 153–165. https://doi.org/10.1080/20464177.2023.2167269

Wu, L., Xing, Y., Yang, K., Li, W., Ren, G., Zhang, D., & Fan, H. (2024). Optimising maise germination forecasts with random forest and data fusion techniques. PeerJ Computer Science, 10, e2468. https://doi.org/10.7717/peerj-cs.2468