Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 24, S. 351-358, Nisan 2021

Derleme Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 24, pp. 351-358, April 2021

Review Article

A Review of Literature on the Quantitative Methods for Olive Yield Forecasting

İnanç Kabasakal[1*], Murat Özaltaş[2]

[1*] Ege University, Faculty of Economics and Administrative Sciences, Department of Business Administration, İzmir, Turkey, (ORCID: 0000-0003-0098-0144), inanc.kabasakal@ege.edu.tr
[2] Ministry of Agriculture and Forestry, Olive Research Institute, Bornova, İzmir, Turkey, (ORCID: 0000-0003-0021-9443), murat.ozaltas@tarimorman.gov.tr

(2nd International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF)-10–12 March 2021)

(**DOI:** 10.31590/ejosat.900368)

ATIF/REFERENCE: Kabasakal, İ. & Özaltaş, M. (2021). A Review of Literature on the Quantitative Methods for Olive Yield Forecasting. *European Journal of Science and Technology*, (24), 351-358.

Abstract

Yield forecasting is a task that provides critical inputs for producers and other stakeholders in agricultural marketing. Various methods have been applied to forecast olive yield in prior studies that primarily analyze datasets involving meteorological and phenological measurements. Our study reviews the prior literature on olive yield forecasting and explores the prominent methods employed in this context. Accordingly, we categorize prior models into two broad groups: pollen index forecasting and olive yield forecasting models. Moreover, our study highlights the popular methods and attributes involved in previous research, and reports the initial findings of the ongoing olive-yield forecasting project held in İzmir, Turkey. Finally, a discussion is presented regarding the techniques utilized and the attributes analyzed in forecasting models.

Keywords: Olive Yield Forecasting, Forecasting Models.

Zeytin Verim Tahmininde Kullanılan Sayısal Modellere İlişkin Bir Literatür Araştırması

Öz

Verim tahmini, tarımsal ürünlerin pazarlanmasında üreticiler ve diğer paydaşlar için önemli girdi sağlamaktadır. Zeytin verim tahmini çalışmalarında çeşitli yöntemler ile meteorolojik ve fenolojik ölçümlere dayalı verilerin incelendiği görülmektedir. Bu çalışmada, zeytin verim tahminleme çalışmalarına ilişkin bir derleme sunularak, bu kapsamda kullanılmış yaygın yöntemler ele alınmaktadır. İncelenen çalışmalar, polen endeksi ve zeytin verimi tahmini odaklı modeller biçiminde iki ana grupta toplanmıştır. Bununla birlikte, sunulan modellerde yaygın olarak kullanılan yöntemler ile incelenen nitelikler ele alınmış, İzmir'de başlayan zeytin verim tahmini projesi kapsamında elde edilen öncü veriler sunulmuştur. Son olarak, zeytin verim tahmini modellerinde incelenen nitelikler ve kullanılan yöntemlere ilişkin değerlendirmelere yer verilmiştir.

Anahtar Kelimeler: Zeytin Verim Tahmini, Tahminleme Modelleri.

[*] Corresponding Author: inanc.kabasakal@ege.edu.tr

1. Introduction

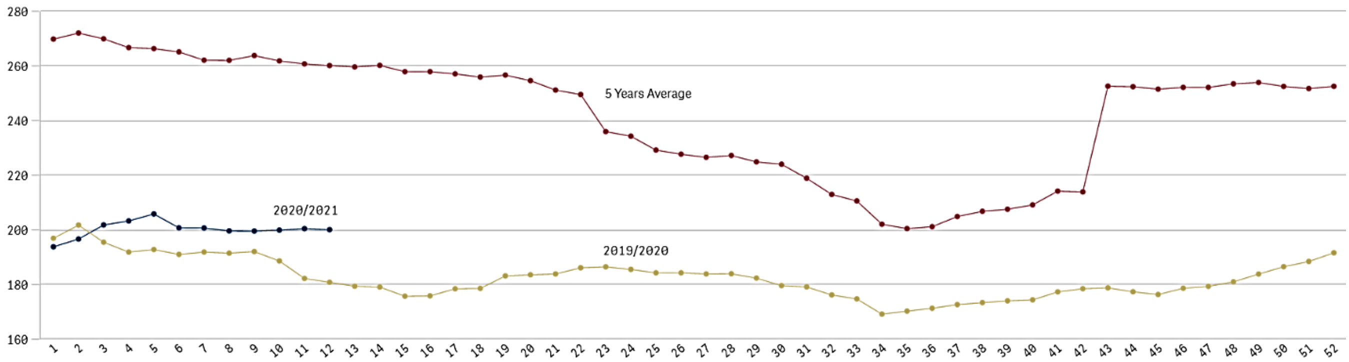
Olive is a valuable product for Mediterranean countries, and olive yield forecasts are essential for both optimization of resources allocated for olive production and planning for the distribution of olive oil (Oteros et al., 2014). Forecasting is important for various managerial decisions in the agriculture sector, including financial planning, pricing, inventory management, and crop rotation optimization (Fornaciari et al., 2005). In a broader perspective, yield forecasts might provide insights about the supply in the future and signal emerging market conditions.

Yield forecasting is a widespread problem in agriculture that relies on data-driven models. Obtaining accurate foresight on yield might signify a shortage or excess of goods in a market. In particular, the forecasting of yield and prices are essential for farmers, agribusiness industries, and governments (Allen, 1994). Crop simulation models are often used for yield forecasting and might provide decision support functions for farmers (Basso et al., 2013). Another prominent forecasting task in agriculture concerns the prices since significant increases or declines might affect multiple stakeholders in supply chains. A recent study by Sabu and Kumar (2020) has developed multiple models to develop arecanut prices and highlighted the better performance of the Long Short-Term Memory networks than the other methods tested.

In a machine-learning study to decide when to spray kiwifruits in New Zealand against leafroller incidences, Hill et al. (2014) highlighted the importance of using technology to reduce data collection costs. Besides, the study utilized various models, including Support Vector Machine, AdaBoost, Naïve Bayes, Random Forest, and Logistics Regression for automated (spray-or-not) decisions. Wireless sensor networks, such as humidity sensors in smart irrigation, are useful in automated agricultural tasks due to their context-awareness, autonomous operating ability, fault-tolerance, and intelligent decision-making capabilities (Ojha et al., 2015:68). In general, research for yield forecasting, including olive, could benefit from that progress over the next years.

The development of olive yield forecasting models can be described as a demanding task that requires research groups to conduct long-term projects. Olive oil is a valuable commodity, and there have been fluctuations in the market price through the years. The graph below demonstrates the virgin olive-oil prices in Spain, which is one of the primary exporters in the global market for olive oil.

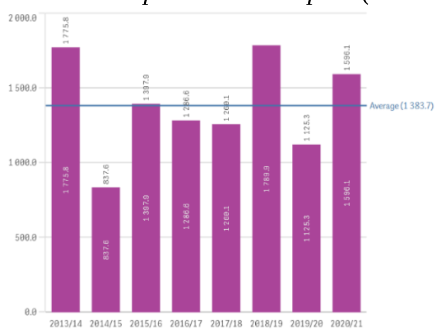
Figure 1 Weekly prices of virgin olive-oil in Spain (€/100kg)



Source: European Commission Agri-Food Data Portal
(https://agridata.ec.europa.eu/extensions/DashboardOliveOil/OliveOilPrices.html)

As in the prices, Figure 2 further demonstrates noticeable fluctuations in total olive oil production in Spain. The decrease in production from 2013/2014 to 2014/2015 was noticeably high (52.8%).

Figure 2 Olive oil production in Spain (x1000 tons)



Source: European Commission Agri-Food Data Portal
(https://agridata.ec.europa.eu/extensions/DashboardOliveOil/OliveOilProduction.html)

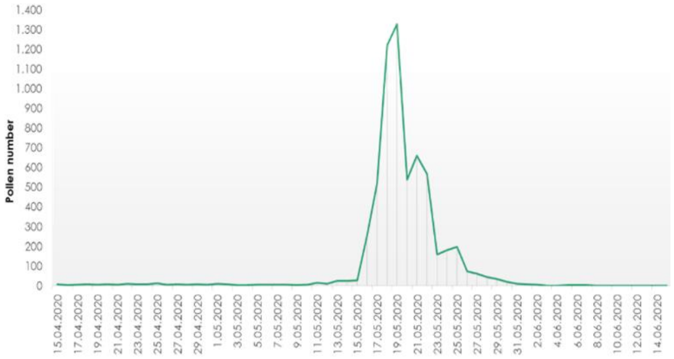
The variations in olive production and the uncertainty in market conditions both necessitate the significance of olive yield forecasting. Besides, accurate forecasts might benefit businesses, producers, and consumers. Forecasts for pollen concentrations are also crucial in medical research, especially for allergy-related studies. Consequently, pollen and olive yield forecasting problems have been receiving intensive attention from various disciplines.

This paper fundamentally aims to investigate the olive yield forecasting methods and extends our prior study (Kabasakal and Özaltaş, 2018) by examining more recent literature. Besides, this study partially presents our preliminary research for our ongoing project to develop an olive yield forecasting model, which is supported by the Turkish Ministry of Agriculture and Forestry. According to the project schedule, pollen concentration measurement has started in April 2020 and still going on by March 2021.

As a further detail, daily meteorological data, including daily measurements for temperature, rainfall, humidity, wind, atmospheric pressure, sunlight hours, and solar radiation, are recorded continuously for our project. Additionally, phenological states are monitored and recorded. Data collection in our project also involves the use of Hirst-type volumetric traps for measuring pollen concentrations, which provides valuable input in olive yield forecasting (García-Mozo, 2011).

Figure 3 depicts the daily airborne pollen concentrations recorded between 15.04.2020 and 14.06.2020 in the pilot region located in Kemalpaşa, İzmir, Turkey.

Figure 3 Olive Pollen Calendar in İzmir, Turkey



Before presenting our review, we note that most olive yield studies aim to provide forecasts and future projections, that are also among popular tasks in machine learning. As highlighted by Géron (2019), the success of machine learning applications typically relies on the data as well as the method being employed. In this sense, our review focuses on the attributes being analyzed in olive yield forecasting models, as much as the methods in use. Accordingly, we present and compare distinctive machine learning models and identify the differences noticed regarding the variables being utilized.

2. Material and Method

A variety of studies have been noted where forecasting models were developed towards significant problems in agriculture. The forecasting of prices and total yield was popular in this context, as well as the prediction of events for early warning systems. Such systems utilize classification and regression models to predict and avoid diseases (Corrales et al., 2015:210) and provide warnings for decision-makers. Yield forecasting techniques are applied for a variety of agricultural products, including olive.

Olive is an anemophilous species and has well adapted to wind pollination. The airborne pollen releasing from olive trees into the atmosphere is among important phenological data (García-Mozo, 2011), often exploited in olive yield forecasting. Osborne et al. (2000) conducted a study that links thermal factors with olive flowering dates in the Mediterranean region. Moreover, the authors highlighted that pollen concentration measurements could be used as a climate-warming indicator.

Another use case for airborne pollen data is the prediction of dates for olive phenophases. A study by Mancuso et al. (2002) developed an artificial neural network to analyze climatic data towards olive phenophase prediction. Pollen-based studies inherently depend on the collection of airborne pollens, even from multiple regions. Moreover, this might be considered a challenge that leads to costs for acquiring and running specialized equipment, and even limits the number of projects on olive yield forecasting. Before presenting prior studies, our study addresses several challenges regarding olive yield forecasting models.

2.1. Challenges in Olive Yield Forecasting

Intuitively, a problem such as olive yield forecasting could be related tightly to environmental variables. Besides, differences might arise from the characteristics of the area for which fruit yield is to be forecasted. Furthermore, even in adjacent locations, data might differ due to specific conditions such as exposure to sunlight, micro-climate effects, elevation, and the type of soil. The variety of data involved in the various models for yield forecasting might show significant differences due to the constraints of the study as well. In addition, model development has its own difficulties to be handled. As highlighted by Oteros et al. (2013:313), developing a method that obtains a proper Pollen Index indicator is another challenge towards yield forecasting.

Another obstacle for yield forecasting involves the tasks and costs required for obtaining and using specialized equipment. Most forecasting models depend on airborne pollen measurements, and pollen data is collected by specific traps. Besides, routine tasks such as maintenance and calibration are occasionally required for continuous operation. Accordingly, olive yield forecasting studies require funds often granted with projects that have time and budget constraints. Moreover, the availability and the number of technical equipment might differ, as well as the staff hired to accomplish the controls, maintenance, and calibration required for seamless operation. Therefore, obtaining extensive volumes of recordings in data collection might be constrained by such factors in olive yield forecasting projects.

2.2. Forecasting Models

In our literature inquiry, we covered studies that relate to olive yield forecasting. Specifically, pollen forecasting models are reviewed as well as yield forecasting models. Pollen forecasting models primarily focus on forecasting the number of airborne pollens before the spring. Such studies utilize meteorological variables and provide a pollen index. Forecasting of olive yield also utilizes a similar subset of variables, mostly aggregated from various meteorological data. Additionally, olive yield forecasting models use pollen forecasts obtained before.

In this section, we summarize the prominent studies for olive yield forecasting and pollen forecasting, respectively. Subsequently, the next section will highlight the differences in models regarding the methods and attributes utilized.

2.2.1. Olive Yield Forecasting Models

Olive yield forecasting can be considered as one of the crop yield forecasting problems in prior studies. To generalize, agrometeorological data is an essential input for crop yield forecasting, and there are numerous parameters to be used accordingly.

Statistical regression with agrometeorological inputs is prominently used in crop forecasting (Basso et al., 2013:10). The study by Çolakoğlu and Tunalıoğlu (2010) presented a multiple regression model to analyze meteorological data for olive yield forecasting. As in the wheat yield forecasting studies by Şimşek et al. (2007) and Yildirak et al. (2015), regression models in crop yield forecasting often include additional indicators obtained regarding the phenological periods of the crop. The study by Cunha et al. (2016) aimed to forecast the volume of wine production in Alentejo, Portugal, and proposed a pollen-based regression model that achieved 81% accuracy.

The most common olive yield forecasting technique until the 1960s was plot census, which relied on a limited set of observations to predict the total olive production in a large region (García-Mozo, 2011:3; Dhiab et al., 2017:541). However, several drawbacks of this approach, including high costs and observer subjectivity, were highlighted by several scholars. García-Mozo (2011) and Dhiab et al. (2017) noticed this issue and underlined that airborne pollen models come forward due to their focus on quantitative data obtained through actual measurements.

In airborne pollen models, an indicator that signifies the pollen count is generated through a straightforward process that involves collecting and counting the pollens. The most common device for data collection is the Hirst-type spore trap; moreover, there are recent developments that pave the way for online data collection with the utilization of new technology, including image recognition, bio-molecular analysis, chemical identification (Oteros et al., 2015:159). The progress in airborne pollen measurement and pollen index monitoring has importance for other disciplines as well. For instance, the study by Buters et al. (2018) highlights the recent developments in pollen monitoring technology within the context of allergy and public health.

As a frontier in the olive yield and pollen index forecasting studies, Galán et al. (2004) analyzed over 20-years of data that involved metrological measurements, pollen concentrations, and other variables derived considering the olive tree phenology. The findings of the study confirmed that pollen emissions are useful predictors of olive yield, approximately eight months before the harvest. The authors proposed multiple regression models, in which the dependent variable is the olive yield, and the independent variables consist of Pollen Index (PI), and monthly rainfall & temperature data recorded over the flowering period and before the harvest. Moreover, the results reported involved multivariate regression models developed at different times. The independent variables involved in those models consist of pollen index as well as various periodical statistics of various meteorological measurements, which are briefly demonstrated in Table 1.

Table 1. Inputs of olive yield forecasting regression models developed by Galán et al. (2004)

Model	Date	Independent Variables
1	July	- Pollen Index - Rainfall amount in May - The minimum temperature in May - The minimum temperature in June - The maximum temperature in June
2	November	- Pollen Index - Rainfall amount in May - The minimum temperature in October - Rainfall amount in October - Rainfall amount in July - The maximum temperature in October - The minimum temperature in July
3	January	- Pollen Index - Rainfall amount in May - The minimum temperature in October - Rainfall amount in October - Rainfall amount in July - The maximum temperature in October - The maximum temperature in December - Rainfall amount in June - The maximum temperature in June - Rainfall amount in September - Rainfall amount in November - The maximum temperature in November - The minimum temperature in September

Galán et al. (2004:46) defined the Pollen Index as the peak pollen concentration among the daily samples. As the results in Table 1 suggest, this attribute is an independent variable involved in the fruit yield forecast model, with a positive correlation coefficient as noted by the authors.

More recently, Oteros et al. (2014) developed a model based on the partial least squares regression method to predict the olive yield. The independent variables of the regression model involve temperatures, humidity, wind, rainfall, and daily solar radiation data. The authors reported that increasing water availability in spring and higher pollen index positively affected the total yield in the areas examined. It was noted that regression models included variables derived by periodic averages of meteorological data, as in the study conducted by Galán et al. (2004).

Orlandi et al. (2016) calculated aerobiological indicators regarding pollination and forecasted olive yields for three areas selected from Spain, Italy, and Tunisia; and proposed olive yield forecasts with the use of multiple regression analysis methods.

Another notable study for olive yield forecasting has been reported by Dhiab et al. (2017) in Tunisia. The model proposed depends on the partial least squares regression technique for analyzing independent variables derived from phenoclimatic measurements. The attributes involve cumulatives derived from measurements of temperature, rainfall, and humidity regarding the olive phenology. Besides, the authors calculated the indices listed in Table 2 that are among the inputs for the regression model:

Table 2. Several independent variables included in the regression model proposed by Dhiab et al. (2017)

Independent Variable	Description
Growing Degree Days (GDD)	A variable derived as a degree of accumulated heat requirements. Calculated as: $GDD = \sum[(T_{max} + T_{min})/2] - T_b$ where: T_{max}: $Maximum\ daily\ temperature$ T_{min}: $Minimum\ daily\ temperature$ T_b: $Respective\ threshold\ for\ the\ site$

Chilling Units (CHU)	A variable derived as a degree of accumulated cold requirements, calculated using the method presented by Crossa-Raynaud (1955).
Hot Days During Fruit Development	Count of days, when the average daily temperatures were higher than the average of maximum temperatures during the fruit development phase
Days for Optimal Pollen Grain Germination	Total days in the flowering period where average temperature is higher than 25°C.

In an earlier study, Fornaciari et al. (2005) also analyzed independent variables, including biological and meteorological parameters such as GDD, CU, daily pollen grain counts. The data was collected from Perugia, Italy, and the authors proposed a linear regression model to forecast olive yield based on the input variables mentioned.

A major element in crop yield forecasting was the use of pollen emission data, which had been pointed as a reliable indicator in relatively early studies of Galán et al. (2004) and Fornaciari et al. (2005). In this manner, forecasting of airborne pollen is also a problem highly relevant to yield forecasting. Accordingly, the next subsection reviews airborne pollen forecasting models in prior research.

2.2.2. Airborne Pollen Forecasting Models

Airborne pollen studies have been widely conducted by researchers of several fields, including aerobiology, agriculture, allergy, and medicine. The pollen measurements are valuable datasets for such domains, where significant trends and seasonal patterns are organized into specific calendars (Camacho et al., 2020) for decision-makers and researchers. Pollen measurements are helpful in building yield forecasting models for fruits.

Various researchers have investigated seasonal pollen levels and published regional pollen calendars. Bishan et al. (2020) published the pollen concentrations in Zhanjiang, South China, and reported the correlation among pollen measurements with several factors, including temperature and relative humidity. Similarly, there have been publications reporting pollen measurements in various regions of Turkey such as Aydın, Muğla, Manisa, (Ulaş and Güvensen, 2019; Güvensen et al., 2020), İstanbul (Celenk et al., 2010) and Antalya (Tosunoglu et al., 2015). O*lea* was among the species that were examined in those publications. It was also noted that the use of volumetric pollen traps was common in those studies.

In addition to pollen measurements, calendars, and regional reports, various publications have presented pollen index forecasts, particularly for olea pollens. A recent study by Tseng et al. (2020) utilized a hidden Markov model for pollen index forecasting that depends on observations of meteorological and biological data that were collected in two decades (1996-2015). An important study based on multivariate regression was publicized by Oteros et al. (2013) to forecast pollen intensity. The authors presented four indices used in their pollen index forecasting model:

- Thermal Index: calculated with the division of minimum temperatures recorded by the temperature range.
- Pre-Flowering Hydric Index: calculated as the total rainfall in February and March.
- Dormancy Hydric Index: calculated as the division of total rainfall by the sunlight hours, both recorded from September to January.
- Summer Index: calculated with a multiplication of the minimum temperature during the period between June and September, and the mean of hours with sunlight from June to August.

As described by the authors (Oteros et al., 2013), the independent variables of the forecasting model were constructed from prior measurements. It should also be noted that selection of attributes in forecasting models involved domain expertise, since the dates in those indices were carefully selected regarding the phenological states of olive. This study was also found noticeable due to the coverage of the dataset analyzed since the measurements for meteorological factors and pollens had been recorded for 29 years in Córdoba, Spain.

García-Mozo et al. (2014) also analyzed an extensive dataset that involved pollen measurements recorded between 1982-2011. The authors proposed ARIMA models to demonstrate long-term trends for pollen-related phenological attributes, including the pollen index. Furthermore, based on a significant rise of airborne pollens in Spain, García-Mozo et al. (2016) analyzed the daily pollen counts, obtained the trend component of airborne pollen data using seasonal-trend decomposition over time-series data, and proposed a linear regression model over the regular pollen trend data along with external variables. The independent variables included in the linear regression model consist of annual, quarterly, and daily temperatures with mean and peak values as well as rainfall data organized similarly.

Another notable study by Galán et al. (2016) stepped forward by identifying airborne pollen trends in 12 locations in Iberia and analyzing the data collected over two decades. Linear regression analysis was used in forecasting, where the results depicted trend lines for the pollen indices regarding each of 12 locations individually.

Based on the projects and publications mentioned so far, we could argue that pollen index forecasting models are usually time-series analyses that analyze actual measurements as well as research-specific indices aggregated from data. As noticed by Tseng et al. (2020), those publications dominantly utilized regression models in pollen forecasting.

The notable studies in our review that focus on either forecasting of olive yield or olive pollen index have been presented in Table 3:

Table 3. A summary of forecasting models regarding airborne pollens and olive fruit yield.

Authors	**Forecasted Variable**	**Methods**
Galán et al. (2004)	Olive Yield	Multiple Regression
Fornaciari, Orlandi, and Romano (2005)	Olive Yield	Linear Regression
Basso, Cammarano, and Carfagna (2013)	Olive Yield	Regression Analysis

Oteros et al. (2013)	Olive Pollen Index	Multivariate Regression
García-Mozo et al. (2014)	Pollen Index for multiple species, including olea	ARIMA
Oteros et al. (2014)	Olive Yield	Partial Least Squares Regression
Orlandi et al. (2016)	Olive Yield	Multiple Regression
Galán et al. (2016)	Olive Pollen Index	Linear Trendline Model
Dhiab et al. (2017)	Olive Yield	Partial Least Squares Regression
Merkoci, Hasimi, and Dvorani (2019)	Olive Yield	Multiple Regression

3. Results and Discussion

After the literature inquiry, it can be noticed that most studies follow two main approaches: olive yield forecasting and airborne pollen forecasting. Considering the objective of our study, both approaches are aligned with the objective of fruit yield forecasting. Despite the difference of focus with olive yield forecasting models, airborne pollen forecasting models are still linked to the problem since they obtain pollen indices serving as crucial indicators for olive yield forecasting.

A noticeable difference in both groups of methods is the schedule of practice: airborne pollen forecasting is conducted a few months before the pollination, whereas olive yield forecasting is conducted after the pollen samples are evaluated. In this regard, we underline that both groups of models might be described as complementary, rather than alternatives in olive yield forecasting.

The quantitative methods adopted in both approaches often depend on regression models. Among the models investigated, we noted that independent variables derived from temperature, rainfall, wind, humidity, and sunlight measurements were prominently included in the forecasting models. Besides, aggregated data such as monthly averages, minimum or maximum values, or total days between specific occasions are occasionally used involved as datasets. As in the variables demonstrated in Table 1 and Table 2, yield-forecasting models make use of indices calculated from meteorological data organized considering olive phenology. The existence of pollen index forecasts also provides critical input for yield forecasting models and enables obtaining forecasts before actual measurements are available.

A notable remark in our review was the scope of data collection in forecasting studies. Several highly popular publications mentioned in our review (Galán et al., 2004; Oteros et al., 2013; Tseng et al., 2020) have covered large datasets, including measurements for various attributes recorded since 1980s. Mainly, regions from Spain and Portugal were the hotspots for olive yield forecasting studies that benefit from such volumes of data. Obviously, the availability of long-term data is a requirement, and a significant advantage in developing both yield and pollen forecasting models with high accuracy.

4. Conclusions and Recommendations

In this study, prior research on olive yield forecasting has been reviewed with a focus on the models proposed for that purpose. After reviewing the literature on such models, prominent studies have been listed with their methodology and output.

As noted in the discussion section, prior research involves forecasting models focusing on olive yields and airborne pollen concentrations. Our study included and categorized those studies as two groups: olive yield forecasting models and pollen concentration forecasting models. The latter models provide indirect but useful results since the pollen index is usually regarded as a critical indicator for olive yield. Apart from the latter group, the olive yield forecasting models ultimately utilized attributes that represent pollen concentration, such as the pollen index. It was noted that both groups of forecasting studies mostly relied on multiple regression techniques.

The forecasting models reviewed primarily used pollen-based variables as inputs. Accordingly, we should note that the availability of accurate and sufficient pollen measurements is vital in olive yield forecasting. Besides, the prominent publications in olive yield forecasting have been possible after long-term projects that enable conducting pollen measurements for decades. The pollen datasets are also useful for research groups from biology, medicine, environmental studies, and other disciplines.

From a managerial perspective, we underline that accurate crop yield forecasts are valuable inputs for decision-makers, both in businesses and governments. Such insights are obtained by researchers and analysts who develop statistical models with the use of data accumulated over many years. Moreover, the availability of large datasets is a requirement to obtain forecasts through statistical models and machine-learning applications. In this regard, the emergence of prominent studies from Spain was no surprise; because of the existence of numerous recordings accumulated for several decades. Accordingly, the significance of olive yield forecasting is obvious for Turkey, which is also one of the major global olive oil producers.

As a final note, we would like to report the current progress in the olive-yield forecasting project by the Olive Research Institute, İzmir, Turkey. Both authors are involved in the project, and Hirst-type volumetric traps are used for data collection since April 2020. A depiction of pollen concentrations recorded in the spring of 2020 was presented above in Figure 3. Besides, there have been promising developments recently, including the progress in sensors and mobile technologies on the route to develop olive-yield forecasting models. The use of smart sensors provides substantial opportunities in agriculture (De Clercq et al., 2018) by facilitating real-time data collection and processing. Remarkably, the availability of image recognition in automatic pollen monitors and new instruments extend the capability to capture pollen measurements with timely peak recordings (Buters et al., 2018). In this regard, machine learning models might be promising in crop yield forecasting, especially when larger volumes of data are available with a high number of attributes (Paudel et al., 2021). Accordingly, further research might be beneficial to develop forecasting models with the use of sensors and machine learning in coordination with researchers, domain experts, and IT specialists.

5. Acknowledge

This study has been conducted as a part of the project (Project-No: TAGEM/TSKA D/B/18/A9/P6/1246) entitled “Exploring the relations between pollen concentration and meteorological factors in olive yield forecasting (Zeytin Verim Tahmininde Polen Konsantrasyonu ve Bazı İklim Verileri Arasındaki İlişkilerin Belirlenmesi)" funded by the Turkish Ministry of Agriculture and Forestry.

References

Allen, P.G. (1994). Economic Forecasting in Agriculture. *International Journal of Forecasting*, 10, 81-135.

Basso, B., Cammarano, D., Carfagna, E. (2013). Review of crop yield forecasting methods and early warning systems. *Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics*, FAO Headquarters, Rome, Italy, 18-19.

Bishan, C., Bing, L., Chixin, C., Junxia, S., Shulin, Z., Cailang, L., Siqiao, Y., Chuanxiu, L. (2020). Relationship between airborne pollen assemblages and major meteorological parameters in Zhanjiang, South China. *Plos one*, 15(10), 1-17. https://doi.org/10.1371/journal.pone.0240160

Buters, J., Schmidt-Weber, C., Oteros, J. (2018). Next generation pollen monitoring and dissemination. *Allergy*, 73, 1944-1945. https://doi.org/10.1111/all.13585

Camacho, I., Caeiro, E., Nunes, C., Morais-Almeida, M. (2020). Airborne pollen calendar of Portugal: a 15-year survey (2002–2017). *Allergologia et immunopathologia*, 48(2), 194-201. https://doi.org/10.1016/j.aller.2019.06.012

Celenk, S., Bicakci, A., Tamay, Z., Guler, N., Altunoglu, M. K., Canitez, Y., Malyer, H., Sapan, N., Ones, U. (2010). Airborne pollen in European and Asian parts of Istanbul. *Environmental monitoring and assessment*, 164(1), 391-402. https://doi.org/10.1007/s10661-009-0901-1

Corrales, D. C., Corrales, J. C., Figueroa-Casas, A. (2015). Towards detecting crop diseases and pest by supervised learning, *Ingeniería y Universidad*, 19(1), 207-228. https://doi.org/10.11144/Javeriana.iyu19-1.tdcd

Crossa-Raynaud, P. (1955). Effets des hivers doux sur le comportement des arbres fruitiers à feuilles caduques: Observations faites en Tunisie à la suite de l'hiver 1954-1955. *Impr. La Rapide*, 1-22.

Cunha, M., Ribeiro, H., Abreu, I. (2016). Pollen-based predictive modelling of wine production: application to an arid region. *European Journal of Agranomy*, 73, 42-54. https://doi.org/10.1016/j.eja.2015.10.088

Çolakoğlu, C.A., Tunalıoğlu, R. (2010). Determination of Relationship between Climate Data and Olive Production Data in Aydın Province. *Journal of Adnan Menderes University Agricultural Faculty*, 7(1), 71-77.

De Clercq, M., Vats, A., Biel, A. (2009). Agriculture 4.0: The Future of Farming Technology. 2018 World Government Summit Report. Source: https://www.worldgovernmentsummit.org/api/publications/document?id=95df8ac4-e97c-6578-b2f8-ff0000a7ddb6

Dhiab, A. B., Mimoun, M. B., Oteros, J., Garcia-Mozo, H., Domínguez-Vilches, E., Galán, C., Abichou, M., Msallem, M. (2017). Modeling olive-crop forecasting in Tunisia. *Theoretical and applied climatology*, 128(3-4), 541-549. https://doi.org/10.1007/s00704-015-1726-1

European Commission (2020). “Agri-Food Data Portal”, Available: https://agridata.ec.europa.eu/extensions/DataPortal/home.html, Data Accessed: 07.02.2021.

Fornaciari, M., Orlandi, F., Romano, B. (2005). Yield forecasting for olive trees: a new approach in a historical series (Umbria, Central Italy). *Agronomy Journal*, 97(6), 1537-1542. https://doi.org/10.2134/agronj2005.0067

Galán, C., Vázquez, L., Garcia-Mozo, H., Dominguez, E. (2004). Forecasting Olive (Olea Europaea) Crop Yield Based on Pollen Emission. *Field Crops Research*, 86(1), 43-51.

Galán, C., Alcázar, P., Oteros, J., García-Mozo, H., Aira, M. J., Belmonte, J., de la Guardia, C.D., Fernández-González, D., Gutierrez-Bustillo, M., Moreno-Grau, S., Pérez-Badia, R., Rodríguez-Rajo, J., Ruiz-Valenzuela, L., Tormo, R., Trigo, M.M., Domínguez-Vilches, E. (2016). Airborne pollen trends in the Iberian Peninsula. *Science of the Total Environment*, 550, 53-59. https://doi.org/10.1016/j.scitotenv.2016.01.069

García-Mozo, H. (2011). The use of aerobiological data on agronomical studies. *Annals of Agricultural and Environmental Medicine*, 18(1), 1-6.

García-Mozo, H., Yaezel, L., Oteros, J., Galán, C. (2014). Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. *Science of the Total Environment*, 473, 103-109. http://dx.doi.org/10.1016/j.scitotenv.2013.11.142

García-Mozo, H., Oteros, J. A., Galán, C. (2016). Impact of land cover changes and climate on the main airborne pollen types in Southern Spain. *Science of the Total Environment*, 548, 221-228. https://doi.org/10.1016/j.scitotenv.2016.01.005

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.), O’Reilly, Canada.

Güvensen, A., Uğuz, U., Altun, T., Eşiz-Dereboylu, A., Şengonca-Tort, N. (2020). Aeropalynological survey in the city center of Aydın (Turkey). *Turkish Journal of Botany*, 44(5), 539-551. https://doi.org/10.3906/bot-1909-38

Hill, M. G., Connolly, P. G., Reutemann, P., Fletcher, D. (2014). The use of data mining to assist crop protection decisions on kiwifruit in New Zealand. *Computers and electronics in agriculture*, 108, 250-257. https://doi.org/10.1016/j.compag.2014.08.011

Kabasakal, İ., Özaltaş, M. (2018). Quantitative Methods for Olive Harvest Prediction: A Classification Based on Prior Research. *Proceedings of the 29th International Scientific - Expert Conference of Agriculture and Food Industry*. September 26-28, Çeşme, İzmir, Turkey, 164-171.

Mancuso, S., Pasquali, G., & Fiorino, P. (2002). Phenology modelling and forecasting in olive (Olea europaea L.) using artificial neural networks. *Advances in Horticultural Science*, 16(3/4), 155-164.

Merkoci, A. L., Hasimi, A., Dvorani, M. (2019). Yield forecasting for olive tree by meteorological factors and pollen emission. *Micro Macro & Mezzo Geo Information*, 12, 7-16.

Ojha, T., Misra, S., Raghuwanshi, N. S. (2015). Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges, *Computers and Electronics in Agriculture*, 118, 66-84. https://doi.org/10.1016/j.compag.2015.08.011

Orlandi, F., Aguilera, F., Galán, C., Msallem, M., Fornaciari, M. (2016). Olive Yields Forecasts and Oil Price Trends in

Mediterranean Areas: A Comprehensive Analysis of the last two Decades. *Experimental Agriculture*, 53(1), 71-83. https://doi.org/10.1017/S0014479716000077

Osborne, C. P., Chuine, I., Viner, D., & Woodward, F. I. (2000). Olive phenology as a sensitive indicator of future climatic warming in the Mediterranean. *Plant, Cell & Environment*, 23(7), 701-710. https://doi.org/10.1046/j.1365-3040.2000.00584.x

Oteros, J., García-Mozo, H., Hervás, C., Galán, C. (2013). Biometeorological and autoregressive indices for predicting olive pollen intensity. *International Journal of Biometeorology*, 57(2), 307-316. https://doi.org/10.1007/s00484-012-0555-5

Oteros, J., Orlandi, F., García-Mozo, H., Aguilera, F., Dhiab, A. B., Bonofiglio, T., Abichou, M., Ruiz-Valenzuela, L., Mar del Trigo, M., Díaz de la Guardia, C., Domínguez-Vilches, E., Msallem, M., Fornaciari, M. (2014). Better Prediction of Mediterranean Olive Production Using Pollen-Based Models. *Agronomy for Sustainable Development*, 34(3), 685-694. https://doi.org/10.1007/s13593-013-0198-x

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., Traidl-Hoffmann, C., Schmidt-Weber, C., Buters, J.T.M. (2015). Automatic and online pollen monitoring. *International Archives of Allergy and Immunology*, 167(3), 158-166. https://doi.org/10.1159/000436968

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016. https://doi.org/10.1016/j.agsy.2020.103016

Sabu, K. M., Kumar, T. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science*, 171, 699-708. https://doi.org/10.1016/j.procs.2020.04.076

Şimşek, O., Mermer, A., Yıldız, H., Özaydın, K. A., Çakmak, B. (2007). AgroMetShell modeli kullanılarak Türkiye'de buğdayın verim tahmini. *Journal of Agricultural Sciences*, 13(03), 299-308.

Tosunoglu, A., Altunoglu, M. K., Bicakci, A., Kilic, O., Gonca, T., Yilmazer, I., Saatcioglu, G., Akkaya, A., Celenk, S., Canitez, Y., Malyer, H., Sapan, N. (2015). Atmospheric pollen concentrations in Antalya, South Turkey. *Aerobiologia*, 31(1), 99-109. https://doi.org/10.1007/s10453-014-9350-6

Tseng, Y. T., Kawashima, S., Kobayashi, S., Takeuchi, S., & Nakamura, K. (2020). Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors. *Science of the Total Environment*, 698, 1-10. https://doi.org/10.1016/j.scitotenv.2019.134246

Ulaş, U., Güvensen, A. (2019). Olea europaea L. polenlerinin Aydın, Manisa ve Muğla atmosferindeki dağılımları. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(3), 936-942. https://doi.org/10.19113/sdufenbed.565330

Yildirak, K., Kalaylıoglu, Z., & Mermer, A. (2015). Bayesian estimation of crop yield function: drought based wheat prediction model for tigem farms. *Environmental and ecological statistics*, 22(4), 693-704. https://doi.org/10.1007/s10651-015-0327-6