Master Thesis in Geographical Information Science nr 179

Olive yield forecasting from remote sensing and climate datasets in the Jaen province (Spain)

Pedro Martinez Duran

2024
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Pedro Martinez Duran (2023). *Olive yield forecasting from remote sensing and climate datasets in the Jaen province (Spain)*
Master's degree thesis, 30/ credits in Master in Geographical Information Science
Department of Physical Geography and Ecosystem Science, Lund University

Olive yield forecasting from remote sensing and climate datasets in the Jaen province (Spain)

Pedro Martinez Duran

Master's degree thesis, 30 credits in Master in Geographical Information Science
Lund University

Supervisors:

Dr. Lars Eklundh
Dr. Hongxiao Jin
Dr. Lanhui Wang

Department of Physical Geography and Ecosystem Science
Lund University

Examination Committee:

Dr. Torbern Tagesson
Dr. Paul Miller

Department of Physical Geography and Ecosystem Science
Lund University

Abstract

The severe decline in olive oil production recorded in the last years is having an economic impact on many Mediterranean countries. The aim of this MSc thesis is to predict the olive yield by means of satellite vegetation indices and meteorological data mostly retrieved by satellite remote sensors. Estimating olive yield production using remote sensing data is reasonably new as few scientific papers focused on this topic so far. Collecting and analysing remote sensing data is cheap and could cover large extensions of olive orchards. It is more important that remote sensing data are usually publicly available, and their analyses do not require extraordinarily complex machinery and installation, apart from powerful computers. In this study several optical sensors have used: the MODIS sensor onboard Terra and Aqua satellites, the optical sensor ETM+ onboard Landsat-7, OLI on Landsat 8, and the MSI onboard on Sentinel-2.

Multiple linear regression technique has been used to evaluate and model for relationship of olive yield (dependent variable) to multiple independent variables, such as vegetation indices, precipitation, temperature, and drought index. The olive yield data were collected from the Survey on Areas and Crop Yields (ESYRCE) Survey. The thesis goal is to estimate which independent variable can best explain the variation in the olive yield in several olive orchards i.e. the dependent variable. The time range considered for the statistical analyses in this study is from November 2001 to November 2020.

The best correlation coefficients of olive yield with 936 monthly parameters (78 parameters for 12 months) were used to run the multilinear regression models. These coefficients corresponded to the vegetation indices and meteorological parameters for all months of the statistical series analysed in this MSc thesis. In this MSc thesis, twenty-four regression models have been run. The best multilinear regression models have a significance level of <0.01 and are those where both meteorological data and vegetation indices parameters have been included, based on the lowest p-value criteria.

The regression models based on the lowest p-values independently of the sensor, generated adjusted R^2 values above 0.85. All regression models show the same pattern: the fewer observations, the higher R^2 values, and lower-resolution climate datasets and MODIS reach lower R^2 values than 0.85. The highest R^2 values correspond to regression models using parameters retrieved from Sentinel-2 data, whereas climate datasets and MODIS sensor data retrieve R^2 values below 0.84.

NDVI and NDMI are the best among 12 tested vegetation indices, suitable for predicting olive yield. The correlation coefficients of NDVI and NDMI from four platforms, Landsat-7, 8, Sentinel-2, and MODIS are more or less similar. For other indices, there is no common vegetation index that can be used for olive yield prediction; ARVI is the best for Sentinel-2, whereas EVI and AVI work well with MODIS, and EVI and GCI gives the highest correlations with Landsat-7.

For climate variables, the water-related ones such as SPEI, topsoil moisture, rainfall, specific humidity, and evapotranspiration clearly show higher correlations than those temperature-dependent. On the other hand, the surface net thermal radiation seems better parameter to assess the impact of thermal stress rather than those temperature-based parameters. In summary, the promising results confirm that vegetation indices and weather data extracted from satellite sensors can be an accurate tool for forecasting olive crop yield on Mediterranean climates.

Keywords: Olive yield, Remote sensing, Meteorological data, Vegetation Indices, Multilinear regression

Table of Contents

List of Figures

List of Tables

1 Introduction, literature review and research questions

1.1 Rationale

In the last years, an important decrease in olive oil production has been recorded all over the world (*International Olive Council*, 2014). Likewise, any crop, oil olive production is always subject to annual variations influenced by factors such as weather conditions, agricultural practices, and economic factors (Anastasiou et al., 2023; Dhiab et al., 2017; Fornaciari et al., 2005; Galán et al., 2008; H. Gao et al., 2023; Idso et al., 1977b; Khan et al., 2018; Merkoci et al., 2019; Sanz-Cortes et al., 2002).

During the last seven harvest seasons, the International Olive Council (IOC) has recorded a steady decline in global olive oil production. This world production dropped 25% from 3.42 million tonnes in 2021/2022 to 2.57 million tonnes in the campaign 2022/2023 (*International Olive Council*, 2014). Moreover, the forecast for the current campaign 2023/24 is expected to be 22% below than the last seven-year average (**Figure 1**).

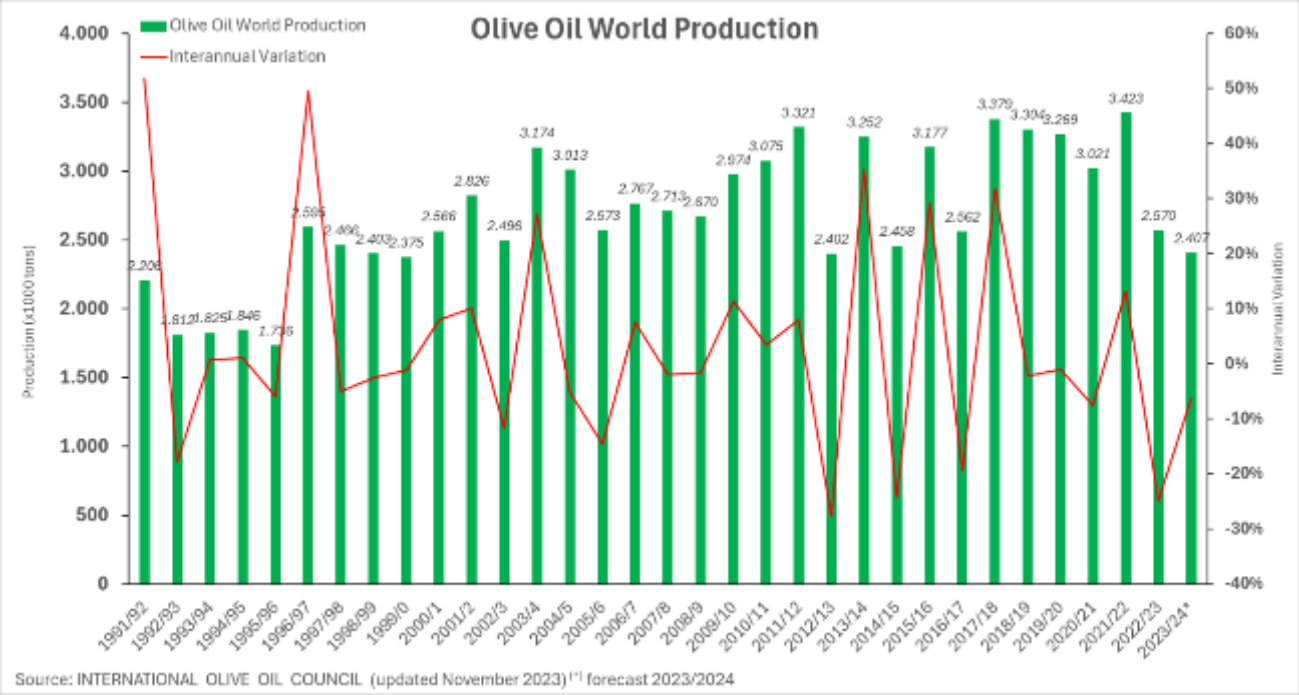


Figure 1 Olive oil world production. Data from the International Olive Oil Council (last update by the International Olive Oil Council in November 2023). The red line shows the interannual variation of the olive oil world production (percentage difference between the previous with the assigned year).

According to the International Olive Council (IOC), Spanish olive oil production accounts for 70% of EU (European Union) production and 45% of the world production, being the largest producer of olive oil in the world. Hence, this scenario of declining olive oil production will have a severe economic impact on the Spanish agri-food industry. Therefore, monitoring olive plantations is economically particularly important as more than 350,000 farmers work on olive growing, the sector supports around 15,000 jobs in the industry and generates more than thirty-two million daily wages per season.

In Spain, the 2022/2023 harvest has been the second lowest (highlighted in orange in **figure 2**) in twenty-one years falling to 780,000 tons, the lowest since the 2012/13 crop year (highlighted in red in **figure 2**). This downward trend is the same for the other countries such as Italy or Tunisa. On the other hand, Greece and Turkey show an increase in the last two years but with the same variability alike the other main producers (**Figure 2**).

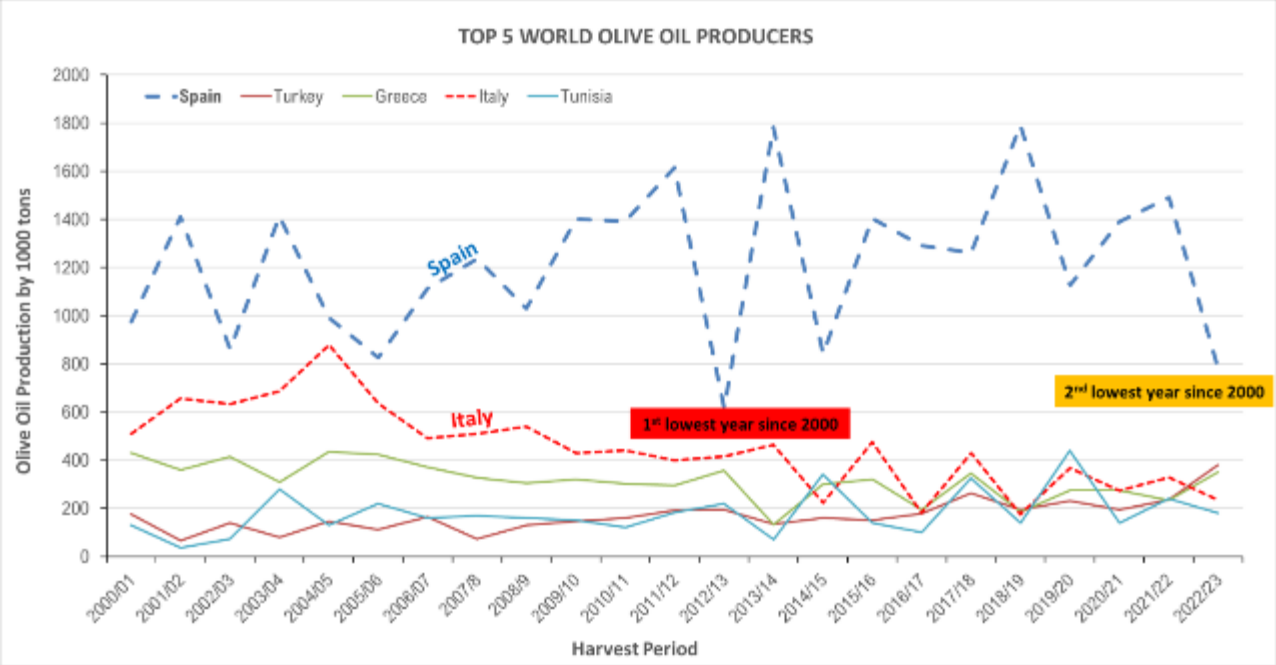


Figure 2 World Top 5 Olive Oil producers (source: International Olive Council)

1.1.1 Why focus specifically on Jaen province?

The work is based on the olive crops of the Jaen province (Andalusia). This southern Spanish province, between latitudes 30° and 45°, is the perfect habitat for olive cultivation because of the dry climate and scorching summer of this Mediterranean region.

45% of the Jaen province is used for olive crops, with more than 600.000 hectares of olive orchards and over 60 million trees. The study area stands for more than 25% of Spain's olive orchards and 42% in Andalusia.

The Jaen province is highlighted with the red polygon in **figure 3**. The 78% of the farming lands in Jaen are olive crops, for the harvest of olives for oil olive production (violet areas in **figure 3**).

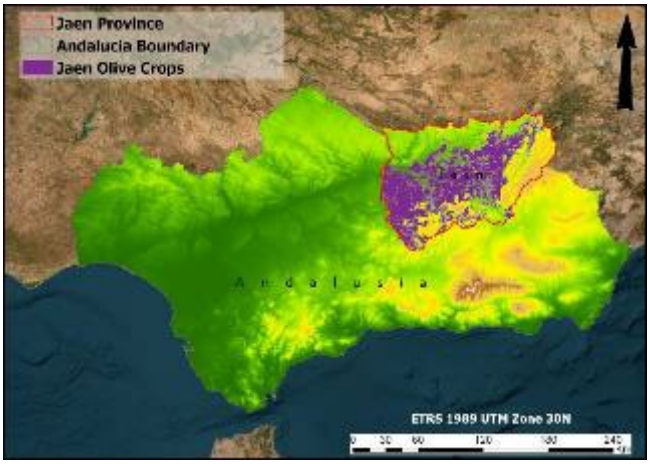


Figure 3 Olive crops highlighted in pink on the Jaen province. Data from the ESYRCE Survey.

Jaen province produces around 50% of Spanish olive oil, corresponding to more than 20% of the world's production. The average production in the province of Jaen exceeds 600.000 tons per year, being a financial value of 300 million euros (Servicio de Estudios y Estadísticas de la Consejería de Agricultura, Pesca, Agua y Desarrollo Rural, 2023). These figures make the Jaen province the largest producer of olive oil globally. Just the province of Jaen produces more than the second largest producer country, Italy, with 20% of the world's olive oil. Therefore, the economic significance of this crop is especially important for Spain and other Mediterranean countries.

1.1.2 Why is olive yield forecasting important?

According to Dawson (2022) and Roux (2023) Olive farmers consider two elements behind this decline in olive production during the last seven years: the first one is the elevated temperatures that coincides with the blossoming season and secondly the consequences of exceptional droughts throughout the whole

Mediterranean area creating water-stress in the crops. These two combined aspects harm in olive trees and subsequently olive oil production drops (Dawson, 2022; Roux, 2023).

Olives are well-known as extraordinary drought-resistant trees, but because of those two considerations olive trees suffer water-stress during the most important stages of fruit development such as sprouting and blooming and consequently trees release fruits to optimise the use of scarce water and defend themselves of this water-stress situation.

This decline is increasing the pressure on the supply, distribution and consequently on the price of olive oil worldwide. Since olive oil is a fundamental element in the cuisine of millions of people, it is particularly important to forecast production to foresee production, minimize supply distortions, and price escalation. In addition, an estimate of production would help to improve the management of agricultural resources such as machinery, labour, storage capacity and ensure an adequate supply to match the demand of olive oil and avoid excess stocks, which are expensive to maintain.

Going further in this analysis, those high-temperatures and long-lasting droughts guilty of the olive yield decline could worsen as they can be the first signals of the climate change. Climate change effects will be especially pronounced in the Mediterranean area as predicted by Sala et al. (2000).

Moreover, the Spanish Meteorological Agency (AEMET) conclude an upward trend in temperatures and potential evapotranspiration and decreasing precipitation. This trend has been progressive throughout the second half of the 20th century with rates of temperature increase of less than 0.2 ºC per decade.

However, in the last years of the of the 20th century and the beginning of the 21st century there have been considerable rates of increase producing significant changes of up to +0.25ºC in average temperature values in the periods 1971-2000 and 1976-2005. Seasonally, the changes have been much more pronounced in spring and summer (**Table 1**).

PRECIPITATION AND MAIN SCENARIO EVOLUTION IN ANDALUSIA					
(mm)	Annual	Winter	Spring	Summer	Autumn
1971-2000	581,87	233,54	148,08	30,90	169,35
1976-2005	579,34	228,97	140,85	26,36	183,26
2011-2040	380,68	158,03	96,14	40,10	86,42
2041-2070	370,49	162,99	86,99	38,43	82,08
2071-2100	346,22	152,79	82,61	40,25	70,56
TEMPERATURE AND MAIN SCENARIO EVOLUTION IN ANDALUSIA					
°Celsius	Annual	Winter	Spring	Summer	Autumn
1971-2000	15,92	8,94	13,96	23,95	16,84
1976-2005	16,17	8,98	14,54	24,29	16,88
2011-2040	18,03	10,84	16,1	25,92	19,25
2041-2070	19,18	11,76	17,01	27,45	20,49
2071-2100	20,46	12,71	18,57	28,82	21,74

Table 1 Rain and Temperature forecast 2100 (AEMET, project DESERNET)

The historical records for Andalusia give a clear message that the climate warming with less rain and more torrential rains coupled with erosion and lose of fertile soil is happening.

In **table 1**, forecast values for Andalusia are shown for successive 30-year periods until 2100. Precipitation values clearly decrease from 581.87 mm for the period 1971-2000 to 346.22 mm by the end of 2100. The opposite trend is depicted for the temperature evolution with a clear increase from an annual temperature average of 16 °C by the end of 2000 to more than 20°C in 2100.

According to the main scenario shown in **table 1**, the climate change impacts in the Andalusian agricultural sector in the second half of this century could be enormous. In the second half of this century, the loss of fertile soil by erosion with desertification would affect the crop yield, though it could be partially compensated with an improvement of the crop techniques and intensive use of fertilizers. However, these solutions are not elastic. In addition, many types of crops will not adapt to the new climatological conditions as fast as climate change will take place in the affected areas (Madueño et al., 2005).

Hence, the loss of fertile soil, erosion, desertification, crops outside their natural climatic habitat could be the perfect storm to trigger a threat to food security. As the economy of Andalusia is highly dependent on the agricultural sector, this scenario will make especially important not only olive yield but any crop type forecasting for companies and growers, but also for millions of consumers around the world.

1.1.3 Why to use remote sensing data?

For crop yield estimate, it is necessary to understand how much sunlight plants receive and how much water crops need. Although, crop growth depends on these two factors, many other factors affect plant growth such as temperature, humidity, and soil type. The quantity and quality of crops produced in the world is a major factor affecting food security. This is especially important in countries where agriculture is still an important part of the economy.

Remote sensing data can be used to estimate quantities of light received by plants and thus predict crop yield and is a powerful tool for estimating crop yield. It provides information on the growing crops and their environment, allowing you to estimate crop production. Therefore, remote sensing techniques can play a significant role to support the oil olive crops worldwide to ease decision-making for increasing productivity. Collecting and analysing remote sensing data is cheap and could cover large extensions of crop lands (country-scale and even continent-scale). Remote sensing data is many times on public domain and their analysis do not require extraordinarily complex machinery and installation. Just powerful computers, software tools and algorithms to manage with satellite imagery and GIS and data analysts with expertise in remote sensing techniques. Hence, remote sensing data is not only useful for crop yield predictions but also for monitoring crops conditions months before harvest.

This tool provides information on the state of crops at specific points in time and space to find issues during the growing months to help changes farming practices and make recommendations based on this information. In other words, remote sensing techniques help decision-making for increasing productivity (Abdul-Jabbar et al., 2023; Ali et al., 2022; Bharadiya et al., 2023; Cubillas et al., 2022; Debaeke et al., 2023; Doraiswamy et al., 2003; Ferencz et al., 2004; Gracia-Romero et al., 2023; Idso et al., 1977b; Ilyas et al., 2023; India, et al., 2020; Jaafar & Ahmad, 2015; Jeong et al., 2023; Ji et al., 2021; Khanal et al., 2020; Li et al., 2007; Lobell, 2013; Muruganantham et al., 2022; Ronchetti et al., 2023; Steven, 1982; X. Zhang et al., 2003).

1.2 Literature review about forecasting olive yield from satellite remote sensing

Numerous approaches exist for estimating crop yields with remote sensing. Most of them rely on vegetation indices computed from remote sensing measurements of light at blue, green, red and near-infrared (NIR) wavelengths and linked through empirical relationships with ground-based yield data (Abbasi et al., 2023; Ali et al., 2022; Amankulova et al., 2023; Bharadiya et al., 2023; Bojanowski et al., 2022; Bolton & Friedl, 2013; Carreño-Conde et al., 2021; R. Chen et al., 2020; H. Gao et al., 2023; Gayke & Rokade, 2021; Gilabert et al., 2002; A. A. Gitelson et al., 2003; Huete, 1988, 2012; Jaafar & Ahmad, 2015; Ji et al., 2021; Jiang et al., 2008; Jiménez-Jiménez et al., 2022; Jurado et al., 2020; Karimli & Selbesoğlu, 2023; Kefi et al., 2016; Kern et al., 2018a, 2018b; Kurucu et al., 2015; Navrozidis et al., 2019; Penuelas et al., 1995; Rondeaux et al., 1996; Solgi et al., 2023; Sousa & Small, 2023; Taloor et al., 2021; Vicente-Serrano et al., 2010; Wall et al., 2008; Zarco-Tejada et al., 2004; L. Zhang et al., 2019; Zhu et al., 2021). Red and near-infrared (NIR) wavelengths are most preferred to estimate crop yield as vegetation, are very reflective in the NIR and absorptive at red wavelengths, and therefore, some combination of the two is a good measure of vegetation vigour (Lobell, 2013).

Hence, focused on vegetation indices, the amount of scientific literature about vegetation indices to predict biomass production, plant health or crop vigour, water stress, yield forecast, and related topics is countless. However, this myriads of papers decreases considerably when the search is focused using the key words “olive yield” or “olive production” and even much smaller if words such as “predict,” “forecast”, or other synonyms are added to the search. For this reason, I tried to screen the most important papers about olive yield prediction as

well as papers covering more general concepts about vegetation indices, methodology, or even comparing papers about yield prediction of different crops.

As a preliminary summary, the search for olive yield forecast using remote sensing and climate data together gives very few records (Anastasiou et al., 2023; Cubillas et al., 2022; Jurado et al., 2020; Torres et al., 2008). This fact is confirmed by the recently published paper (Anastasiou et al., 2023) about trends in remote sensing technologies in olive cultivation. This work gives an excellent description of the status quo of the work published for olive yield forecast.

These authors achieved research for publications about remote sensing and olive crops from the period of January 2001 to December 2021 using several sources such as *Scopus* or Web of Science. Anastasiou et al. (2023) filtered the first results to exclude publications unrelated to the study's aim based on the title and abstract, duplicates, or publications in a language other than English. Finally, the papers that matched the filters were 56 research studies (**Figure 4**). From those 56 papers, 42% used multispectral sensors in olive cultivation (Anastasiou et al., 2023).

I used the work from Anastasiou et al. (2023) to make the first screening of papers focused on remote sensing and climate data to assess or forecast olive crops. As this study is based on multispectral sensors, I discarded from Anastasiou et al. (2023) those papers focused on UAV, UGV and manned flight and ground measurements. Finally, I used the word "yield" to constraint the search and only two papers focused on olive yield production were retrieved:

- Authored by Torres et al. (2008) aimed to describe a software development CLUAS and the information provided by this software.
- Authored by Cubillas et al. (2022) based on meteorological data acquired for a weather station 2km away from the studied orchard.

The paper from Torres et al. (2008) is based on ground-truth data taken in an olive orchard of about 2 ha and remote images with a spatial resolution from 0.25 to 1.5 m but does not correlate and estimate olive yield in time series.

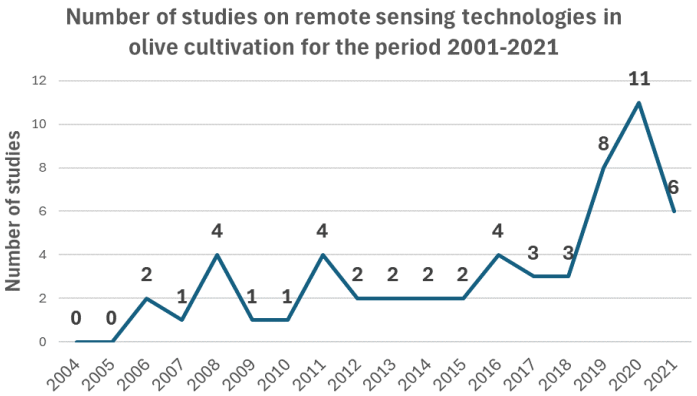


Figure 4 Studies on remote sensing technologies in olive cultivation for the period 2001-2021. Data extracted from Anastasiou et al. (2023)

The last one, the paper from Cubillas et al. (2022) deserves special attention given that it pursues the same aims of this MSc. study, the olive yield prediction in an orchard of the Jaen province. What makes this paper important is that olive yield of this studied orchard is provided by the farmer and this data is published in the article as public domain data. However, Cubillas et al. (2022) only used meteorological data from a weather station owned by the Spanish State Meteorological Agency (AEMET), which is only 2 km from the studied orchard. As the orchard could be localised and yield data published, this paper has been used to validate the regression models proposed in this study.

Thus, apart from using climate data as something not to be missed, the combination with remote sensing data does not retrieve focused on olive yield production using remote sensing data. Hence, this literature review was forced to open the scope to a broader search for more general concepts about vegetation indices, methodology, or even comparing papers about yield prediction of different crops. As I will show in this literature review, the closest research and methodologies to forecast olive crop yields are based on climate data with olive airborne pollen concentrations.

Below, I describe the main or more relevant papers about olive yield prediction near the study area, Andalusia (Spain), mostly based on airborne pollen concentrations. Though challenging, airborne pollen can be estimated using remote sensing technologies linked to high temporal resolution surface observations (Buters et al., 2024), no paper has been found combining olive airborne pollen estimations to predict olive yield.

As shown below, all papers have remarkably similar conclusions. One of the earliest ones is the work by Fernandez-Mensaque et al. (1998). These authors combined airborne olive pollen and accumulated rainfall between 1 September and the following 15 April to obtain four mathematical equations to forecast the crop six months in advance. Authors found that the degree of accuracy in forecasting becomes remarkably high when, besides the pollen counts, the rainfall prior to the period of pollen emission is considered (September to the beginning of blooming at the end of April and May). This is because the fruiting rate depends on the water stress the plant suffers in the months before flowering. In other words, the lack of water reserves in these months increases the number of ovarian abortions following the process of fertilization (Fernandez-Mensaque et al., 1998).

In the same year, Minero et al. (1998) published a similar paper about forecasting olive crop production based on monitoring airborne pollen in Andalusia. The authors showed a positive correlation between production and the amount of pollen, but the length of the main pollination period was found to be negative for production. This is possibly because a short pollination leads to better abscission of the fruits, resulting in a higher production. Additionally, and following the same results from Fernandez-Mensaque et al. (1998), Minero et al. (1998) found that rainfall preceding (1 September±15 April) and following (1 September±30 November) flowering is also related positively to production. Also, the maximum temperatures during the main pollination period (April-May) are negatively related to crop size. This is possible because flowering in the olive is impaired by water stress in the preceding months, and the amount of fruit pulp is less if autumn rainfall before fruit picking is scarce (Rallo, 1994).

Similar conclusions are drawn by Galán et al. (2008). These authors combined olive tree phenology, airborne pollen concentrations, meteorological data, and fruit production data in the province of Cordoba (Andalusia, Spain) over 20 years (1982–2002) to obtain models for predicting fruit production. The pollen emission was shown to be a reliable bio-indicator to forecast olive fruit production up to 8 months in advance, while rainfall in May was the most important meteorological parameter affecting final fruit production. These authors ran three statistical models, with different elapsed times between crop estimation and harvest, were developed: 8, 4 and 2 months, and all showed high determination coefficients (73-98%) with a significance of 99% (Galán et al., 2004).

Galán et al. (2008) extended the work abovementioned, enhancing and enriching their statistical models. The authors assessed that the annual pre-peak pollen index (pollen levels before they reach their peak during a pollen season) was the variable most influencing the final olive crop. On the other hand, spring and summer rainfall and both maximum and minimum temperatures in summer and autumn were the major weather-related parameters affecting final fruit production; however, statistical analysis revealed differences between sites regarding the timing and the degree of their influence (Galán et al., 2008).

García-Mozo et al. (2008) followed the same methodology using atmospheric pollen and meteorological data. The statistical analysis run by those authors showed that the annual pollen index was the variable influencing most of the final olive crop. On the other hand, the maximum temperature in March was the meteorological variable affecting most of the annual olive crop, whereas the rainfall registered in October influences the final fruit production (García-Mozo et al., 2008).

A similar approach is the bioclimatic model developed to forecast olive yield in the Alentejo province using three indices: regional pollen index, plant water requirements, and a phytopathological index (Ribeiro et al., 2009). At the flowering stage, 66% of the olive yield of the study area was able to be explained by the RPI and this forecast improved by 26% by adding the plant water requirements and finally the phytopathological index enhanced the

model another 5%. Therefore, the bioclimatic model justified around 97% of the olive production with a very tiny variation around 4% between the monitored and calculated yield. These authors replicated the same work using only airborne pollen concentrations (Ribeiro et al., 2017). By using a linear regression model, Ribeiro et al. (2017) found that airborne pollen concentrations as the independent variable showed an accuracy of 87% in estimating olives fruit production in Alentejo (Portugal).

Following the same methodology, the work from (Oteros et al., 2014) describes pollen index and water availability during spring as the most significant variables in all models and essential for optimal crop production in the Mediterranean basin. These authors showed that higher pollen indices and water availability during spring are related to an increase in final fruit production with determination coefficients (R^2) around 0.91 for the areas near our study area. Moreover, higher pollen indices positively correlate with air temperature during early spring and autumn. Furthermore, fruit production decreases due to increasing air temperature during winter and summer. As some locations of this study are near our selected orchards, these findings are particularly useful information for this research.

More studies about airborne pollen content and climate data are from Fátima Aguilera & Ruiz-Valenzuela (2013) where they used the annual olive yield as the dependent variable and both aerobiological and meteorological parameters as the independent variables. They found that pollen index and accumulated precipitation were the most important parameters for explaining fluctuations in fruit production.

These findings are in line with the works described previously in this review, especially in the role played by the airborne pollen index and accumulated rainfall until May as the blooming season (Aguilera & Ruiz-Valenzuela, 2013; Fernandez-Mensaque et al., 1998; Galán et al., 2004, 2008; Ribeiro et al., 2017). On the other hand, several papers also focus on how production will be affected by climate change (**Table 1**). An interesting study related to olive yield modelling for future climate scenarios exists. (Viola et al., 2013) Based on a numerical model, it showed significant reductions in olive yield because of rainfall reduction and temperature increase.

Other scientific works focused on olive yield forecast based on airborne pollen and meteorological-related variables but outside of the study area are Fornaciari et al. (2005), Dhiab et al. (2017), Merkoci et al. (2019), and Achmakh et al. (2020) from Italy, Tunisia, Albania, and Morocco, respectively. Thus, as mentioned previously and seen during this literature review, no specific scientific papers focused on forecasting olive yield production have been found using the mix of satellite sensors and climate data.

In a broader sense, there are many studies focused on the study about floral phenology of *Olea europaea L*. Though, remote sensing of flowers is possible to determine final productivity (B. Chen et al., 2019; Dixon et al., 2021; Gonzales et al., 2022; Han et al., 2020), this approach has not been found to the floral phenology of *Olea europaea L* in this literature research.

Regarding floral phenology F. Aguilera & Valenzuela (2009) is one of the most comprehensive studies I could find. These authors show that the beginning of the flowering period depends on altitude, and statistical analyses indicate that the temperature, humidity, cumulative rainfall, and cumulative solar radiation are the meteorological parameters that most affect olive floral phenology (Aguilera & Valenzuela, 2009). Moreover, olive crop responses are constant due to similar meteorological conditions independent of altitude variations (Orlandi et al., 2010).

Slightly opposite to Orlandi et al. (2010),Fátima Aguilera & Valenzuela (2012) advocate for the influence of microclimatic conditions for reproduction in olive trees. These authors found that the most favourable conditions are those with low temperature and high precipitation records during the months prior to flowering of the olive trees. They suggest that olive trees tend to increase their pollen production rate as altitude increases, which can be interpreted as a reproductive strategy to ensure fertilisation (Aguilera & Valenzuela, 2012). The main thoughts extracted after this review are:

- The airborne olive pollen index is the most popular and seems to be the most accurate tool to estimate yields in olive crops. However, remote sensing offer advantages that airborne pollen studies cannot because collecting and analysis remote sensing data is cheap and could cover large extensions at country-scale and even continent-scale (see **1.1.3**).
- Climate data is necessary, but not all climate data or months have the same weight when running correlation or statistical regressions. Some of them could have less predictive power for modelling.
- Multivariate and linear regression is the favourite methodology.
- No remote sensing data have been used so far in studies of olive yield forecasting.

These conclusions agree with the rationale of this work when using climate data and statistical methodology but are also challenging as no papers have exclusively used remote sensing data for the olive yield predictions.

Remote sensing has been used for agricultural applications for over three decades and widely used with most common crops such as wheat, soybeans, corn, or rice to estimate crop yield (Hatfield, 1983; Idso et al., 1977a, 1977b; Knipling, 1970; Steven, 1982; Tucker et al., 1985). Hence, given the overabundance of papers, my strategy for this literature search was based on the number of citations and selecting the most recent articles to give a general overview.

One of the most cited papers is from Bolton & Friedl (2013), using the same vegetation indices and data from Moderate Resolution Imaging Spectroradiometer (MODIS) to develop empirical models predicting maize and soybean yield in the Central United States between 2004 and 2009. The authors found that the Enhanced Vegetation Index (EVI2) better predicted maize yields than the Normalized Difference Vegetation Index (NDVI) between 64-75 days afterward green up season stage of the maize (Bolton & Friedl, 2013).

Also, the information on crop phenology derived from MODIS and the combination of NDWI and EVI2 indices significantly improved the model. The models showed that correlations between vegetation indices and yield were highest 65-75 days after green-up for maize and 80 days after green-up for soybeans (Bolton & Friedl, 2013).

Regarding using satellite data to assess crop yield, the paper from Lobell (2013) gives an exciting overview to understand the difference between yield potential and average yield and what causes these gaps on different crops. The authors claim that combining ancillary ground data with yield maps and remote sensing data provides a unique opportunity to overcome both spatial and temporal scaling challenges and thus improve understanding of crop yield gaps (Lobell, 2013).

Another with a high score of citations is from Doraiswamy et al. (2003) which evaluate a method of integrating satellite imagery data in a crop growth model with climate data to simulate spring wheat crop yields in North Dakota. The sensor is the Landsat TM to calculate NDVI, and the study revealed three optimum periods when remote sensing data were most effective: during the early vegetative phase, flowering, and senescence (Doraiswamy et al., 2003). Also, remote sensing data was accurate in detecting crop water stress conditions as it delays growth and promotes early senescence.

Finally, it is worth mentioning an article combining meteorological data and vegetation indices to forecast corn production at the Jilin and Liaoning Provinces of China (Zhu et al., 2021). The study provides a reference method for selecting modelling variables and offers insights into climate's impact on potential crop yield.

Also, these results show that meteorological indices and remote sensing data contribute significantly to the accuracy of the models with R^2 between 0.73 and 0.82. Authors identified as the most important variables for yield estimation are temperature, precipitation, gross primary production (GPP), Normalized Difference Vegetation Index (NDVI), and standardized precipitation index (Zhu et al., 2021).

1.3 Objectives and purpose of this M.Sc

Based on remote sensing and climate datasets, the purpose of this M.Sc. is to study relationships among vegetation indices and weather data such as precipitation and temperature extracted from satellite sensors to forecast olive crop yield.

Consequently, the two main questions of this MSc. Project are:

- Question: 1. How accurately can olive crop yield be predicted from remote sensing observations using vegetation indices derived from earth observing satellite data?
- Question: 2. By using MODIS, Landsat and Sentine-2 sensors, this study would like to identify which of these sensors is the most accurate to forecast olive crop yield.
- Question: 3. How can crop yield be predicted from the above vegetation indices derived from remote sensing combined with weather data such as monthly average temperature and rainfall, and altitude?

2 Background about phenology and reflectance-based greenness indices

2.1 Phenology, land surface phenology and plant phenology

Phenology is the study of periodic plant and animal life cycles and their variations in relation to the seasons, such as flowering, nesting, and migrations. Then, phenology describes the environmental cycles of living organisms by characterising vegetation changes through time, such as the start and end of a season and the duration of a growing season. These plants' life-cycle processes and changes, such as photosynthesis, CO_2 uptake, and transpiration, strongly impact atmospheric processes, establishing a climate-vegetation-atmosphere relationship (Post et al., 1982; Sellers et al., 1986; Wenzel et al., 2015).

Parameters define phenological cycles, also referred to as phenometrics, which are how to measure or characterise the phenological processes understood as recurring plant and animal life cycles and their variations in relation to the seasons. Therefore, these parameters are normally yearly metrics derived from the study of the seasonal trajectories or periodic plant and animal life cycles mentioned previously.

These parameters describe the behaviours and patterns on the specific stages of the seasonal growth curve (seasonal trajectory), such as the date of the start of the season, the length of the season, or the date of the end of the season (**Figure 5**).

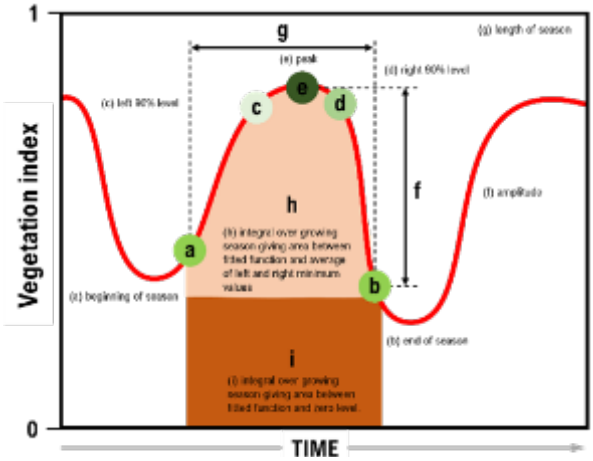


Figure 5 The main phenological parameters

There are many parameters to characterise the phenological cycles, which can vary depending on agencies (Copernicus, USGS, etc), but in general terms, the more common ones are shown in **figure 5**. Consequently, phenology is an excellent tool for unveiling and track trends on farm areas (Helman, 2018).

Phenology uses several sources of information that could be defined as land surface phenology and plant phenology (Helman, 2018). Plant phenology is based on plant observations and describes the periodic life-cycle events of individual species of plants, explaining their growth and maturity periods, such as bud break and the emergence of leaves and flowers. On the other hand, land surface phenology is the study of plant phenology at a regional to global scale measured from data acquired using optical sensors installed on satellite platforms. These remote optical sensors are extremely valuable to reveal regional and even global phenological trends that would be hidden by the naked eye using traditional ground-based methods. Therefore, land surface phenology presents an opportunity for large scale monitoring than the one obtained with only ground-based plant phenology.

2.2 Remote sensing phenology

Remote sensing techniques collect data from satellite sensors that measure wavelengths of light absorbed and reflected by vegetation. Multispectral optical sensors are multichannel detectors with few spectral bands. Each channel is sensitive to radiation within a narrow wavelength band such as the blue, green, red, near infrared, or short-wave infrared band.

Leaves reflect wavelengths of green and near-infrared light; this last one is invisible to human eyes but can be recorded by satellite remote sensing sensors. The reflectance intensity of these wavelengths changes as plant canopies changes from early spring growth to late-season maturity, so remote sensing sensors record small reflectance variations.

Remote sensing phenology is when satellite sensors are used to track phenological events to back, underpin, and complement ground observations by recording the reflectance variations. Hence, remote sensing phenology is paramount for tracking climate change as phenology changes over large areas are challenging to detect by traditional ground methods. Though remote sensing sensors are paramount to phenology, there are some limitations or caveats to bear in mind when using data from remote sensing sensors. (Helman, 2018) describes those limitations and warns that those satellite-related limitations may lead to misinterpretation of land surface phenology observations if not considered (see chapter **Influence of the Mediterranean climate on the phenology of olive trees 3**).

Apart from clouds, aerosols, sensor drift, sensor sun viewing geometry, one of the most important ones is related to the spatial resolution of the sensors. The lower the spatial resolution, the bigger the pixel (more ground area covered); sometimes, this could be a promising approach for regional or large areas. In contrast, the higher the resolution, the smaller the pixel or ground area covered, which means that the reflectance captured by the pixel comes from a smaller area. Hence, the smaller the area, the fewer species are within the pixel; therefore, there is less room for confusion.

The spatial resolution is paramount not only to capture detailed features but also to know exactly what you are collecting to analyse only what you think you are seeing. This limitation means that the variability of species (or a mosaic of plant species as Helman describes) and its subsequent different composition within a single pixel could change the signal because of the different phenology of all species involved within the same pixel. Hence, the bigger the pixel, we must expect a mixed signal due to the ground diversity covered by a big pixel.

On the other hand, the smaller the pixel, the more it is expected to reduce the ground diversity, and accordingly, the signal would be less mixed (Helman, 2018). Consequently, the limitation of working with spatial low-resolution images is that it creates more mixed signal responses that can be misinterpreted as the phenological change of a single sort of plant, which is associated with changes in all the species' phenology within a single pixel. In other cases, changes may be real but could correspond to changes in the earliest species within the pixel.

Furthermore, the reason for the mixture of signals could be from vegetated and non-vegetated areas such as bare soils and/or waterbodies. Also, the issue described by (Helman, 2018) is interesting about the whiskbroom MODIS scanner with a different spatial inter-pixel correlation causing a “pixel-shift” resulting in changes in the real coverage area of each pixel. Another important and tricky limitation is dealing with complex systems with multi-canopy layers.

2.2.1 Reflectance-based greenness indices

Reflectance-based greenness indices are derived from the response of leaves and how they intensely reflect wavelengths of green and near-infrared light recorded by optical sensors. The reflectance-based greenness indices are based on how plants interact with solar radiation during photosynthesis and other pigment-related activity in the leaf tissues as not all solar radiation energy is beneficial for the plant life cycle. The reflection of solar radiation by vegetation displays low values in the blue and red bands, with slightly higher values in the green band, and remarkably high values in the near infrared band (**Figure 6**). This is one of the most striking characteristics for the vegetation spectrum, the steep increase of reflectance from red to near infrared (**Figure 6**).

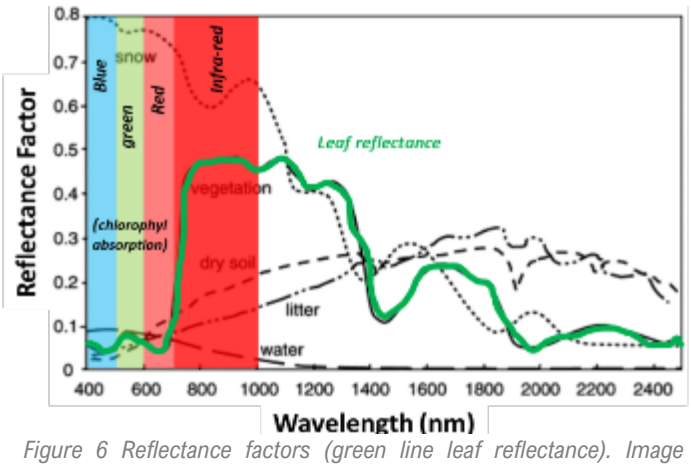


Figure 6 Reflectance factors (green line leaf reflectance). Image modified from A. R. Huete (2004).

The reason is because the light spectrum used by plants for photosynthesis is constrained by a spectral range, which is the portion of the light spectrum used for photosynthesis. This range is known as photosynthetic active range (McCree, 1972).

In general, green leaves absorb light within the red spectrum, with maximum absorption near 620 nm but those leaves reflect and transmit light in the near infra-red band around 800 nm. The near infra-red band is ineffective for photosynthesis and damages the plant.

Therefore, constructed on this behaviour, several combinations of these reflectance properties have conformed useful spectral vegetation indices and used to track changes in plant-related traits, such as leaf growth (and leaf area index) and CO_2 uptake from remote sensing using optical sensors (Helman, 2018). The most used vegetation indices are the NDVI and the EVI (enhanced vegetation index).

Both indices are normalised ranging between -1 and 1 for non-vegetated and fully vegetated surfaces, respectively, which makes them particularly useful for comparative studies. These indices take advantage of the sharp contrast in light absorption across the narrow wavelength range of 620-800 nm (Helman, 2018). The advantage of the EVI over the NDVI is its use of the blue band (500 nm), which minimises residual aerosol variations (Huete, 2012), though strongly affected by atmospheric disturbance. Additionally, the EVI is less dependent on the red and as such it does not saturate as rapidly as NDVI in dense vegetation, so this enables to be used over dense vegetated areas (Helman, 2018).

In this study, apart from using the NDVI and the EVI, other vegetation indices have been calculated to investigate which ones are the most suitable indices to forecast olive yield (see annex **Equations of the vegetation indices applied in this study 11.1**). The list of vegetation indices used, and a brief explanation and their main references is below:

- **NDVI.** The Normalized Difference Vegetation Index measure of healthy, green vegetation. The combination of its normalized difference formulation and use of the highest absorption and reflectance regions of chlorophyll makes it robust over a wide range of conditions. It can, however, saturate in dense vegetation conditions when LAI becomes high (Rouse et al., 1973).
- **EVI.** The Enhanced Vegetation Index is useful in high LAI regions where NDVI may saturate. It uses the blue reflectance region to correct for soil background signals and to reduce atmospheric influences, including aerosol scattering and improving sensitivity in high biomass regions (Huete, 2012).
- **EVI2.** The 2-band Enhanced Vegetation Index was designed to sensor systems without blue band such as the Advanced Very High-Resolution Radiometer (AVHRR) and may reveal different vegetation dynamics in comparison with the AVHRR NDVI dataset (Jiang et al., 2008).
- **NDMI.** The Normalized Difference Moisture Index (NDMI) detects moisture levels in vegetation being a reliable indicator of water stress in crops. Water stress correspond to the negative values, whereas positive ones no water stress (B. Gao, 1996).
- **GNDVI.** The Green Normalized Difference Vegetation is an index of plant photosynthetic activity. It is a chlorophyll index and used at later stages of development and is more sensitive to chlorophyll concentration than NDVI. It is one of the most widely used vegetation indices to determine water as values between -1 and 0 are associated with the presence of water or bare soil (A. Gitelson & Merzlyak, 1994).
- **GCI.** The green chlorophyll index is used to estimate leaf chlorophyll content across a wide range of plant species (A. A. Gitelson et al., 2003).
- **SIPI.** The structure intensive pigment vegetation index maximizes sensitivity to the bulk carotenoids to chlorophyll ratio, while minimizing the impact of the variable canopy structure. It is particularly useful in areas with high variability in the canopy structure, or leaf area index. SIPI values range from 0 to 2, where healthy green vegetation ranges from 0.8 to 1.8 (Penuelas et al., 1995).
- **SAVI.** The Soil Adjusted Vegetation Index (SAVI) is a modification of the NDVI with a correction factor for soil brightness (Huete, 1988).

- **ARVI.** The atmospherically resistant vegetation index is an enhancement to the NDVI that is resistant to atmospheric factors (for example, aerosol). The value of this index ranges from -1 to 1, with higher pixel values corresponding to healthier and greener vegetation (Kaufman & Tanre, 1992).
- **AVI.** The Advanced Vegetation Index reacts more sensitively to vegetation quantity and can highlight subtle differences in canopy density (Jacobsen et al., 1995).
- **OSAVI.** As SAVI, the optimized soil adjusted vegetation index is the soil-adjusted vegetation index is used for agricultural monitoring when soil brightness plays a key role such as in areas where vegetative cover is low, and the soil is exposed (Rondeaux et al., 1996).
- **ReCI.** The red-edge chlorophyll vegetation indices responsive to chlorophyll content in leaves that is nourished by nitrogen. ReCI shows the photosynthetic activity of the canopy cover but is not suitable for the season of harvesting (A. A. Gitelson et al., 2006).

All these twelve vegetation indices have been calculated directly from the surface reflectance data (blue, green, red, infrared, and shortwave infrared) from Landsat-7, Landsat-8, MODIS and TOA reflectance for Sentinel-2 (Units, wavelength ranges and resolution in **table 2** and **Equations of the vegetation indices applied in this study** in point **11.1**).

For instance, the NDMI (Normalized Difference Moisture Index) have been used to examine the effects of water stress on the crop yield as the study area normally suffers from very dry conditions during summer months and severe drought conditions time to time. The NDMI is calculated using the near-infrared (NIR) and the short-wave infrared (SWIR) reflectance:

NDMI = (NIR – SWIR) / (NIR + SWIR)

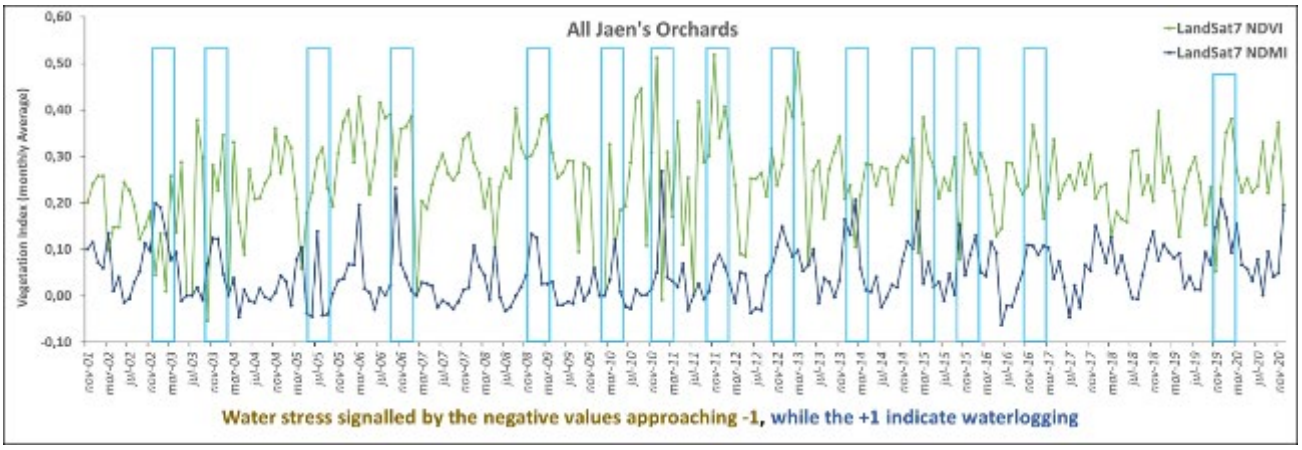


Figure 7 Monthly average of NDVI and NDMI index calculated from Landsat reflectance data. See the slightly shifted (NDMI, before and NDVI slightly after, see blue rectangles) opposite correlation of both indices during November-March. The NDMI valleys (lower values) are recorded starts with the blooming season and during the summer months opposite to the fruit formation phase of the olive trees.

The short-wave infrared spectral channel (SWIR) is sensitive to the vegetation water content and the mesophyll structure of leaves. On the other hand, the near-infrared band (NIR) picks up the bright reflectance off the leaf internal structure and leaf dry matter content. When combined, the accuracy of data on the vegetation water content becomes much higher. NDMI values vary throughout the growing season because the plants' reflectance is slightly different for every phenological stage. There also exists an interesting correlation between NDMI and NDVI (**Figure 7**). During November and March there is an opposite correlation, which is slightly shifted as the NDMI peak is recorded slightly before the peak of the NDVI.

Currently, many vegetation indices developed, but also many other spectral indices have been used in studies of water bodies, soils, snow cover, surfaces of artificial materials. Not all spectral indices can be applied to all sensors since the basic requirement is that the sensor covers the bands used in the index when selecting data to calculate indices.

3 Influence of the Mediterranean climate on the phenology of olive trees

When using remote sensors to capture phenological stages, these limitations are found in the setting of the studied olive orchards in Jaen. There are three characteristics to take into consideration:

1. Olive trees are evergreen.
2. The study area is under the Mediterranean climate.
3. Olive tree canopies are not big enough, so the understorey layer covers a significant area of the orchard floor.

The scenario described by Helman (2018) for Mediterranean climates is followed by the olive orchards in the Jaen province. Winter months, from November to February register much higher vegetation index values than summer ones, where these vegetation indices are normally at the lowest levels of the time series. This is due to the high green coverage of ephemeral herbaceous plants on the woodland floor at the end of the rainy season.

Moreover, as the olive tree canopies are not big enough, the contribution to the greenness proxy signal of the understorey herbaceous layer is highly significant because it covers almost the entire area of the woodland floor, and greenness proxies within a single pixel reach the highest values (Helman, 2018). This is a constant throughout the time series from November 2001 to December 2020.

This fact must be bear in mind as the winter months' vegetation index can lead to errors when trying to explain the phenological activity of olive trees and mislead the regression analysis. In this case, the understorey herbaceous layer can overestimate greenness proxies (e.g., NDVI or EVI) during the rainy season.

Therefore, it is vital to bear in mind that spatial resolution, apart from a horizontal limitation, also represents a vertical limitation as the bigger is the pixel, the wider is the canopy and ground area to be captured by the pixel. In **figure 8** is shows the monthly averages of the NDVI, EVI, and NDMI vegetation indices from all the olive orchards in the Jaen province.

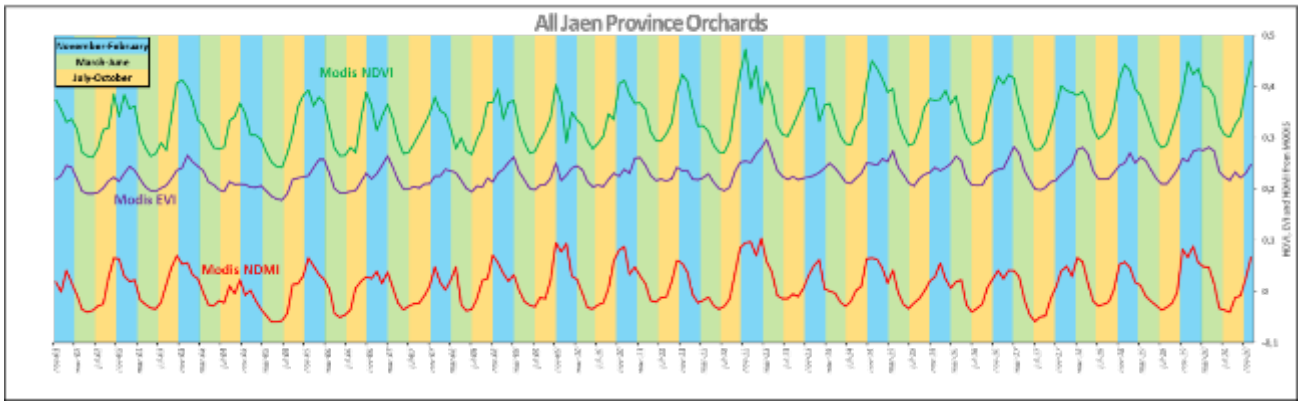


Figure 8 Monthly average of the NDVI, EVI and NDMI vegetation indices for all Jaen's orchards. The coloured bands divide the years into thirds (see legend).

This point is explained because the highest rainfall rates are usually between the end of October and February. In **figure 9**, the rainfall rate plotted with the main vegetation indices shows the highest rates in winter months. This creates a higher greenness proxy because the understorey layer makes an important contribution to the pixel value, more than the overstorey development (**Figure 9**).

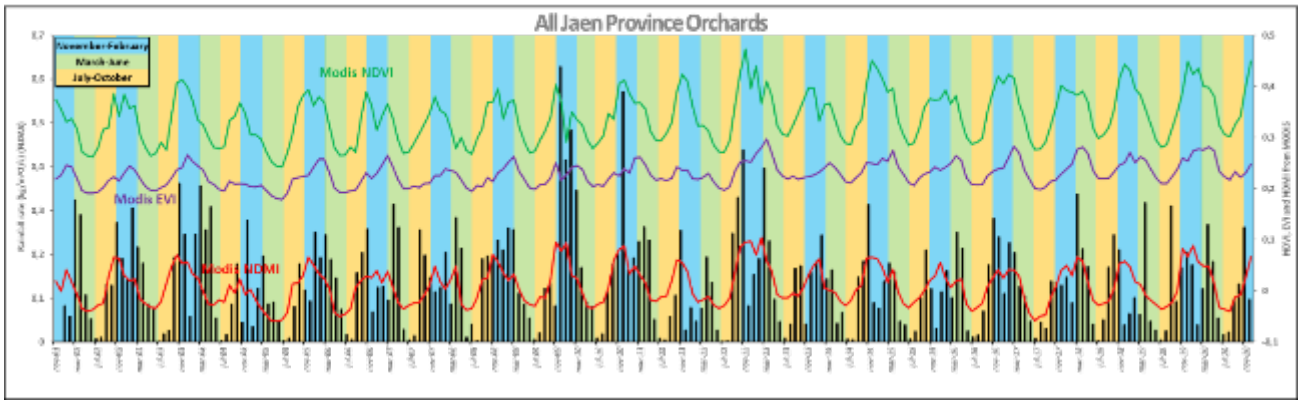


Figure 9 Rainfall rate from FLDAS dataset (black bars) plotted with the monthly average of the NDVI, EVI and NDMI vegetation indices for all Jaen's orchards. The coloured bands divide the years into thirds (see legend).

This scenario is confirmed because olive tree canopies are not big enough, so the understorey layer covers a significant area of the orchard floor compared to the canopy of the olive trees. To evaluate the extent of the understorey layer, the Meta's Segment Anything Model (SAM) for segmenting objects in images was used in the orchard 4134152 (ESRI, 2023). The SAM model calculated that the canopy area of olive trees only covers 27% of the whole orchard. In other words, 61,201.96 square meters are covered by the canopy, whereas more than 225,000 square meters represent bare soil (**Figure 10**).

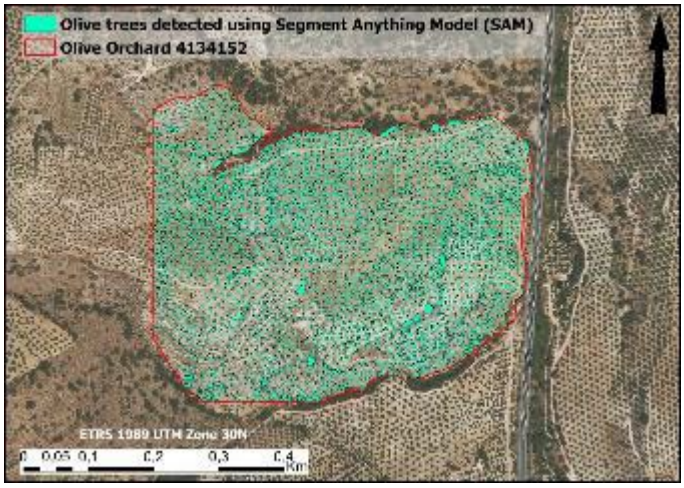


Figure 10 Meta's Segment Anything Model (SAM) applied to identify olive trees in the orchard 4134152.

Therefore, this scenario of multi-canopy layers is reflected in the satellite-derived signal and confirms how much the understorey layer influences the pixel value, especially in winter months. During winter the understorey layer is covered by small emerging bushes, scrubs, and grass becoming the most dominant plant species in this biome (**Figure 8** and **Figure 9**).

Helman (2018) mentioned that the herbaceous layer is marginal in terms of biomass, but its contribution to the vegetation index signal is the most significant because it covers almost the entire woodland floor. **Figure 11** shows the pixel size of the four remote sensors used in this study to represent how much of the understorey layer contributes to the surface reflectance registered by these sensors. Consequently, it is vital to bear in mind this multistorey scenario because the spatial resolution, apart from a horizontal limitation, also represents a vertical limitation as the bigger is the pixel, the wider is the canopy and ground area to be captured by the pixel (Helman, 2018).

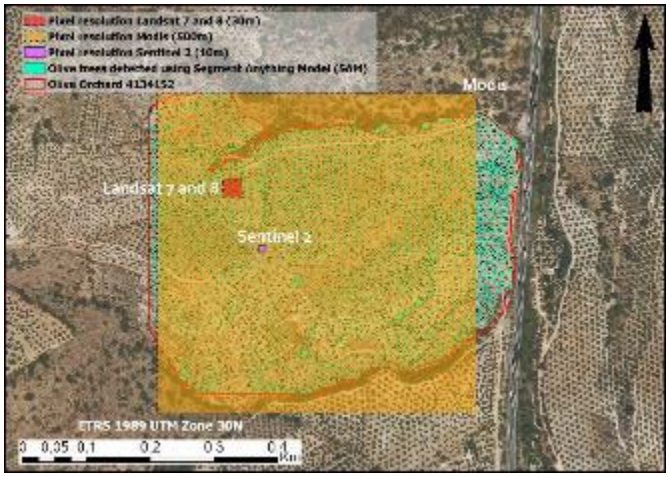


Figure 11 Representation of the pixel resolutions of MODIS (500m), Landsat 7 and 8 (30m) and Sentinel 2 (10m) overlaid with the canopy of olive trees in the orchard 4134152.

On the other hand, given that 45% of the Jaen province is used for olive crops, pixel edges have been treated following the GEE algorithm “ee. Reducer.mean()”. This GEE method incorporates all pixels if at least 0.5% of the

pixel is within the polygon, so not too strict in terms of special spatial considerations and calculations (further information in point **5.2**). This was decided because the surroundings of the study areas are more olive orchards (**Figure 11** and **Figure 12**), so that should not give any other signal rather than olive orchards. Hence, the Mediterranean climate of the Jaen province, coupled with the evergreen nature of the olive trees and their restricted canopy extent, forces us to choose vegetation indices that consider these characteristics. Also, it is important to be cautious and bear in mind this scenario when making comparisons among the different vegetation indices and especially when discussing and concluding.

Figure 12 Aerial view of Jaen province showing the so-called sea of olives trees covering the 45% of the Jaen province.

3.1 Olive tree and its phenological processes

The olive tree (*Olea europaea*) is a woody, medium-sized perennial, evergreen tree (Zohary & Spiegel-Roy, 1975). Mediterranean climate is the natural olive tree habitat, characterized by mild winters and warm and dry summers (Besnard et al., 2009; Sanz-Cortes et al., 2002). The average time for olive trees to begin producing olives is around 4-5 years and can live and produce olives for over a century. The olive tree's phenological processes begin with the buds sprouting until the fruit ripens and the olive tree's dormancy until the following winter. The phenological stages of the olive tree in the study region are:

- March/April: Sprouting to bud development.
- May: Olive tree blooming.
- May/June: Fruit formation or Curdling.
- June/November: Fruit development.
- November/December: Veraison and harvesting.
- January/February: Dormancy.

3.1.1 Sprouting to bud development (March/April)

At the end of winter, the sprouting process begins with the appearance of the buds. In this phase, these shoots can be of two types: vegetative or flower shoots. The vegetative shoots will become new stems with leaves and new buds. The flower shoots are responsible for the reproduction of the olive tree, producing flowers from which the fruits will later sprout. Olive buds are susceptible to low temperatures but also to extreme hot weather as well. In spring they grow the most, then stop growing in summer and will grow again in autumn. For this reason, this stage is crucial as March and April are one of the key months for monitoring purposes with remote sensors.

Figure 13 Flowering of olive trees. No changes were made in this image.
Author: Luis Fernández García, Creative Commons Attribution-Share Alike 4.0, (https://creativecommons.org/licenses/by-sa/4.0/)

3.1.2 Olive tree blooming (May)

The onset of the flowering period depends on altitude (Aguilera & Valenzuela, 2009). These authors demonstrated by statistical analysis that temperature, humidity, cumulative rainfall, and cumulative solar radiation are the meteorological parameters that most affect olive floral phenology.

Also, these authors showed that the pollen and blooming in Jaen province can be seen between May and June, but May is considered the blooming month (**Figure 13** and **Figure 14**). This blooming stage will not last more than a week. **Figure 14** highlights the province of interest for this study (Jaen) with the dashed green line, where 90% of the blooming stage occurs in May.

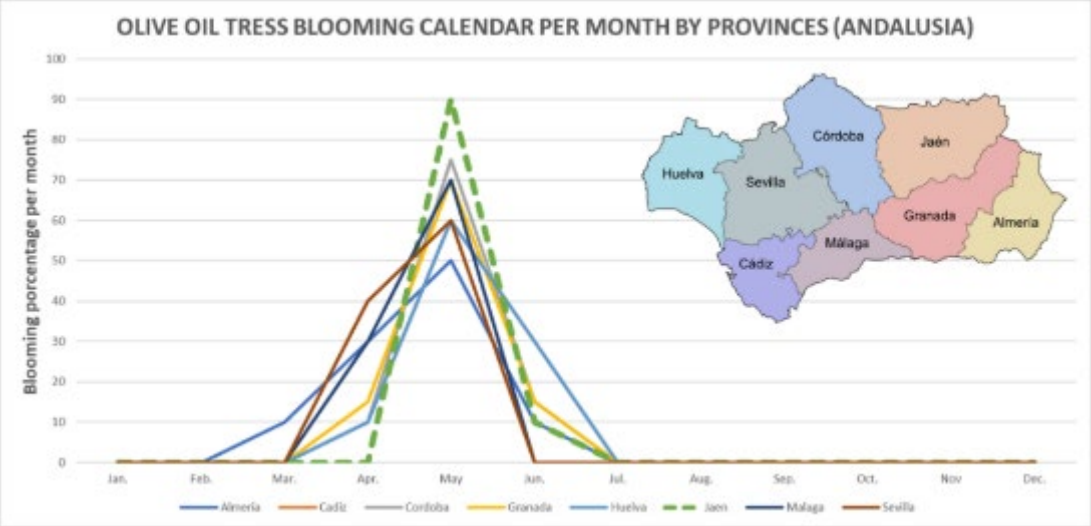


Figure 14 Blooming Calendar for Olive oil tress (2014, 2015 and 2016) in Andalusia (Data extracted from Piferrer et al., 2017)

This stage is one of the most important phases as buds of olive trees bloom, giving rise to a substantial number of flowers whose objective is to be fertilized, thus becoming new fruits. At the end, very few flowers will be fertilized and become fruits. In an estimated calculation, it will be about one or two per cent.

3.1.3 Fruit formation or Curdling (May/June)

Once the flowering stage is over, the petals come off their flowers once fertilized, and the new olives are about to grow. This phase, called fruit formation, is followed by a natural fall of some new fruits (**Figure 15**) as normally more fruits are set than can be supported by sufficient nutrient supply. Hence, the olive tree makes its own selection and eliminates the fruits that it will not be able to feed, thus ensuring its survival during the hot and dry summer.

Figure 15 Fruit formation. No changes were made in this image. Author: Ewen Cameron, Creative Commons Attribution-Share Alike 4.0, (https://creativecommons.org/licenses/by-sa/4.0/)

3.1.4 Fruit development (June/November)

During this stage, the oil formation takes place inside the fruit and is when the olive fruits grow. The fruit bone begins to harden. This process is challenging for olive trees since they must face a lack of water and elevated temperatures, making this period complex and discontinuous.

Summer months are also good for tracking purposes with remote sensors as water stress could be well identified with specific vegetation indices and water-dependent meteorological datasets. In the Jaen province, summer is usually dry and hot, so the olive tree stops its activity for most of the day, being active in the early and late hours of the day when temperatures drop. To save water, the olive closes its stomata, thus avoiding the evaporation that occurs in the leaves.

3.1.5 *Veraison and harvesting (November/December)*

By the end of the summer, fruits begin to change colour, and each variety of olive ripens at various times. In general, olives lose the green tone, changing to a yellowish or pink tone, until reaching an intense garnet or jet-black tone, when it has reached its maximum degree of maturity.

During the final phase of maturation, the flesh of the fruit accumulates water and sugars, and by the end of autumn, the veraison phase is completed, and the olive has accumulated all its oil. At the end of this stage is when harvest is done (**Figure 16**). It is important to highlight that all olive fruit production of the province of interest for this study (Jaen) is for olive oil end-product (**Figure 17**).

Figure 16 Maturation stage of the olive fruits. No changes were made in this image.
Author: Filipe Cera, Creative Commons Attribution-Share Alike 4.0, (https://creativecommons.org/licenses/by-sa/4.0/)

The harvesting calendar for the years 2014, 2015 and 2016 by the Spanish Ministry of Agriculture shows subtle variations about the olive crops in the different provinces (Piferrer et al., 2017).

The harvest of olive oil trees starts earlier in western provinces (Huelva, Sevilla, and Cadiz) and picked later in eastern provinces with higher altitudes (**Figure 17**). In Jaen the harvest is done in November highlighted with the dash green line.

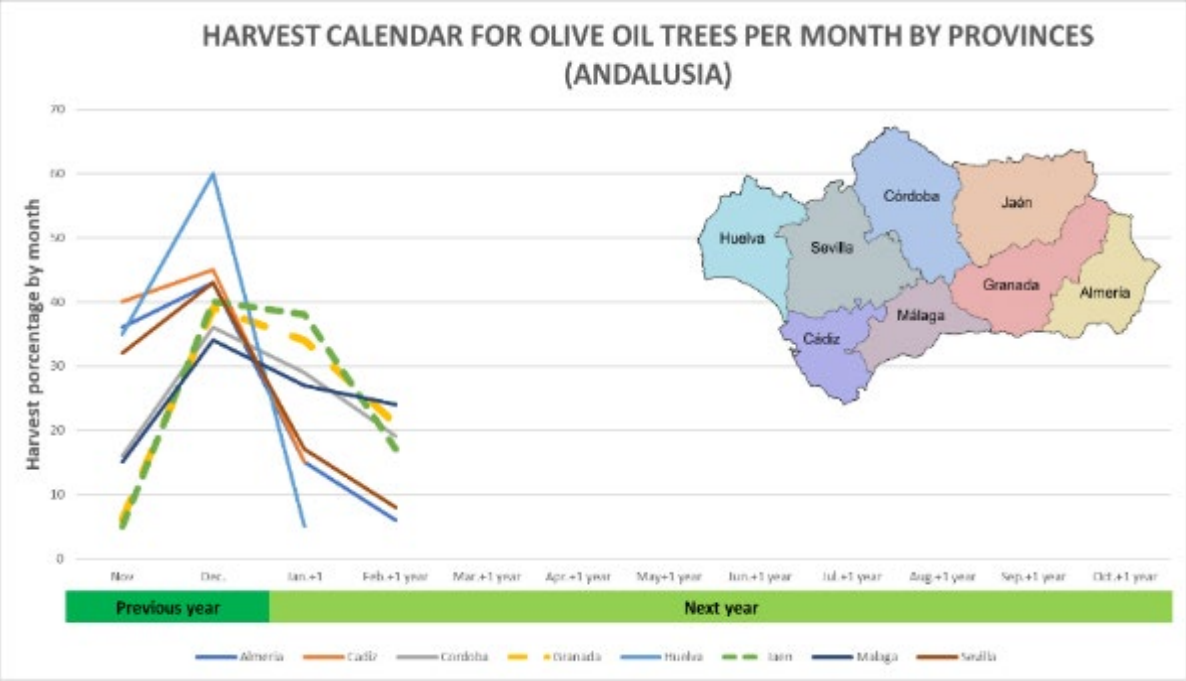


Figure 17 The Harvest Calendar for olive oil production (2014, 2015 and 2016) in Andalusia (Piferrer et al., 2017)

3.1.6 Dormancy (January/February)

The end of the harvest session and the beginning of winter are also the beginning of dormancy until the next cycle. These months are used for pruning and ploughing crops. Another way to describe the phenological processes of olive trees is the BBCH scale.

The BBCH scale has been used for a wide range of crops where similar growth stages for each plant are given the same numerical code. The BBCH describe the phenological growth stages of olive trees using the BBCH scale with eight principal growth stages and thirty-two secondary growth stages (Sanz-Cortes et al., 2002). These stages are summarised in the annex **The BBCH-scale 11.2**. Winter months are the least suitable months for monitoring olive yield. The rainfall season occurs during winter months and the understorey layer is covered with herbs, shrubs. Therefore, ephemeral herbaceous plants cover the understorey layer at the end of the rainy season contributing to increase vegetation indices since olive tree canopies are not big enough and the understorey layer covers a significant area of the acquired pixel (Helman, 2018).

4 Data sources

4.1 List of Datasets

As described previously, this research project aims to evaluate how accurately remote sensing and weather data are combined to forecast olive crop yield on several polygons known for olive yield. Consequently, the datasets are primarily derived from remote sensors. The list of datasets needed to conduct the aims of this study is as follows (**Table 2**):

- Reflectance data from MODIS, Sentinel-2, Landsat-7 and Landsat-8
- Climate data from the Regional Government of Andalusia
  - Precipitation data
  - Temperature data
  - Scouting data such as orthophotos, land use, and DEM used to extract scouting data such as altitude, slope, and aspect.
- SPEI index
  - SPEI 1-months' time scale
  - SPEI 3-months' time scale
  - SPEI 6-months' time scale
  - SPEI 12-months' time scale
- Olive yield data from the Jaen province collected by the Ministry of Agriculture
- From the TerraClimate dataset several bands have been extracted:
  - Actual evapotranspiration derived using a one-dimensional soil water balance model.
  - Climate water deficit derived using a one-dimensional soil water balance model.
  - Vapour pressure deficit.
  - Soil moisture derived using a one-dimensional soil water balance model.
  - Precipitation accumulation
  - Minimum and Maximum temperature
- From MODIS collection MOD15A2H V6.1 have been extracted
  - Leaf Area Index (LAI)
  - Fraction of Photosynthetically Active Radiation (FPAR)
- MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m
  - NDVI vegetation index
  - EVI vegetation index
- FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System
  - Soil moisture in the interval between 0-10 cm
  - Soil moisture in the interval between 10-40 cm
  - Soil moisture in the interval between 40-100 cm
  - Rainfall rate accumulation
  - Specific humidity
- ERA5-Land Monthly Aggregated - ECMWF Climate Reanalysis
  - Temperature of air at 2m above the surface of land
  - Temperature of the soil in layer 2 (7-28 cm)
  - Temperature of the soil in layer 3 (28-100 cm)
  - Temperature of the soil in layer 4 (100-289 cm)

- Net thermal radiation at the surface
- Amount of evaporation from bare soil at the top of the land surface
- Amount of evaporation from the canopy interception reservoir at the top of the canopy
- Amount of evaporation from vegetation transpiration

The **table 2** summarises the main parameters of the datasets used in this study. The time span of these data sets is from November 2001 to November 2020. The links of these datasets can be found in the annex **11.5**.

DATASETS from SPANISH AND REGIONAL GOVERMENTS				
Dataset Name	**Resolution**		**Units/Type/Res**	**Source**
Temperature	500 meters		Celsius	GIS Department of the Regional Government of Andalusia
Precipitation	500 meters		mm	
Orthophotos	0.25 meters		0,5 m resolution	
DEM	5 meters		5 m resolution	
Land Use	shapefile		shapefile data	
Olive yield Information	shapefile		Kgr/year	Ministry of Agriculture of Spain
GOOGLE EARTH ENGINE DATASETS				
Dataset Name	**Resolution**	**Bands used for this study**	**Units/Band width**	**Source**
ERA5-Land Monthly Aggregated - ECMWF Climate Reanalysis	27830 meters	"temperature_2m" (Temperature at 2 meters above the surface) "soil_temperature_level_2" (Temperature of the soil 7-28 cm) "soil_temperature_level_3" (Temperature of the soil 28-100 cm) "soil_temperature_level_4" (Temperature of the soil 100-289 cm) "surface_net_thermal_radiation_sum" (Net thermal radiation at the surface) "evaporation_from_bare_soil_sum" (evaporation from bare soil at the top of the land surface) "evaporation_from_the_top_of_canopy_sum" (evaporation from the canopy interception reservoir at the top of the canopy) "evaporation_from_vegetation_transpiration_sum" (evaporation from vegetation transpiration)	K K K K J/m^2 m of water equivalent m of water equivalent m of water equivalent	European Union/ESA/Copernicus
FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System	11132 meters	"SoilMoi00_10cm_tavg" (soil moisture 0-10 cm) "SoilMoi10_40cm_tavg" (soil moisture 10-40 cm) "SoilMoi40_100cm_tavg" (soil moisture 40-100 cm) "Rainf_f_tavg" (rainfall average) "Qair_f_tavg" (humidity)	Volume fraction Volume fraction Volume fraction kg/m^2/s Mass fraction	Goddard Earth Sciences Data and Information Services Center
MOD15A2H.061: Terra Leaf Area Index/FPAR 8-Day Global 500m	500 meters	"Lai_500m" (Leaf Area Index) "Fpar_500m" (Fraction of Photosynthetically Active Radiation)	m^2/m^{-2} %	NASA LP DAAC at the USGS EROS Center
MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m	250 meters	NDVI EVI	Min:-0.2 and Max: 1 Min:-0.2 and Max: 1	NASA LP DAAC at the USGS EROS Center
SPEIbase: Standardised Precipitation-Evapotranspiration Index database, Version 2.8	55660 meters	"SPEI_01_month" "SPEI_03_month" "SPEI_06_month" "SPEI_12_month"	Min: -2.33 - Max: 2.33	Spanish National Research Council (CSIC)
TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho	4638.3 meters	"tmmn" (Minimum temperature) "tmmx" (Maximum temperature) "vpd" (Vapour pressure deficit) "def" (Climate water deficit) "soil" (Soil moisture) "aet" (Actual evapotranspiration) "pr" (Precipitation accumulation)	°C °C kPa mm mm mm mm	University of California Merced
USGS Landsat 7 Level 2, Collection 2, Tier 1	30 meters	Surface reflectance "SR_B1" (Blue) from 28/05/1999 Surface reflectance "SR_B2" (Green) from 28/05/1999 Surface reflectance "SR_B3" (Red) from 28/05/1999 Surface reflectance "SR_B4" (Infrared) from 28/05/1999 Surface reflectance "SR_B5" (Shortwave Infrared) from 28/05/1999	0.45 - 0.52 µm 0.52 - 0.60 µm 0.63 - 0.69 µm 0.77 - 0.90 µm 1.55 - 1.75 µm	U.S. Geological Survey
USGS Landsat 8 Level 2, Collection 2, Tier 1	30 meters	Surface reflectance "SR_B2" (Blue) from 18/03/2013 Surface reflectance "SR_B3" (Green) from 18/03/2013 Surface reflectance "SR_B4" (Red) from 18/03/2013 Surface reflectance "SR_B5" (Infrared) from 18/03/2013 Surface reflectance "SR_B6" (Shortwave Infrared) from 18/03/2013	0.452 - 0.512 µm 0.533 - 0.590 µm 0.636 - 0.673 µm 0.851 - 0.879 µm 1.566 - 1.651 µm	U.S. Geological Survey
MOD09A1.061 Terra Surface Reflectance 8-Day Global 500m	500 meters	Surface reflectance "sur_refl_b01" (Red) from 18/02/2000 Surface reflectance "sur_refl_b02" (Near infrared) from 18/02/2000 Surface reflectance "sur_refl_b03" (Blue) from 18/02/2000 Surface reflectance "sur_refl_b04" (Green) from 18/02/2000 rface reflectance "sur_refl_b06" (Shortwave Infrared) from 18/02/20	620-670nm 841-876nm 459-479nm 545-565nm 1628-1652nm	NASA LP DAAC at the USGS EROS Center
Sentinel-2 MSI: MultiSpectral Instrument, Level-1C	10 meters	TOA reflectance "B2" (Blue) from 23/06/2015 TOA reflectance "B3" (Green) from 23/06/2015 TOA reflectance "B4" (Red) from 23/06/2015 TOA reflectance "B8" (Near infrared) from 23/06/2015 TOA reflectance "B11" (SWIR) from 23/06/2015	496.6nm-492.1nm 560nm-559nm 664.5nm-665nm 835.1nm-833nm 1613.7nm-1610.4nm	European Union/ESA/Copernicus

Table 2 Summary of datasets used in this study. In the GEE datasets, the name of the bands shown in this table in quotation marks are the same that appear in the datasets. More descriptions of these datasets are in the following text.

4.2 Climate and scouting data from the Regional Government of Andalusia

The main source of information has been the Regional Government of Andalusia. This regional government offers a solid website portal, the REDIAM catalogue where users can download these datasets with their metadata. The spatial reference of both datasets is: ETRS 1989 UTM ZONE 30N (EPSG:25830). As almost all datasets have been obtained with this projected coordinate system, the ArcGIS Pro project, their calculations, and data extractions from the ArcGIS Pro project have been operated under this PCS ETRS 1989 UTM ZONE 30N. The access and use constraints are licenced under the Creative Commons Attribution 4.0 International Public License (CC BY 4.0). From the REDIAM catalogue of the Regional Government of Andalusia the datasets obtained are as follows:

- Precipitation data
- Temperature data
- Scouting data such as orthophotos, and DEM where altitude was extracted.

4.2.1 Temperature dataset

The temperature dataset is in raster format of the average monthly temperature, which reflects the monthly evolution of temperatures in Andalusia since 1951 (**Figure 18**). No information about the spatial interpolation techniques used to build up these rasters is provided in the metadata (see annex **11.5.1**).

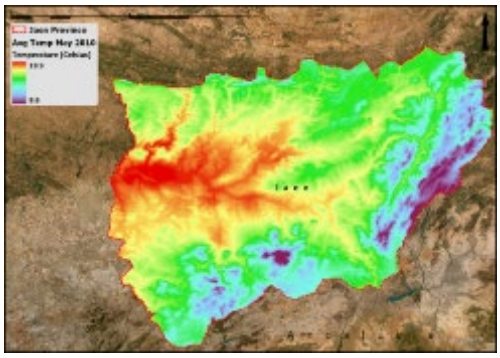


Figure 18 Example of the average monthly temperature dataset (May 2010).

4.2.2 Precipitation dataset

The precipitation dataset is made of layers of geographic information in raster format that represent the average monthly rainfall in the reference period 1991-2020 (**Figure 19**). This information has been prepared from data from the weather stations of the State Meteorological Agency and the Ministry of the Environment (see annex **11.5.2**).

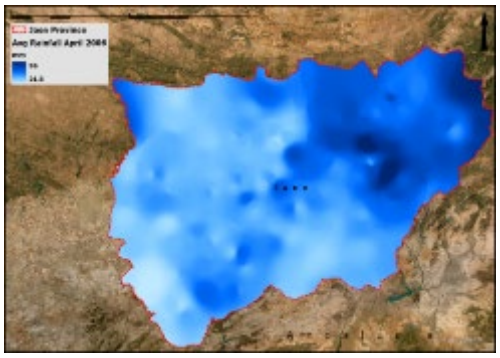


Figure 19 Example of the average monthly precipitation dataset (April 2006).

4.2.3 Regional of Orthophotos of the study area

The regional orthophotos of Andalusia were generated within framework of the National Plan for Aerial Orthophotography (NPAO) 2020-2021. It is based on a photogrammetric-LIDAR flight of 2019 with a flight pixel size of 35 cm.

There are available in three channels (RGB) and in ETRS89 h30 and h29 with a resolution of 0.25 m. In **figure 20**, it can be seen the mosaic of orthophotos covering the whole of the Jaen province (see annex **11.5.3**).

This dataset was used to calculate the extend of the understorey layer in comparison to the canopy of the olive trees. As described in chapter **3**, the Meta's Segment Anything Model (ESRI, 2023) for segmenting objects in images was used for this purpose (**Figure 10**).

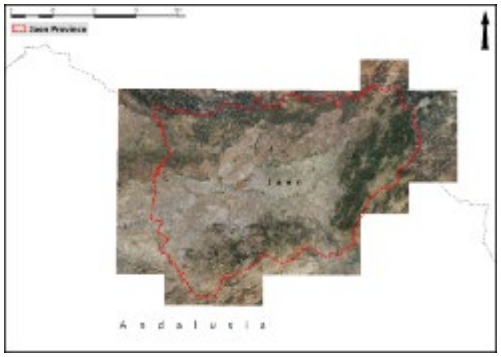


Figure 20 NPAO dataset of orthophotos mosaic covering the Jaen province.

4.2.4 Digital Elevation Model

The digital elevation model was used to extract heights of the orchards used for the analysis to see the effects of altitude in the oil yield production and forecasting. This DEM delivered the government of Andalusia was obtained by automatic stereo-correlation from the 2013 photogrammetric flight of National Plan for Aerial Orthophotography (NPAO) with a resolution of 5m (see annex **11.5.4**).

4.2.5 Land Use dataset

The land use map is based on the SIGPAC Spanish methodology. The SIGPAC is the Geographic Information System for the Identification of Agricultural Plots, created through collaboration between the Spanish Agrarian Guarantee Fund (FEGA) and the different Autonomous Communities, within their territories, as an element of the Integrated Management System and Control of direct aid schemes (see annex **11.5.5**).

SIGPAC Land Use	Description	SIGPAC Land Use	Description
IV	Greenhouse crops	VF	Mix vineyard and Fruit tress
TA	Arable land	VI	Vineyards
TH	small orchards	VO	Mix of vineyards and Olive trees
CF	Mix citrus-fruit trees	PA	Grass with trees
CI	Citrus tress	PR	Shrub grass
CS	Mix of citrus-fruit peel trees	PS	Pasture
CV	Mix Citrus-Vineyard	ZC	Concentrated area not included in the orthophoto
FL	Mix of nuts and olive groves	AG	Streams and water surfaces
FS	Nuts	CA	Roads
FV	Mix of nuts and vineyards	ED	Buildings
FY	Fruit tress	FO	Forests
OC	Mix of olive and citrus trees	IM	Non-productive land
OF	Mix of olive and fruit trees	ZU	Urban areas
OV	Olive trees	EP	Landscape element

Table 3 SIGPAC Land Use classification

4.2.6 Olive yield Information in the study area

The production data was sourced from the ESYRCE, acronym that is Spanish is: "*Encuesta sobre Superficies y Rendimientos Cultivos*" or Survey on Areas and Crop Yields (ESYRCE).

This survey about areas and crop yields has been managed annually since 1990 in collaboration with the statistical services of the Autonomous Communities. It is based on field research, in which information is taken directly at the foot of the plot in a georeferenced sample of the national territory, conducted in the months of May to August. The results obtained constitute an objective data source that complements other statistical information from the Ministry to obtain official data, which are later published in the statistical yearbook of the Ministry of Agriculture.

Figure 21 Example of a segment unit of the ESYRCE methodology. Public domain image taken from the report "Análisis de las plantaciones de olivar en Andalucía. Año 2015, 2015"

The methodology is based on the cells into which the Spanish national territory is divided. A cell is understood to be the portion of the territory constituted by each square of 100 ha (blue square in **figure 21**), delimited by the basic grid lines (red square in **figure 22**) of the UTM projection of the National Topographic Map in the official projection system (European Datum 1950).

These cells are divided by observation units which known as segments (**Figure 22**). Segments are cell fractions formed by the square of 49 ha supported by the South-West (S.W.) angle of the corresponding cell.

Figure 22 Sampling methodology of the ESYRCE system. Public domain image taken from the report "Análisis de las plantaciones de olivar en Andalucía. Año 2015, 2015"

In the report ESYRCE (2019) is described the ESYRCE methodology for the olive crops (only in Spanish). The data format is delivered in shapefile.

For the olive crops in Jaen, yield data is available from 2002 until 2020. The segments selected by the Survey on Areas and Crop Yields (ESYRCE) to yield data from olive orchards are shown in **figure 23**. The yield data is accompanied by scouting data such as age of the olive trees, information about the irrigation (only rainfall or artificial irrigation), or density of the trees.

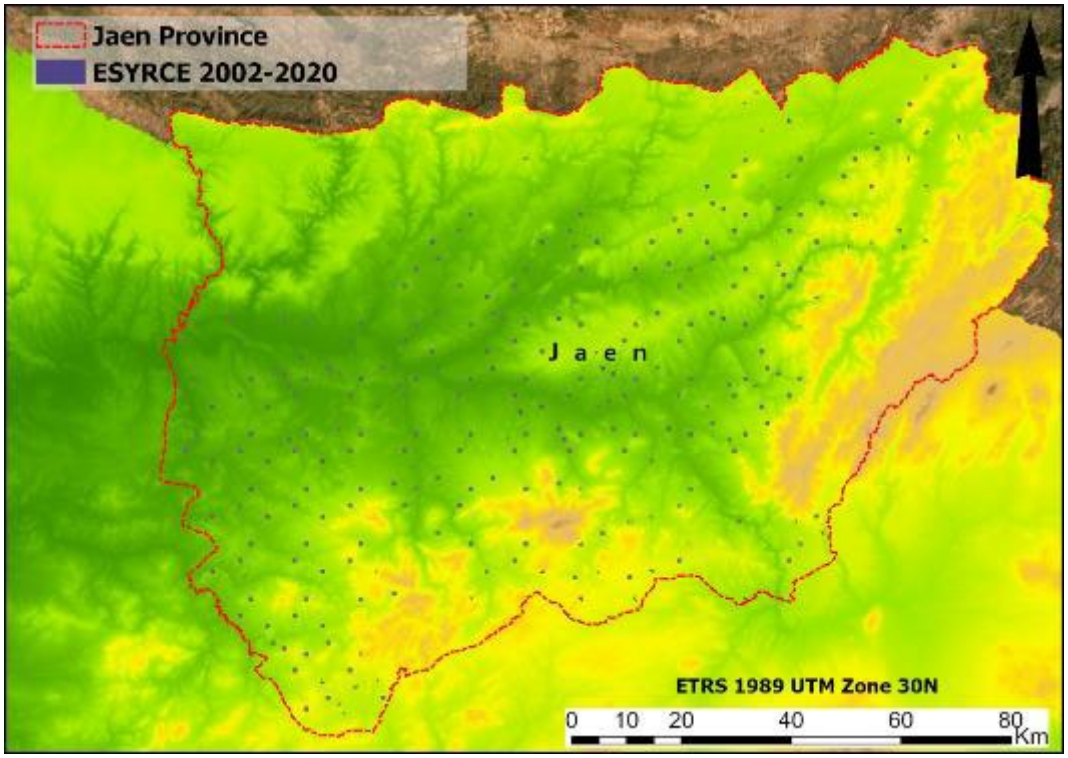


Figure 23 Segments given by the Survey on Areas and Crop Yields (ESYRCE) with yield data from olive orchards.

The first approach to this dataset showed that the selected orchards in this study are small and very fragmented. Also, some polygons or orchards change their area from year to other, so is difficult to find orchards with a constant area over the years and even more difficult to find orchards with a long yield record of years.

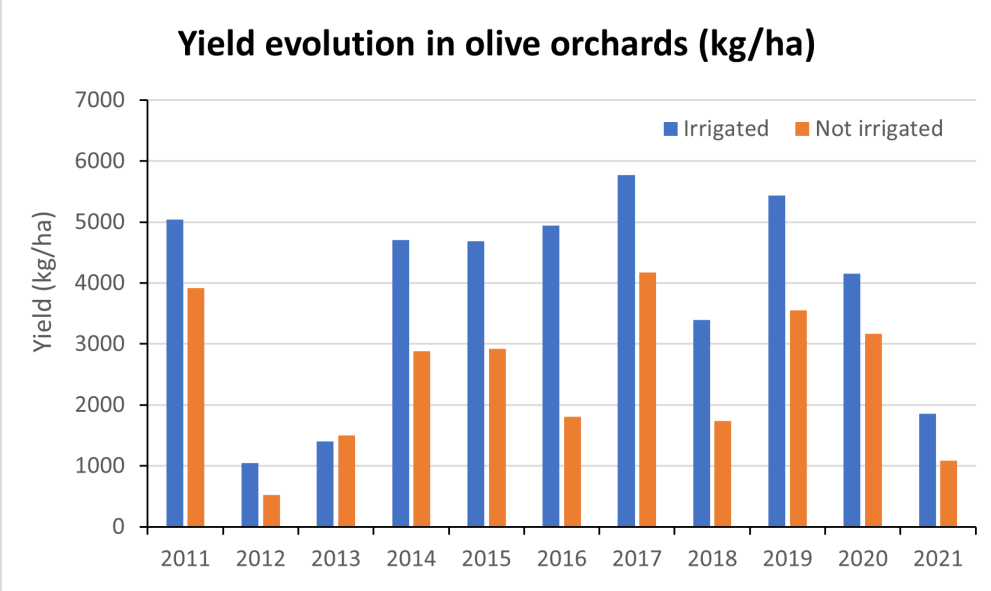


Figure 24 Yield of irrigated olive orchards in comparison with non-irrigated orchards between 2010-2020. Data from the Survey on Areas and Crop Yields (ESYRCE).

Obviously, the information about irrigation systems is particularly important, so this condition was also considered. According to the Survey on Areas and Crop Yields (ESYRCE), irrigated olive orchards show yield a 35% higher than non-irrigated olive groves (**Figure 24**).

Therefore, polygons with irrigation systems have been discarded to not distort the regression analysis with precipitation data, so only polygons with natural irrigation have been considered. For this study, apart from the irrigation criteria, other two criteria have been used when screening the Survey on Areas and Crop Yields (ESYRCE): orchards with the longest number of years with production records and trees that are more than 5 years old as this is the age when olive trees begin to produce olives (see **Olive tree and its phenological processes 3.1** and Besnard et al. (2009). At the end, this screening process has given three orchards or segments. In **table 4** and **figure 25** the three segments or orchards selected are shown with their area in square meters, and age of the olive trees in these orchards in which the multilinear regression analysis will be applied.

Year/Yield	Seg: 5034172 - 68 years Area:5151.66 sqm (kg/ha)	Seg: 4134152 - 84 years Area:226753.70 sqm (kg/ha)	Seg: 4434212 - 32 years Area:53334.77 sqm (kg/ha)	Olive Yield in Jaen Province (Tons)
2002		4.500		1.626.898
2003	1.500	6.000		2.908.130
2004	1.000	2.600		1.887.716
2005	100	1.300		1.064.730
2006	500	3.400		2.212.656
2007	1.000	4.000		2.231.270
2008	800	4.200		1.967.425
2009	2.000	6.000		2.526.530
2010				2.736.540
2011	870		6.800	3.024.010
2012	800		1.200	696.248
2013				3.600.070
2014	200	1.250	200	1.023.103
2015	250	4.375	5.600	2.400.287
2016	2.500	4.000	10.500	2.402.787
2017	250	3.000	6.500	1.766.935
2018	800	5.000	6.500	3.206.580
2019	800	2.500	1.850	1.765.585
2020	600	5.000	3.220	2.779.928
2021				

Table 4 The three orchards selected under the three criteria described from the Survey on Areas and Crop Yields (ESYRCE), the production for all olive crops in the Jaen province and the orchard from Cubillas et al (2022) used as the validation model which extend until 2021 as olive yield data is extracted from this paper.

The olive yield of the first three orchards selected from the ESYRCE dataset is shown in kilograms per hectare. Secondly, yield is in tons for the whole orchards of Jaen province. The statistical time is limited by the Survey on Areas and Crop Yields (ESYRCE), ranging from 2002 to 2020. For metadata information and links to download the olive yield dataset, see annex **11.5.6**.

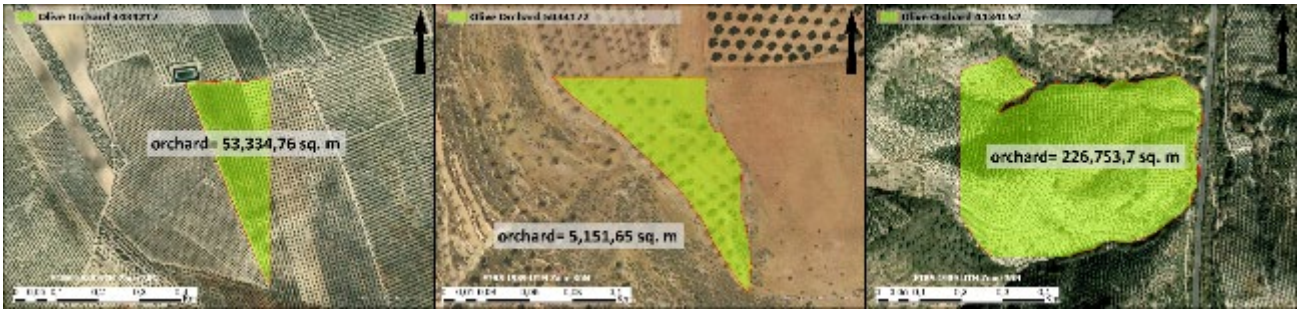


Figure 25 Locations and size of the three olive orchards.

4.3 Datasets extracted from Google Earth Engine (GEE)

4.3.1 The Standardised Precipitation-Evapotranspiration Index

The Standardised Precipitation-Evapotranspiration Index (SPEI) is a global index based on climatic data to track drought conditions. This SPEI index is used to identify when drought events start, their length and scale in comparison to normal conditions. The spatial resolution of the SPEI global dataset is 0.5° (Beguería et al., 2023). This index provides timescales between 1 to 48 months to identify and calculate the scale of drought events all over the world. In this study the intervals used are 1, 3, 6 and 12 months (**Table 2**). For metadata information and links to download the dataset, see annex **11.5.7**.

4.3.2 TerraClimate

The TerraClimate is a climatological database built by merging the high spatial resolution WorldClim dataset with the Climate Research Unit (CRU) of the University of East Anglia and the Japanese 55-year Reanalysis dataset (JRA55).

This dataset is represented by a global coverage and ~4-km of spatial resolution and provide 14 monthly sub datasets such as rainfall, maximum and minimum temperatures, actual evapotranspiration, Palmer drought index, soil moisture, vapor pressure, wind speed among others (Abatzoglou et al., 2018). From those 14 datasets, the variables were rescaled and extracted for the regression analysis are (**Table 2**):

- Precipitation accumulation.
- Minimum and Maximum temperature.
- Actual evapotranspiration.
- Water deficit.
- Soil moisture.

For metadata information and links to download the dataset, see annex **11.5.8**.

4.3.3 MOD15A2H.061: Terra Leaf Area Index/FPAR 8-Day Global 500m

From the MOD15A2H V6.1 MODIS I have extracted the Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (FPAR) products, which is an 8-day composite dataset at 500m resolution (**Table 2**). The algorithm chooses the "best" pixel available from all the acquisitions of the Terra sensor from within the 8-day period. The monthly average was calculated to harmonise to all datasets. The purpose of using these two variables is to evaluate the influence of these variables in the olive yield. For metadata information and links to download the dataset, see annex **11.5.9**.

4.3.4 MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m

From the MOD13Q1 V6.1 provides several vegetation indices from which I have drawn the NDVI and EVI. The MODIS NDVI and EVI products are computed from atmospherically corrected surface reflectance's that have been masked for water, clouds, heavy aerosols, and cloud shadows (**Table 2**). For metadata information and links to download the dataset, see annex **11.5.10**.

4.3.5 FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System

The FLDAS dataset (McNally et al., 2017) is a custom instance of the NASA Land Information System (LIS) framework. This dataset was designed for drought monitoring in Africa and assist with food security assessments in sparse data situation such as developing countries.

FLDAS uses the Noah Land Surface Model (NASA Land Information System), which simulates land surface states and fluxes, including soil moisture. It includes information on many climate-related variables and in this study the variables used are as follows (**Table 2**):

- Soil moisture content at 10cm depth
- Soil moisture content at 40cm depth
- Soil moisture content at 100cm depth
- Humidity
- Total precipitation

For metadata information and links to download the dataset, see annex **11.5.11**.

4.3.6 ERA5-Land Monthly Aggregated - ECMWF Climate Reanalysis

ERA5-Land monthly aggregated dataset is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades. This reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics (Muñoz-Sabater et al., 2021). For this study the bands were rescales accordingly to the scale used by the Earth Engine Data Catalog. The bands used for this study are as follows (**Table 2**):

- Temperature at 2 meters above the surface
- Temperature of the soil 7-28 cm
- Temperature of the soil 28-100 cm
- Temperature of the soil 100-289 cm
- Net thermal radiation at the surface
- Evaporation from bare soil at the top of the land surface
- Evaporation from the canopy interception reservoir at the top of the canopy
- Evaporation from vegetation transpiration)

For metadata information and links to download the dataset, see annex **11.5.12**.

5 Methodology

The work is aimed to forecast olive yield on olive crops from the Jaen province using Landsat-7, Landsat-8, Sentinel-2 and MODIS sensor reflectance data coupled with climate data such as precipitation, temperature, evaporation, soil moisture, drought index among others. Given the different time series variables, this correlation will be evaluated using multiple linear regression analyses grouped by family datasets.

Each family dataset is defined by the source of the data or type of sensor (in this study four remote sensors are used) and another family dataset is the meteorological parameters. Hence, there are five family datasets. The statistical time range considered for this study is from November 2001 to November 2020, as the harvest month for olive oil trees is in November for the Jaen province (see **3.1.5** and **figure 17**). Therefore, the month of November of the previous year (e.g., 2001) will affect the olive production of the following year (e.g., 2002).

5.1 Data curation and harmonisation before statistical analysis

Before running multivariate regression analysis to assess how those independent variables influence the olive yield, it is necessary to curate and harmonise all datasets. This process includes several steps. As a summary, in sequential order, they are as follows:

- Screening the Survey on Areas and Crop Yields (ESYRCE) and selection of the segments based on the criteria mentioned beforehand.
- Extract the yield data by segments.
- Redraw the polygons for the selected orchards (**Figure 23**, **Figure 25**, **Figure 29** and **Table 4**) to create individual shapefiles for each orchard.
- Upload those individual shapefiles into Google Earth Engine platform.
- Based on those shapefiles, extract the reflectance, SPEI, TerraClimate and MOD15A2H.061 data from GEE.
- Identification of outliers if any from those datasets extracted from GEE. No outliers were removed.
- Calculations of the vegetation indices.
- Resampling all raster datasets from the Regional Government into the same resolution 250m.
- Snap those raster datasets into the same reference point. In this case the reference raster used is the DEM 5m resolution resampled into 250m resolution.
- Based on the shapefiles of the selected orchards, extract in ArcGIS Pro the temperature, and precipitation from the raster datasets from the Regional Government.
- Extract from ArcGIS Pro altitude of the selected shapefiles with olive yield known.
- Calculate all datasets by monthly averages to harmonise data and the regression analysis.

5.2 Vegetation indices and meteorological data extracted with Google Earth Engine (GEE)

The first step was extracting the data using Google Earth Engine (GEE). In this study, the reflectance data meteorological data have been obtained by using Google Earth Engine (GEE). GEE is a cloud-based geospatial analysis platform that provides easy, web-based access to an extensive catalogue of satellite imagery and other geospatial data in an analysis-ready format. This service enables users to visualize and analyse satellite images. Google Earth Engine collects more than 40 years of historical and current global satellite imagery.

An online Integrated Development Environment (IDE) is the tool to access to the vast catalogue for making requests to the Earth Engine servers to visualise and running spatial analyses using the JavaScript API (Application Programmable Interface). As Google Earth Engine enables users to upload their own raster and vector data for analysis. I generated a code in JavaScript to extract top-of atmosphere (TOA) and surface reflectance data from Sentinel-2, Landsat-7, Landsat-8, MODIS sensors and the meteorological datasets

mentioned previously (**Datasets 4.3**). The JavaScript code written to extract the values from the datasets described in **table 2** is shown in the annexes (see annex **11.4** and for metadata see annex **11.5**).

The coding enables to upload the polygons that outline each olive orchard and the Jane province. For data extraction, the user selects in the coding the period of interest for each polygon. In this case, this period or time span correspond with the olive yield records obtained for each olive orchard or Jaen province (**Table 4**) from the Survey on Areas and Crop Yields (see further explanation in **4.2.6**). Also, the user decides when coding on the mathematical approach to treat all pixels covering the polygon of interest and those at the edges of the polygon of interest. In all datasets, the monthly average was the approach to extract and calculate the monthly values. Regarding the edge pixels, the algorithm used when running the statistics of each polygon is the "**ee.Reducer.mean()**" as a weighted reducer. This GEE method includes all pixels if at least 0.5% of the pixel is within the polygon as long as the image is not zero value. When using this method, a scale should be given to the extracted output dataset. This is because the source data and their analyses could involve data from diverse sources with different scales. Therefore, by defining beforehand the scale of the output dataset, that scale will not be established from the diverse inputs (see coding examples in **Google Earth Engine Scripts 11.4**). Automatically GEE, based on their pyramiding policy, establishes how pixels at a specified level of the pyramid is processed from the aggregation of a 2x2 block of pixels at the following lower level (**Figure 26**).

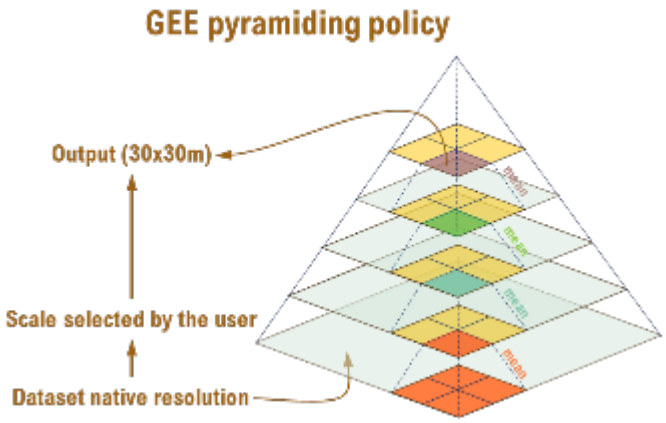


Figure 26 Sketch showing how GEE uses the scale specified by the output to determine the appropriate level of the image pyramid to use as input.

Regarding the remote sensors to retrieve reflectance, the atmospherically corrected surface reflectance was used for Landsat-7, Landsat-8, and MODIS sensors to reduce sensitivity to variations of aerosol atmospheric content.

By contrast, the top-of atmosphere (TOA) reflectance was the option used for Sentinel-2 to cover as many years as possible of the statistical time series. The reason is that Sentinel-2 Level-2A orthorectified atmospherically corrected surface reflectance dataset availability starts on 28-03-2017, covering only three years of the time range with crop yield data provided by the Spanish Ministry of Agriculture. The reason to use the Sentinel-2 regardless the shorter time span was to evaluate the impact in the analysis of the higher resolution of this sensor.

As Google Earth Engine (GEE) offers the raw values, for all reflectance data a scaling factor was applied for the Landsat-7 and 8 to convert the raw digital numbers (DNs) recorded by the Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and OLI/TIRS sensors into reflectance values. Also, an offset scale of -0.2 to correct for the along-track pointing error of the Landsat-7 satellite, which can cause the imagery to be misaligned. As Landsat-8 mission data availability starts on 2013, Landsat-7 was also used to some extent try to cover the time range from 2002 (first year of crop yield data availability) to have a continuous time series of vegetation indices (**Table 2**). Next, these vegetation indices were calculated by the date of acquisition and later reduce them by a monthly average to mathematically harmonise these index values with the statistical time range from November 2001 to November 2020 (see point **5.1** for further information).

The meteorological variables were already extracted from GEE as monthly mean values. Consequently, the 78 parameters obtained for the months of all years of the statistical series are as follows: 16 vegetation indices from MODIS, 12 vegetation indices from Landsat-7, 12 vegetation indices from Landsat-8, 12 vegetation indices from Sentinel-2, and 26 meteorological variables.

Hence, those 78 parameters give rise to 936 values parameter/year of the statistical time series to evaluate as predictors. Therefore, according to the olive yield records extracted from the Survey on Areas and Crop Yields (see **4.2.6**), the polygons of interest showed this number of years with yield records: ,orchard 5034172 retrieved 16 years, orchard 4134152 retrieved 15 years, orchard 4434212 retrieved 9 years, and the whole Jaen province retrieved 19 years of yield records. Later, these reflectance values and meteorological data were exported into Excel where the Savitzky-Golay and exponential smoothing techniques were used to identify and remove outliers in these datasets if any. Regarding the Savitzky-Golay filter, three moving windows of 5, 7 and 9 numbers were applied in the datasets. For the exponential smoothing technique, I used three exponential windows (5, 7 and 9) to smooth time series data, assigning exponentially decreasing weights to past observations. However, I could not find from my own any valid scientifically supported criteria to classify several data as outliners, so at this step I did not change or remove potential outliners in order to avoid data alteration. Moreover, I have to say that I just identified very few potential outliners that could not alter the final monthly averages.

Finally, vegetation indices were calculated by monthly average to mathematically harmonise these four different remote sensors by the same date format. The meteorological parameters were retrieved from GEE as monthly mean values.

5.3 Pearson's Correlation Coefficient Calculations

Next step was to calculate the correlation coefficients by evaluating all those 936 parameter/year values of the statistical time series as predictors against the olive yield. An example of how these correlation coefficients were calculated is shown in **figure 27**. Secondly, those with the highest correlation coefficient by parameter and month were selected to begin screening which ones showed the better correlation and identify the most important ones. These correlation coefficients are sorted by family dataset (by satellite sensor and meteorological data), highlighting the highest coefficient with the greenest and the lowest with the reddest colour and the corresponding month to each coefficient (**Figure 27**).

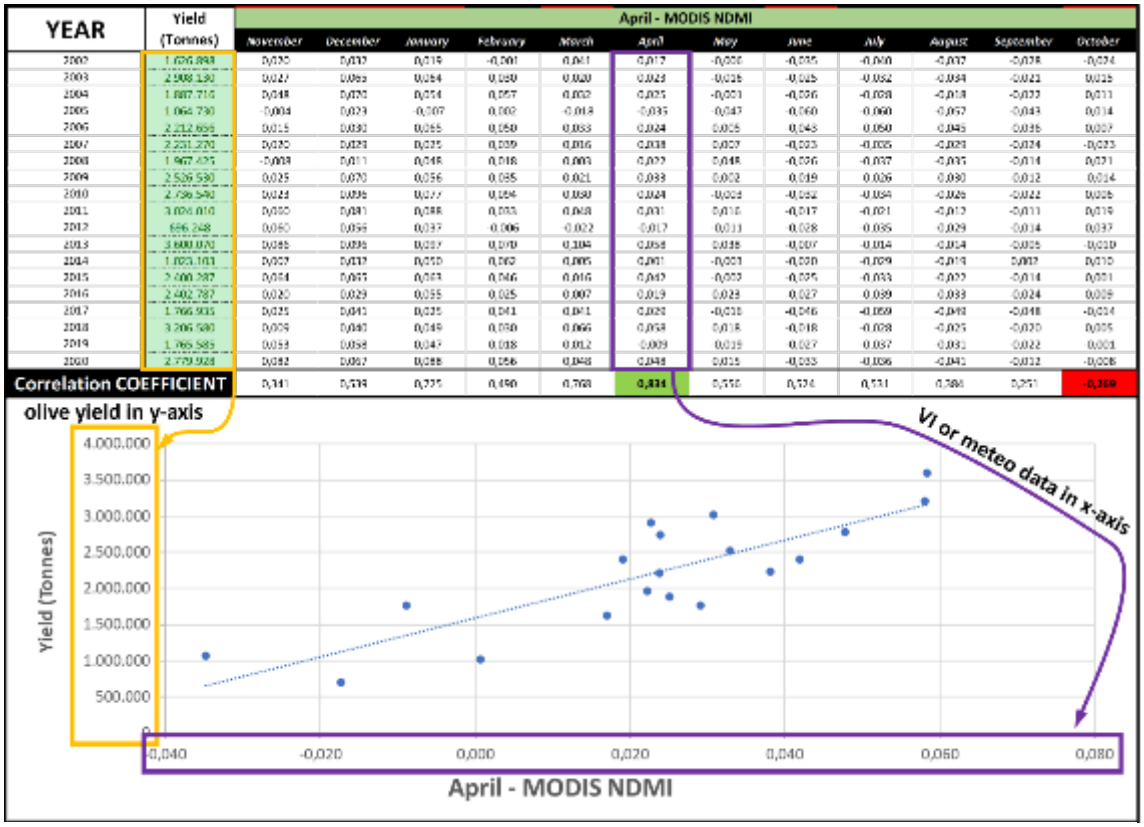


Figure 27 Screenshot of the spreadsheet layout used to calculate correlation coefficients. In this example, the monthly April mean values of NDMI extracted from MODIS are evaluated against the olive yield of the Jaen province

5.4 Multilinear regression background and methodology

Multiple linear regression is a statistical technique used to model the relationship between multiple independent variables (also known as x-predictors, features, or input variables) and a dependent variable (also known as the y-response or output variable).

This methodology enriches the concept of simple linear regression, which deals with only one independent variable and one dependent variable. In practice, multiple linear regression is normally used for prediction and inference in various fields, such as economics, finance, social sciences, and data analysis. It enables to understand the complex relationships among multiple variables. In this case, this research study is looking for relationship among the multiple variables such as reflectance, precipitation, temperature, and drought time series as independent variables with olive yield data from the Survey on Areas and Crop Yields (ESYRCE) Survey as the dependent variable. In multilinear regressions, the goal is to estimate the coefficients or weights associated with each independent variable that best explain the variation in the dependent variable, in this case the olive yield in a specific polygon.

5.4.1 Analysis of p-values

The first step ahead was to work out the p-values for all (936 statistical parameters corresponding to the correlation coefficients of each monthly values of the vegetation indices and meteorological parameters for all the February's, March's, Aprils, etc. of the statistical time series and yield production for each selected orchard and the whole olive crop area of the Jaen province) to determine statistical significance relationships of any those 936 potential predictors against the olive yield data.

The p-values for each parameter are better than the linear or simple correlation coefficients to evaluate the null (H_0) and the alternative (H_1) hypotheses. Whereas the null hypothesis (H_0) states no relationship exists between the two variables, the alternative hypothesis (H_1) confirms that the independent variable affected the dependent variable, and the results are significant in supporting that the results are not due to random chance.

By definition, p-value or probability value describes how likely the data would have occurred by random chance (i.e., that the null hypothesis is true) or the opposite (i.e., that the alternative hypothesis is true, and the null is false). In other words, the p-value quantifies the evidence against a null hypothesis. These thresholds determine if a parameter affects to the dependent variable and consequently, the smaller the p-value, the less likely the results happened by random chance, and the stronger is the evidence to assess that the null hypothesis should be rejected.

For this study, it has been used the most common significance thresholds: 0.05 and 0.01. A p-value of 0.01 or below means a highly statistically significance and represents robust evidence of a real effect rather than just random variation. Precisely, a p-value of 0.01 means there is only a 0.1% of chance to obtain a result at least as extreme as the one observed, assuming the null hypothesis is correct. Consequently, using the threshold of 0.05 broadens the chance to reject the null hypothesis.

5.4.2 Multilinear regressions approach

Following the calculation of p-values for the vegetation indices and meteorological datasets, the last step to identify the best predictors to forecast olive yield is the multilinear regression analysis.

The purpose of the multilinear regressions is to find which parameters or predictors described previously (the 936 statistical parameters) have relationship with the dependent variable, which in this case is the production data of the olive orchards supplied by the Spanish Ministry of Agriculture.

To avoid overfitting regression models, the multilinear regressions have been run step by step, or in other words adding in each step a new parameter to see how the R^2 and especially the adjusted R^2 evolve as an indicator that any new parameter added to the regression model enriches or improves the model. The risk on multilinear regression models is overfitting the model.

This happens when adding more independent variables to a regression model tends to increase the R^2, and consequently invites to add even more variables. This overfitting effect can return an unwarranted high R^2 value. Hence, it could be hard to finally identify which are the predictors that effectively improve the regression model.

The adjusted R^2 is used to determine how well grounded is the correlation and how much it is determined by the addition of independent variables. The adjusted R^2 is a modified version of R^2 because it is adjusted based on the number of predictors in the model and always increases when a new added term in the regression and betters the model more than would be expected by chance.

Typically, the adjusted R^2 is positive and is always lower than the R^2. As an example of how these regression models have been performed in this study for all areas of interest and dataset families, see on **table 5** the regression tree for the overall lowest p-value<0.01 in All olive Jaen's orchards. This multilinear regression model is made up of nine steps, adding a new predictor in each step. These predictors are selected from the lowest p-values to higher one. In this case, there are enough predictors with p-value<0.01, so no needed to use p-value <0.05.

From step 1 to step 6, the adjusted R^2 increases (see below in **figure 28**) as highlighted in orange in **table 5**. After step 6 with an adjusted R^2=0.7687 any combination of the following parameters with the following lowest p-values (all of them from MODIS sensor) moves down the adjusted R^2.

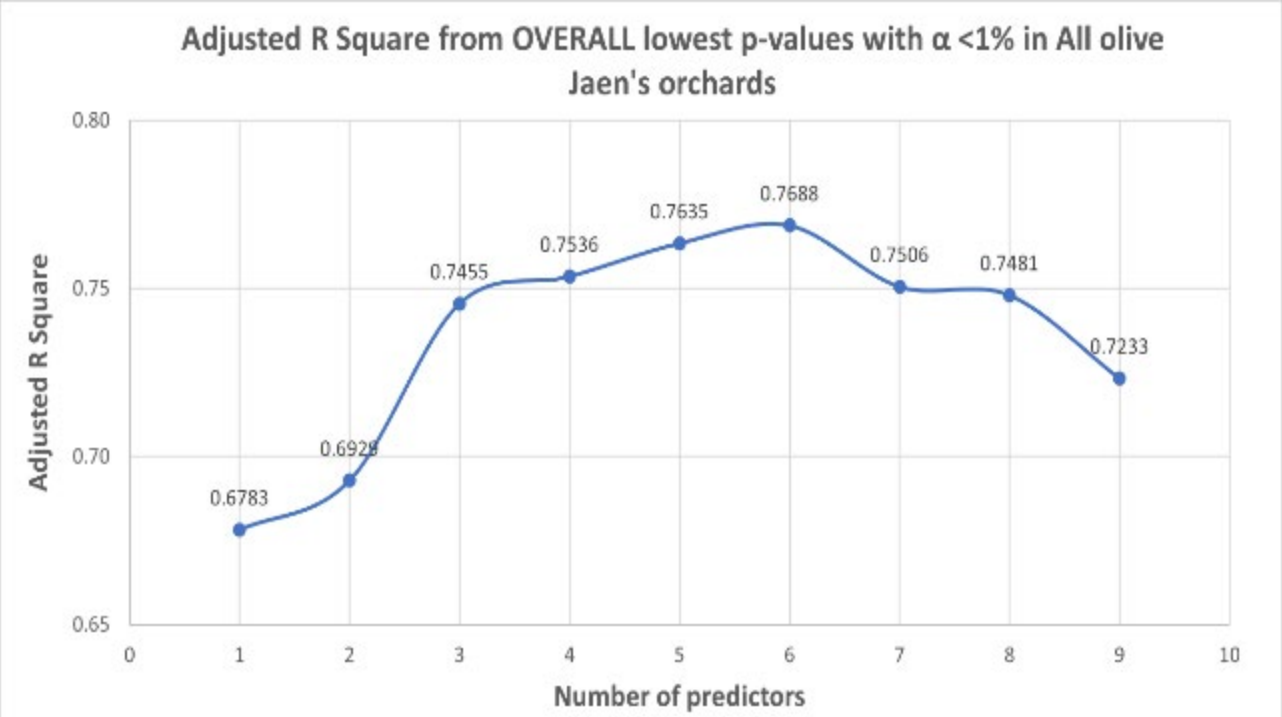


Figure 28 Adjusted R^2 value for the regression model of overall lowest p-value<0.01 in all olive Jaen's orchards. In table 4 is shown the evolution of the adjusted R square as the number of predictors increase

Even, in step 7 and 8 the predictors "MODIS EVI – April" and "NDVI from MODIS – April" were interchanged to see any increase in the adjusted R^2=0.7687 got in step 6 but any combination was unsuccessful. The last predictor added, "EVI from MODIS – April", decreased significantly the adjusted R^2 to 0.7233.

Therefore, the last three predictors: “MODIS EVI – April”, “NDVI from MODIS – April”, and “EVI from MODIS – April” did not improved the regression model. The adjusted R^2 increased until the sixth predictors but beyond this point no other new predictors added to the regression model enriches or improves the model (**Table 5**).

Adjusted R Square from OVERALL lowest p-values with α <1% in All olive Jaen's orchards

Multiple R	0.83							
R Square	0.70							
Adjusted R Square	**0.68**							
Standard Error	438503.72							
Observations	19							
STEPS	1							
Multiple R	0.83	0.85						
R Square	0.70	0.73						
Adjusted R Square	**0.68**	**0.69**						
Standard Error	438503.72	428418.50						
Observations	19	19						
STEPS	1	2						
Multiple R	0.83	0.85	0.89					
R Square	0.70	0.73	0.79					
Adjusted R Square	**0.68**	**0.69**	**0.75**					
Standard Error	438503.72	428418.50	389986.85					
Observations	19	19	19					
STEPS	1	2	3					
Multiple R	0.83	0.85	0.89	0.90				
R Square	0.70	0.73	0.79	0.81				
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**				
Standard Error	438503.72	428418.50	389986.85	383748.81				
Observations	19	19	19	19				
STEPS	1	2	3	4				
Multiple R	0.83	0.85	0.89	0.90	0.91			
R Square	0.70	0.73	0.79	0.81	0.83			
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**			
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16			
Observations	19	19	19	19	19			
STEPS	1	2	3	4	5			
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92		
R Square	0.70	0.73	0.79	0.81	0.83	0.85		
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**		
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77		
Observations	19	19	19	19	19	19		
STEPS	1	2	3	4	5	6		
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	
Observations	19	19	19	19	19	19	19	
STEPS	1	2	3	4	5	6	7	
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
Multiple R	0.83	0.85	0.89	0.90	0.91	0.92	0.92	0.92
R Square	0.70	0.73	0.79	0.81	0.83	0.85	0.85	0.85
Adjusted R Square	**0.68**	**0.69**	**0.75**	**0.75**	**0.76**	**0.77**	**0.75**	**0.75**
Standard Error	438503.72	428418.50	389986.85	383748.81	375986.16	371748.77	386082.73	388051.65
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8

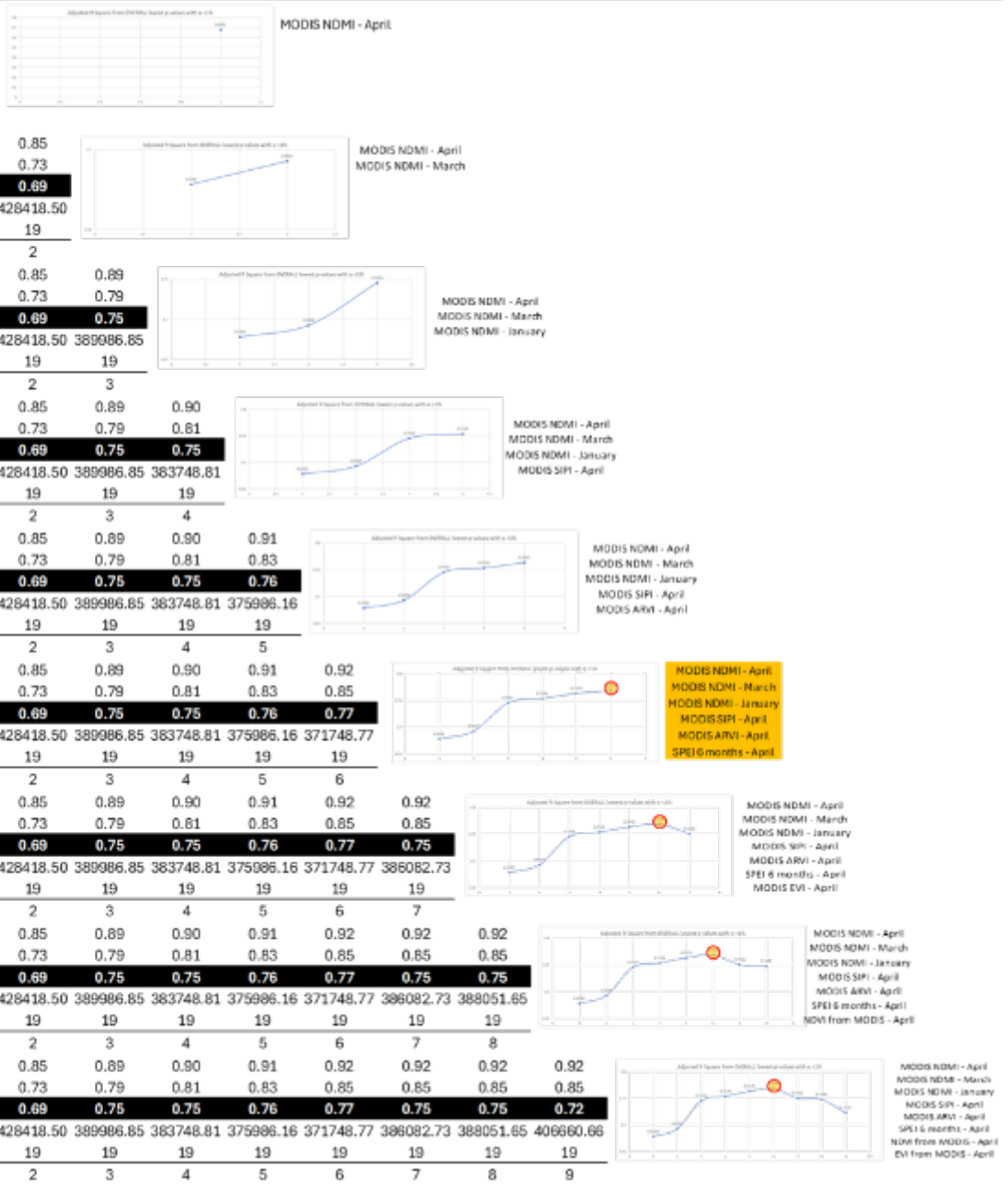


Table 5 Multilinear regression model using step by step the overall lowest p-value with α <1% in All olive Jaen's orchards. Graphics track the evolution of the adjusted R^2 value. Step 6 (highlighted in orange) marks the highest adjusted R^2 value, and the orange point tracks the decline of the adjusted R^2 value from step and seven and onwards.

6 Results

6.1 Results from the statistical correlation coefficients

Firstly, the monthly mean values of the VIs and meteorological data were calculated for each orchard and the whole olive crop area of the Jaen province covering the statistical time series of each one (**Table 4**). This step was conducted by extracting the mean of all pixels covering each orchard polygon and the polygon representing the Jaen province. Secondly, a statistical correlation was calculated between olive yield and monthly averages of the VIs and meteorological data for each orchard and the whole Jaen province. These correlation coefficients were analysed separately between vegetation indices extracted from the remote sensors (see annex **11.3**) and meteorological parameters selected from the different meteorological datasets described (see **4.3**).

6.1.1 Correlation coefficients of the vegetation indices in the selected polygons

The correlation coefficients of the VIs classified by the remote sensors is shown in Table 5. From this table, several assumptions can be drawn. The first one is that the highest correlation averages correspond to Sentinel-2 with 0.834 for the whole olive crop area of the Jaen province and 0.836 for the orchard 4134152 and 0.769 and 0.624 for the other three selected orchards. Sentinel-2 is followed by Landsat-8. On the other hand, MODIS and Landsat-7, which both cover the whole statistical time series of each selected polygon, show on average rather similar results, but MODIS is better than Landsat-7 in the whole olive crop area of Jaen, whereas Landsat-7 gives higher coefficients in the smaller orchards (**Figure 25**).

These facts indicate that the higher pixel resolution of Sentinel-2 (10m) and Landsat-8 (30m) are important in these results. On the other hand, the MODIS sensor with the lowest resolution (500m) retrieves higher correlation coefficients over large areas but gives the poorest results in smaller polygons where the pixel size is bigger than the orchard itself (**Figure 11**).

Focusing more specifically on the vegetation indices, the NDVI gives good correlations in the four sensors with Sentinel-2 showing the highest correlation coefficients as higher resolution gives higher correlations. The Green Normalized Difference Vegetation (GNDVI) also gives high correlation values with the highest resolution sensors such as Sentinel-2 and Landsat-8. Also, the Advanced Vegetation Index (AVI) works well in most sensors, except the Landsat-7. Regarding the most common months of high correlations, November, April, May, July, and August are the five most common ones for the vegetation indices. These correlations correspond to the facts that April is the month for sprouting to bud development, and July and August are the hottest and driest months.

One factor to consider is the number of observations. MODIS and Landsat-7 cover the whole statistical time series in all studied areas, whereas Landsat-8 and Sentinel-2 only cover 7 and 5 observations, respectively. It is assumed that the number of observations is directly proportional to increased uncertainty levels, so the lower number of observations, the higher uncertainty but easier to get higher correlation values. In order to compare how each sensor works and how the number of observations affects the correlation values with respect to the other sensors, it is needed to have shortened the longer time series to evaluate their suitability due to the difference in the number of observations.

Nonetheless, these correlation coefficients values give acceptable hints that olive yield and some of those 936 values parameter/year of the statistical time series have relationships. Therefore, to find how statistically significant a relationship is, p-values were calculated. Employing p-values will show which parameters or predictors are most suitable for multilinear regressions.

CORRELATION COEFFICIENTS SUMMARY STATISTICS										
All Jaen's orchards		HIGHEST CORRELATION COEFFICIENT PER MONTH	Orchard_4434212 (alt:429 m, slope:4.3)		Orchard_5034172 (alt:814 m, Slope:4.5)		Orchard_4134152 (alt:855 m, slope:20.7)		Orchard from Cubillas et al (2022) paper	
0,629	April	NDVI MODIS	0,697	May	0,255	January	0,512	July	0,595	May
0,703	April	EVI MODIS	0,506	May	0,096	August	0,383	May	0,438	August
0,678	April	EVI2 MODIS	0,567	May	0,142	January	0,386	August	0,428	August
0,834	April	NDMI MODIS	0,335	May	0,261	August	0,618	May	0,412	January
0,590	August	GNDVI MODIS	0,385	May	0,138	August	0,375	June	0,340	March
0,325	June	GCI MODIS	0,308	March	0,196	April	0,392	February	0,695	July
-0,009	October	SIPI MODIS	0,532	December	0,124	October	0,314	February	0,233	December
0,680	April	SAVI MODIS	0,558	May	0,136	January	0,384	August	0,419	August
0,715	April	ARVI MODIS	0,423	May	0,112	August	0,419	May	0,384	August
0,607	April	AVI MODIS	0,693	May	0,313	January	0,564	July	0,506	August
0,685	April	OSAVI MODIS	0,509	May	0,123	August	0,386	June	0,391	August
0,687	April	ReCI MODIS	0,497	May	0,125	August	0,436	September	0,397	August
0,692	April	NDVI from MODIS	0,485	May	0,039	August	0,468	May	0,487	November
0,688	April	EVI from MODIS	0,709	May	0,159	January	0,502	May	0,695	November
0,637	August	Frac. P.Active Radiation MODIS datasets	0,295	May	0,215	June	0,430	June	0,351	February
0,642	August	Leaf Index Area from MODIS datasets	0,197	May	0,287	June	0,500	June	0,369	April
0,612	**April**	**Average (whole orchard's time serie)**	**0,481**	**May**	**0,171**	**August**	**0,442**	**May**	**0,446**	**August**
0,542	June	NDVI LandSat7	0,546	June	0,297	May	0,674	May	0,526	May
0,232	February	EVI LandSat7	0,811	April	0,192	April	0,543	January	0,727	April
0,407	March	EVI2 LandSat7	0,694	April	0,161	May	0,237	May	0,680	February
0,088	November	NDMI LandSat7	0,067	October	0,081	October	-0,009	November	0,492	October
0,365	April	GNDVI LandSat7	0,423	April	0,182	April	0,087	August	0,752	February
0,253	July	GCI LandSat7	0,726	June	0,216	August	0,583	June	0,738	April
0,238	December	SIPI LandSat7	0,661	January	0,168	October	0,343	February	0,535	March
0,407	March	SAVI LandSat7	0,701	April	0,176	May	0,207	May	0,673	February
0,484	April	ARVI LandSat7	0,653	May	0,318	May	0,292	May	0,686	February
0,177	August	AVI LandSat7	0,709	November	0,239	Septembe	0,553	January	0,682	November
0,404	April	OSAVI LandSat7	0,677	April	0,176	April	0,195	May	0,706	February
0,431	April	ReCI LandSat7	0,583	April	0,193	May	0,246	May	0,760	February
0,336	**April**	**Average (whole orchard's time serie)**	**0,604**	**April**	**0,202**	**May**	**0,329**	**May**	**0,663**	**February**
0,790	May	NDVI LandSat8	0,558	January	0,765	Novembe	0,472	June	0,772	November
0,811	December	EVI LandSat8	0,435	July	0,211	February	0,769	September	0,786	November
0,818	July	EVI2 LandSat8	0,431	July	0,330	January	0,793	September	0,679	January
0,764	March	NDMI LandSat8	0,127	February	0,820	April	0,723	May	0,609	November
0,888	July	GNDVI LandSat8	0,439	July	0,447	January	0,894	September	0,699	January
0,566	July	GCI LandSat8	0,677	November	0,338	January	0,810	January	0,749	November
0,520	December	SIPI LandSat8	0,669	January	0,634	March	0,201	October	0,709	October
0,822	July	SAVI LandSat8	0,437	July	0,336	January	0,783	September	0,675	January
0,808	July	ARVI LandSat8	0,551	July	0,219	July	0,741	September	0,677	January
0,593	July	AVI LandSat8	0,567	December	0,664	Novembe	0,792	September	0,706	January
0,851	July	OSAVI LandSat8	0,473	July	0,325	January	0,778	September	0,664	January
0,867	July	ReCI LandSat8	0,513	July	0,188	July	0,764	September	0,686	January
0,757	**July**	**Average (n=7)**	**0,490**	**July**	**0,440**	**January**	**0,693**	**September**	**0,701**	**January**
0,838	May	NDVI Sentinel2	0,775	July	0,923	April	0,881	June	0,790	February
0,898	July	EVI Sentinel2	0,780	November	0,376	May	0,872	July	0,724	November
0,864	July	EVI2 Sentinel2	0,799	November	0,607	June	0,894	September	0,738	January
0,880	March	NDMI Sentinel2	0,574	April	0,835	April	0,626	October	0,831	February
0,907	August	GNDVI Sentinel2	0,835	July	0,881	Novembe	0,779	July	0,845	January
0,588	August	GCI Sentinel2	0,764	November	0,468	August	0,703	July	0,576	August
0,764	November	SIPI Sentinel2	0,795	December	0,618	August	0,876	November	0,734	August
0,869	July	SAVI Sentinel2	0,803	November	0,607	June	0,884	September	0,744	January
0,943	August	ARVI Sentinel2	0,668	November	0,440	Septembe	0,928	July	0,842	January
0,772	July	AVI Sentinel2	0,796	July	0,661	June	0,868	September	0,756	December
0,882	July	OSAVI Sentinel2	0,820	November	0,576	June	0,849	July	0,802	January
0,907	August	ReCI Sentinel2	0,818	November	0,497	June	0,878	July	0,870	January
0,834	**July**	**Average (n=5)**	**0,769**	**November**	**0,624**	**June**	**0,836**	**July**	**0,771**	**January**

Table 6 Vegetation indices correlation coefficients summary for each sensor in each selected orchard (the first three right-hand side columns) and the whole olive crop in Jaen province (left-hand side column). The last right-hand side column corresponds to the orchard from Cubillas et al (2022) where the olive yield was directly supplied by the farmer to the authors and used as validation model. The table shows the month with the highest correlation coefficient with the yield data supplied by the Spanish Ministry of Agriculture. Legend: the greener the highest correlation coefficient value and the redder, the lowest correlation coefficient value.

6.1.2 Correlation coefficients of the meteorological data in the selected polygons

Similar analysis was done with the correlation coefficients retrieved from meteorological datasets. The summary is shown in **table 7** shows those correlation coefficients emphasizing the highest coefficient with the greenest and the lowest with the reddest colour and besides the corresponding month to each coefficient (**Table 7**). The correlation coefficients between the monthly values retrieved from the meteorological data for all months of the statistical time series for each selected orchard and the whole olive crop area of the Jaen province are shown in **table 7**. At a glance, **table 7** shows that the water linked parameters such as drought index (SPEI), topsoil moisture, rainfall, specific humidity, and evapotranspiration give the highest correlation coefficients. By contrast, those parameters related to temperature such as temperature monthly average, soil temperature, and so forth provide lower correlation coefficients. This fact points out that rainfall or irrigation and water stress are more crucial factors than temperature.

An interesting parameter with a good correlation coefficient with the sprout and blooming months (March, April and May, see **table 7**) is the surface net thermal radiation from the ERA5-Land Monthly Aggregated (ECMWF Climate Reanalysis).

This parameter is the ratio between downward and upward thermal radiation at the surface of the Earth and can be linked with the number of fertile flowers. Higher thermal radiation can burn fertile flowers, subsequently reduce the amount of olive flowers that will finally produce the olive fruit and impact and very closely correlated with the final harvest (see points **1.1.2**, **3.1.1** and **3.1.4**), (Dawson, 2022; Roux, 2023).

CORRELATION COEFFICIENTS SUMMARY STATISTICS

All Jaen's orchards		METEO PARAMETERS WITH HIGHEST CORRELATION COEFFICIENTS PER MONTH	Orchard_4434212 (alt:429 m, slope:4.3)		Orchard_5034172 (alt:814 m, Slope:4.5)		Orchard_4134152 (alt:855 m, slope:20.7)		Orchard from Cubillas et al (2022)	
0,629	March	SPEI 1 month	0,668	May	0,330	May	0,461	April	0,798	April
0,680	April	SPEI 3 months	0,570	May	0,425	November	0,447	April	0,863	June
0,705	April	SPEI 6 months	0,514	July	0,365	November	0,520	April	0,799	September
0,652	May	SPEI 12 months	0,451	August	0,282	May	0,508	October	0,228	May
0,619	March	Precipitation Monthly Average (Raster data)	0,615	July	0,436	August	0,579	November	0,732	July
0,329	September	Temperature Monthly Average (Raster data)	0,591	September	0,328	December	0,326	November	0,619	January
0,401	September	Maximum Monthly Temperature (TerraClimate)	0,532	September	0,289	September	0,294	August	0,682	November
0,434	March	Minimum Monthly Temperature (TerraClimate)	0,329	January	0,504	January	0,429	December	0,539	January
0,261	September	Climate Water Deficit (TerraClimate)	0,771	September	0,449	August	0,212	February	0,664	August
0,569	March	Soil Moisture (TerraClimate)	0,069	November	0,081	October	0,496	October	-0,482	February
0,664	January	Actual Evapotranspiration (TerraClimate)	0,665	May	0,299	May	0,653	November	0,349	February
0,602	March	Monthly Rainfall (mm) (TerraClimate)	0,669	May	0,312	May	0,581	November	0,691	April
0,223	November	Vapor pressure deficit (kPa) (TerraClimate)	0,644	September	0,338	September	0,281	July	0,433	September
0,297	September	Temperature at 2m (ERA5)	0,620	September	0,382	January	0,334	November	0,606	December
0,258	December	Soil temperature (7-28 cm) (ERA5)	0,621	September	0,574	January	0,393	November	0,610	January
0,269	January	Soil temperature (28-100 cm) (ERA5)	0,661	October	0,506	January	0,404	March	0,597	January
0,188	March	Soil temperature (100-289 cm) (ERA5)	0,605	October	0,258	February	0,359	March	0,710	February
0,310	March	Evaporation from bare soil at the top of land surface (ERA5)	0,682	October	0,265	September	0,268	April	0,684	January
0,641	March	Net thermal radiation (ERA5)	0,588	May	0,491	January	0,484	April	0,612	April
0,461	June	Evaporation from vegetation transpiration (ERA5)	0,534	May	0,534	February	0,264	March	0,863	February
0,176	July	Evaporation from the top of canopy (ERA5)	0,641	September	0,243	June	0,177	July	0,472	November
0,625	April	Specific humidity (FLDAS)	0,840	July	0,585	January	0,564	August	0,562	January
0,508	August	Rainfall rate (kg/m^2/s) (FLDAS)	0,595	May	0,223	May	0,468	April	0,646	April
0,642	August	Soil Moisture 0-10 cm (FLDAS)	0,551	March	0,354	May	0,458	October	0,593	April
0,613	April	Soil Moisture 10-40 cm (FLDAS)	0,462	June	0,467	November	0,430	May	0,655	May
0,572	May	Soil Moisture 40-100 cm (FLDAS)	0,299	September	0,153	November	0,215	June	0,147	September
0,474	March	Average (whole orchard's time serie)	0,569	September	0,364	May	0,408	April	0,564	January

Table 7 Correlation coefficients summary of the meteorological datasets for each sensor in each selected orchard (the first three right-hand side columns) and the whole olive crop in Jaen province (left-hand side column). The last right-hand side column corresponds to the orchard from Cubillas et al (2022) where the olive yield was directly supplied by the farmer to the authors and used as validation model. The table shows the month with the highest correlation coefficient with the yield data supplied by the Spanish Ministry of Agriculture. Legend: the greener the highest correlation coefficient value and the redder, the lowest correlation coefficient value.

6.2 Results of p-values from vegetation indices and meteorological parameters

The analysis of the p-values for all parameters (936 statistical parameters corresponding to the correlation coefficients of each monthly values of the vegetation indices and meteorological parameters for all the February's, March's, April's, etc. of the statistical time series and yield production for each selected orchard and the whole olive crop area of the Jaen province) was used to determine the statistical significance relationships of any those 936 potential predictors against the olive yield.

In **table 8** is shown a summary significant p-values calculated for all 936 statistical parameters corresponding to the correlation coefficients of each monthly values of the vegetation indices and meteorological parameters of the statistical time series and yield production for each selected orchard and the whole olive crop area of the Jaen province.

AREAS	All Jaen's orchards		Orchard 4434212 (alt:429 m, slope: 4.3)		Orchard 5034172 (alt: 814 m, Slope: 4.5)		Orchard 4134152 (alt:855 m, slope: 20.7)	
Significance level (α)	p-values below significance level (α) <1%	p-values below significance level (α) <5%	p-values below significance level (α) <1%	p-values below significance level (α) <5%	p-values below significance level (α) <1%	p-values below significance level (α) <5%	p-values below significance level (α) <1%	p-values below significance level (α) <5%
Meteo (n=19)	41	101	2	12	0	0	0	5
MODIS (n=19)	41	81	0	4	0	0	2	8
LandSat7 (n=19)	0	8	1	12	0	3	1	11
LandSat8 (n=7)	1	17	0	1	0	2	1	8
Sentinel2 (n=5)	1	18	1	2	0	12	0	11
TOTAL	84	225	4	31	0	17	4	43

Table 8 Summary of p-values below significance level (α) <0.01 and <0.05 for each studied polygon and by data type

The first conclusion to draw from **table 8** is that the Sentinel-2 and Landsat-7 datasets and being Landsat-8 sensor with the lowest number of parameters below a significance level of 0.05. MODIS sensor seems working well in meteorological datasets only for the biggest polygon, the whole Jaen province.

By area, the whole olive crop area of the Jaen province yields the 75% of the lowest number of parameters below a significance level of 0.05 being followed by the largest orchard labelled as 4134152 (**Figure 29**, and left-hand side at **figure 25**). The other two orchards, much smaller barely deliver parameters below significance levels below 0.01 (**Figure 25**). These results are mirroring the correlation coefficients described previously basically by the mathematical derivation of the p-values calculations.

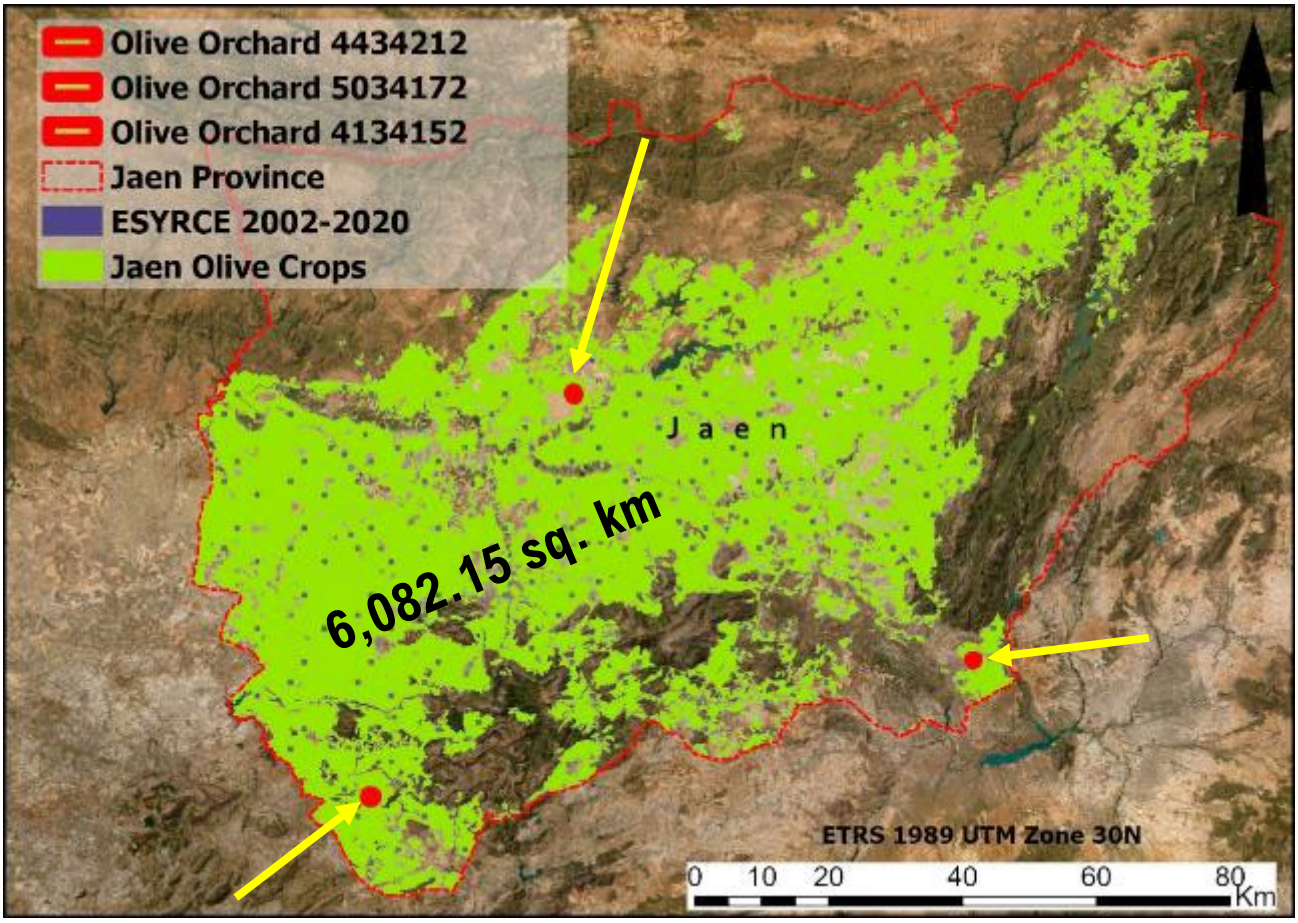


Figure 29 Olive crops in Jaen province. The yellow arrows indicate the location of the three orchards initially selected in this study (see Table 3, Figure 23, Figure 25, and Figure 34).

The size of the study area could play a role as the bigger the area, the more is the number of parameters below the significance level of 0.05 (**Table 8**).

Landsat-7, having the same number of observations that MODIS and meteorological data (n=19), yields an extremely small number of parameters below the threshold of 0.05 in the whole Jaen province in comparison with MODIS. On the other hand, Landsat 7 performs better than MODIS in the small polygons. Hence, the positive relationship between the number of observations and uncertainty is not clear at this case. In terms of monthly occurrence of p-values below $\alpha = <0.01$ and $\alpha = <0.05$ are shown in **figure 30** and **figure 31** reflect the same results shown in **figure 30** but here allocated by month and study areas. Again, the whole olive crops of the Jaen province yield the 70% of the parameters giving p-values below $\alpha = <0.01$ and $\alpha = <0.05$ and this is followed by the second largest polygon analysed labelled as 4134152 at the top (symbolised in orange) of **figure 30**.

Figure 31 clearly show that the highest occurrence for the statistical p-values below α = <0.01 and α = <0.05 are during the main important phenological months in the olive tree, especially during the sprouting to bud development, olive tree blooming, and fruit formation. Therefore, this fact indicates that the remote sensing datasets used to forecast olive yield are covering the most important periods for the olive crops, especially MODIS and meteorological datasets (**Table 8**).

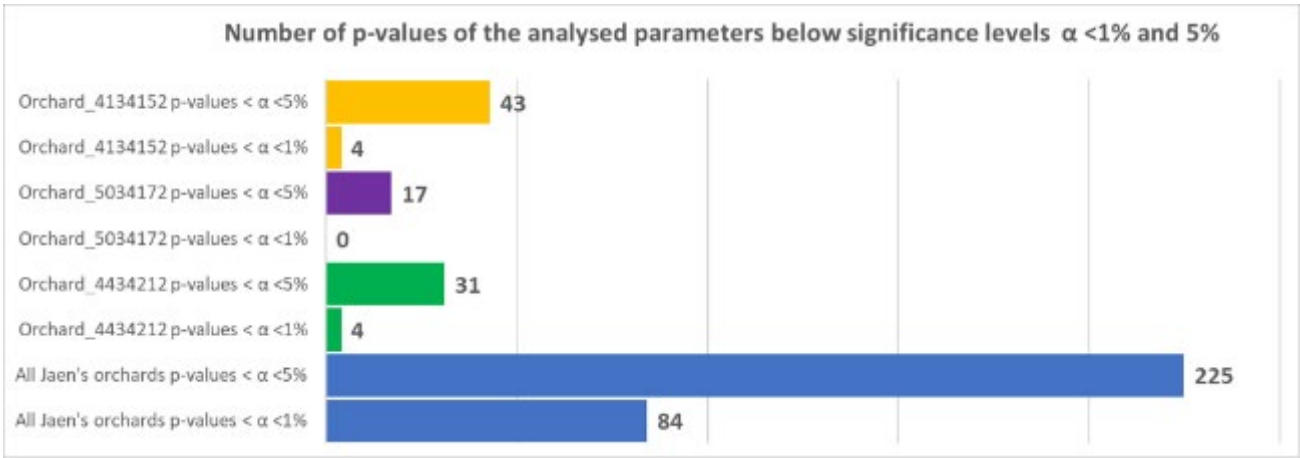


Figure 30 Number of p-values of the analysed parameters below significance levels α <1% and 5% distributed by the study areas.

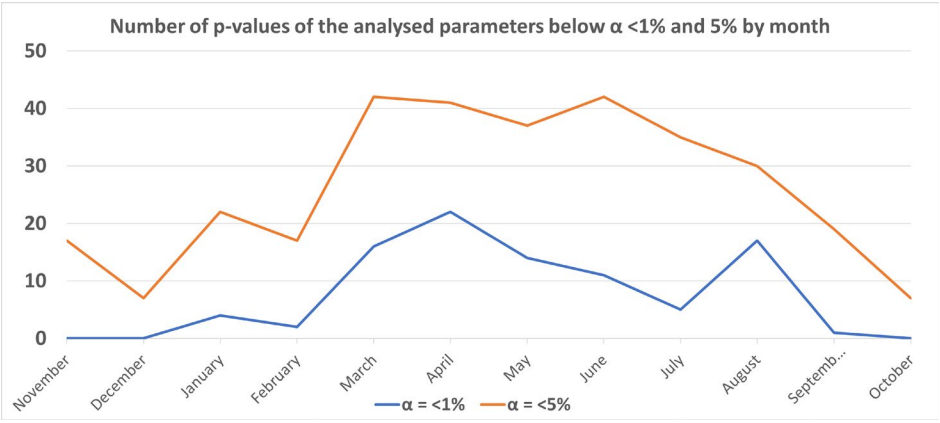


Figure 31 Number of p-values of the analysed parameters below significance levels α <1% and 5% distributed by the statistical time range from November to October

Consequently, this screening analysis based on the p-values delivered the best parameters among those 936 statistical parameters corresponding to the correlation coefficients of each monthly values of the vegetation indices and meteorological parameters for all months of the time series. The next step has been to run the multilinear regressions for each area of interest based of these selected parameters.

6.3 Results of the multilinear regressions on the selected olive orchards

In this study, twenty-four regression models have been run (Annexe **11.6**), separated and based on the selected areas (5 areas) and by dataset families (5 dataset families). Sometimes, there have been enough p-values α <0.01, so extending the analysis to p-values <0.05 is unnecessary. Also, the regression models labelled as "Overall lowest p-values" corresponds to multilinear regression analysis based on the lowest p-values independently the sensor or data type, parameters from satellite and climate data are combined. Here, the purpose is to select the parameters with the lowest p-values mixing meteorological data with vegetation indices.

Also, it is worthwhile to remind as mentioned in the literature review **1.2**, that the paper from (Cubillas et al., 2022) has been used to validate the regression models as the olive yield of this orchard was provided by the farmer and published in this paper as public domain data.

The reason to include the orchard from Cubillas et al. (2022) is that olive yield data, as supplied by the farmer (see further in heading **1.2**), is fully accurate.

Secondly, the farmer also delimits the orchard polygon, and there is no error from weird polygons, as seen in orchard 4434212, the second smallest orchard (see orchard on the left-hand side of **figure 34**).

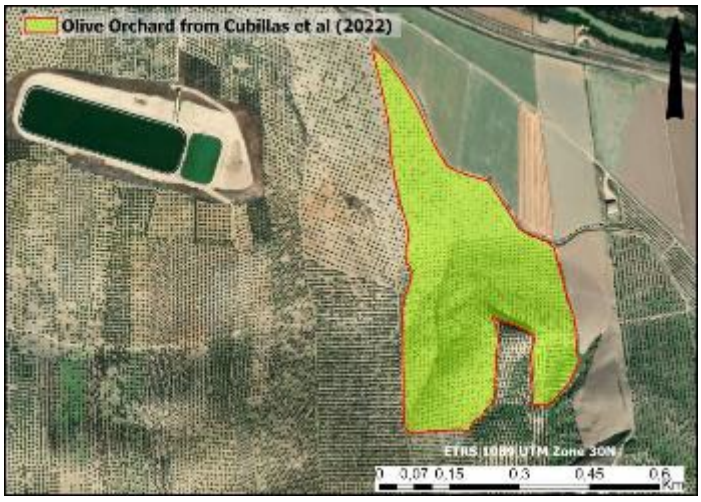


Figure 32 Orthophoto showing the orchard from Cubillas et al. (2022) with the olive yield data supplied by the farmer

These facts make this orchard perfect to compare with the yield data supplied by the Spanish Ministry of Agriculture and the whole olive yield for the whole Jaen province. The best multilinear regression models with a significance level <0.01 for the three selected areas are shown below (**Figure 33**).

Adjusted R Square from OVERALL lowest p-values with α <1% in All olive Jaen's orchards						
Multiple R	0,834	0,853	0,888	0,899	0,911	0,920
R Square	0,696	0,727	0,788	0,808	0,829	0,846
Adjusted R Square	**0,678**	**0,693**	**0,746**	**0,754**	**0,763**	**0,769**
Standard Error	438503,716	428418,505	389986,855	383748,812	375986,164	371748,774
Observations	19	19	19	19	19	19
STEPS	1	2	3	4	5	6
X Predictors						
	MODIS NDMI - April	MODIS NDMI - April MODIS NDMI - March	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April SPEI 6 months - April

OVERALL lowest p-values with α <1% Orchard from Cubillas et al (2022)				
Multiple R	0,912	0,951	0,968	0,998
R Square	0,831	0,904	0,936	0,997
Adjusted R Square	**0,803**	**0,865**	**0,889**	**0,992**
Standard Error	1882,222	1558,381	1414,563	377,007
Observations	8	8	8	8
STEPS	1	2	3	4
X Predictors				
	Leaf Index Area - June	Leaf Index Area - June Evap trans (ERA5) - February	Leaf Index Area - June Evap trans (ERA5) - February SPEI 3 months - June	Leaf Index Area - June Evap trans (ERA5) - February SPEI 3 months - June LandSat7 SIPI - December

OVERALL lowest p-values with α <1% Orchard 4134152				
Multiple R	0,628	0,895	0,970	1,000
R Square	0,394	0,801	0,941	1,000
Adjusted R Square	**0,243**	**0,668**	**0,852**	**0,998**
Standard Error	903,272	598,132	399,345	44,425
Observations	6	6	6	6
STEPS	1	2	3	4
X Predictors				
	LandSat7 NDMI - March	LandSat7 NDMI - March LandSat8 NDMI - February	LandSat7 NDMI - March LandSat8 NDMI - February Sentinel2 SIPI - July	LandSat7 NDMI - March LandSat8 NDMI - February LandSat7 NDVI - May Sentinel2 SIPI - July

Figure 33 Multilinear regression models with a significance level <0.01 for the three final selected areas.

These ones correspond to multilinear regression where meteorological data and vegetation indices parameters have been mixed based on the lowest p-values criteria. Other regression models for these two areas use a significance level between 0.01 > α <0.05 with fewer predictors. The three regression models with p-values α <0.01 give a promising approach to accurately forecast olive crop yield from exclusively remote sensing datasets such as vegetation indices and climate data. The two small orchards reach R^2 values near 0.99, while the largest polygon, the olive crops of Jaen province, yield an R^2 value near 0.85.

Also, looking at the most synchronized months and parameters among these three models, there is a disparity that is difficult to harmonise in one model. Overall, looking at the whole picture, it means that the regression model for the whole Jaen olive crops (at the top of **figure 33**) at the end of April, the regression model could be useful as an early assessment of 6 months in advance for the olive yield with R^2 values near 0.85.

In other cases, it has not been possible to make regression models as no p-values α <0.05 were retrieved from the statistical analysis (grey columns in **table 9**), such as for the orchard from Cubillas et al. (2022) with MODIS, Landsat-8 and Sentinel-2, the orchard 4434212 with Landsat-8 data or olive orchard 5034172 for MODIS and Landsat (**Table 9**).

All of them have been run using the same approach explained previously (see point **5.4.2**) and shown in **table 5** and **figure 28**. The behaviour of the adjusted R^2 has been the main parameter to avoid overfitting, as explained in **Multilinear regressions 5.4.2**. In **table 9** are shown the 24 regression models that have been run (Annexe **11.6**) by selected area (5 areas) and by dataset families (5 dataset families).

Orchard from Cubillas et al (2022)

Sensors&Parameters	Overall lowest p-values	MODIS	Landsat 7	Landsat 8	Sentinel 2	Climate datasets
R Square	0.997		0.91			0.93
Significance level (α)	<1%		1% > α <5%			1% > α <5%
X Predictors number	4		3			3
Data resolution (m)	Variable		30			1000
Observations (n)	8		8			8
Time range (years)	2014-2021		2014-2021			2014-2021

All Jaen's olive orchards

Sensors&Parameters	Overall lowest p-values	MODIS	Landsat 7	Landsat 8	Sentinel 2	Climate datasets
R Square	0.85	0.93	0.76	0.92	0.99	0.70
Significance level (α)	<1%	<1%	1% > α <5%	1% > α <5%	1% > α <5%	<1%
X Predictors number	6	5	7	3	3	4
Data resolution (m)	Variable	500	30	30	10	1000
Observations (n)	19	19	19	8	5	19
Time range (years)	2002-2020	2002-2020	2002-2020	2013-2020	2015-2020	2002-2020

Olive orchard 4434212

Sensors&Parameters	Overall lowest p-values	MODIS	Landsat 7	Landsat 8	Sentinel 2	Climate datasets
R Square	0.998	0.92	0.96		0.93	0.99
Significance level (α)	1% > α <5%	1% > α <5%	1% > α <5%		1% > α <5%	1% > α <5%
X Predictors number	6	2	3		1	7
Data resolution (m)	Variable	500	30		10	>500
Observations (n)	9	9	9		5	9
Time range (years)	2011-2012 & 2014-2020	2011-2012 & 2014-2020	2011-2012 & 2014-2020		2014-2020	2011-2012 & 2014-2020

Olive orchard 5034172

Sensors&Parameters	Overall lowest p-values	MODIS	Landsat 7	Landsat 8	Sentinel 2	Climate datasets
R Square	0.999			0.67	0.96	0.67
Significance level (α)	1% > α <5%			1% > α <5%	1% > α <5%	1% > α <5%
X Predictors number	3			1	2	4
Data resolution (m)	Variable			30	10	>500
Observations (n)	5			7	5	16
Time range (years)	2016-2020			2014-2020	2016-2020	2003-2009, 2011-2012, 2014-2020

Olive orchard 4134152

Sensors&Parameters	Overall lowest p-values	MODIS	Landsat 7	Landsat 8	Sentinel 2	Climate datasets
R Square	0.9996	0.782	0.868	0.998	0.9999979	0.819
Significance level (α)	<1%	1% > α <5%	1% > α <5%	1% > α <5%	1% > α <5%	1% > α <5%
X Predictors number	4	5	6	4	4	7
Data resolution (m)	Variable	500	30	30	10	1000
Observations (n)	6	14	14	7	6	14
Time range (years)	2015-2020	2002-2008, 2014-2020	2002-2008, 2014-2020	2014-2020	2015-2020	2002-2008, 2014-2020

Table 9 Summary of the twenty-four regression models distributed by area and dataset families. The first column corresponds to multilinear regression analysis based on the lowest p-values independently the sensor or data type, parameters from satellite and climate data are combined. Orchards highlighted with a red square and grey-shaded 4434212 and 5034172 were discarded for the final discussion to minimise uncertainties by removing potential errors.

As summary, several facts can be extracted from **table 9**:

- All study areas retrieved regression models with R^2 values above 0.90 with Sentinel-2 data.
- Regression models based on overall lowest p-values for all areas retrieved R^2 values above 0.85.
- All orchards, by average, their regression models retrieve R^2 values above 0.82.
- The orchard 4134152, the second largest (right-hand side at **figure 25**), shows the highest R^2 values.
- By dataset families, climate datasets and MODIS retrieve overall the poorest R^2 values below 0.84.
- Fewer variable conditions included (observations) obtain higher R^2 values.
- Lower-resolution datasets reach lower R^2 values.

As it can be seen in **table 9**, the regression models based only on p-values α <0.01 (in **table 9**, the first column on the left-hand side, where satellite and climate data are mixed and independently the sensor or data type) retrieved by average in the three areas R2 values above 0.94. A slightly higher R^2 average value is obtained by Landsat-8 with 0.96, whereas the highest R^2 values for Sentinel-2.

The **figure 33** show the combination of VI and/or climate plus sensor parameter used for each of them. For the "All Jaén orchards p-values α <1%", NDMI calculated from MODIS is by far the best indicator. Moreover, linked with SIPI, ARVI and SPEI this combination could be used as an early warning to farmers in April, so 7 months in advance of the harvest month to allow farmers to predict production and take measures in case they were necessary such irrigation.

For the model “orchard 4134152 p-values < α <1%”, again the NDMI from Landsat 7 AND 8 is the most reliable index followed by the SIPI and NDVI. This combination could give an early warning to farmers in May, 6 months in advance of the harvest month.

The Orchard from Cubillas et al (2022) p-values α <1% was used to validate the approach showing high adjusted R^2 values above 0.90 at the same level as orchard 4134152 with the same number of predictors.

7 Discussion

7.1 Data quality and resolution

The first step was to make an overall evaluation of the statistical correlation coefficients obtained and see main trend or weird discrepancies among the studied polygons. Those correlation coefficients shown in **table 6** and **table 7** clearly indicate that the polygons corresponding to the orchards 5034172 and 4434212 (the first two orchard from the left **figure 25** and **figure 34**) show much lower values than orchard 4134152 (orchard on the right-hand side of **figure 25**) or the orchard from Cubillas et al. (2022) (**Figure 32**) and even much lower when compared with the whole olive crops of the Jaen province (**Figure 29**).

This is highlighted because the smallest orchard (5034172) gives the lowest correlation values (see the orchard on the right-hand side of **figure 34**). As seen in **figure 34**, the MODIS resolution (500m) is much bigger than the orchard 5034172, which covers large nearby surrounding areas.

Orchard (5034172) gives the lowest coefficient values with MODIS (0.171) but higher coefficient values with higher pixel resolution such as Sentinel-2 with an average of 0.624 (**Table 6**). The same trend is shown with Landsat-7 and meteorological data with a variable resolution from 500m to 55,660m for SPEI. This fact is reflected in **table 9**, where no p-values <0.05 were retrieved for MODIS and Landsat-7 and climate datasets, and Landsat-8 obtained R^2 below 0.68.

Also, there are a few uncertainties about orchard 4434212, the second smallest orchard (see orchard on the left-hand side of **figure 34**). The first one is its weird polygon shape, which partially cover orchard 1 (blue dash line in **figure 34**) by splitting diagonally this orchard 1 and take the top right corner of orchard 2 (purple dash line in **figure 34**), which is different according the orthophoto. Secondly, its yield data supplied by the Ministry when compared with its extent which seems higher. Moreover, for orchard 4434212, no p-values were retrieved from Landsat-8, and the regression models for Sentinel-2, Landsat-7 and MODIS were run with only 1, 3 and 2 predictors, respectively. Therefore, very few predictors hamper the proper assessment of a robust and consistent model. The issue with its size could be an error of the dataset and I did not fix it to not alter the original data.

These considerations convinced to discard the results from these two orchards: 4434212 and 5034172 to minimise uncertainties by removing potential errors sourced from the yield dataset or because of orchard size against sensor resolution issues (see **table 9**, orchards highlighted with a red square and grey-shaded 4434212 and 5034172). Hence, it was decided to take these orchards (4434212 and 5034172) out of the final analysis and conclusions and only consider the biggest individual orchard labelled as 4134152, the orchard from Cubillas et al. (2022) as validation model, and the whole olive crop of the Jaen province.

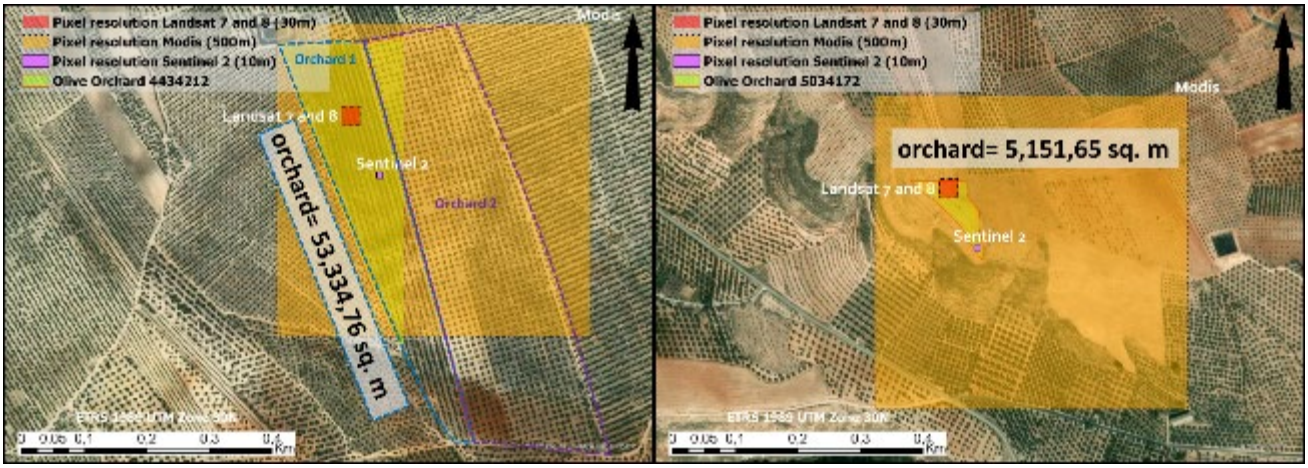


Figure 34 Location and size of the two smallest orchards overlapped with the pixel resolution of the four satellite sensors. On the left, the size issues explained in the text by orchard 4434212 depicted with the blue and purple dash lines.

7.2 Vegetation indices

About the vegetation indices, NDVI and NDMI are the best indicator among all 12 vegetation indicators, suitable for predict olive yield. The predictabilities of NDVI and NDMI from four platforms, Landsat-7, -8, Sentinel-2, and MODIS are more or less similar. For other indices, there is not a common indicator can be used for olive yield prediction, and difficult to find a clear pattern to describe which indices have high statistical significance. Anyhow, it can be said that ARVI is the best for Sentinel-2, whereas EVI and AVI work well with MODIS, and EVI and GCI gives higher correlations with Landsat-7. Moreover, it is important to remember that MODIS and Landsat-7 are based on the whole statistical time series (from 14 to 19 observations), which could make comparing sensors problematic. For Sentinel-2, the coefficient average is above 0.83, and this coefficient decreases as pixel resolution decreases with the other sensors (**Table 6**). However, in a broad sense, it can be highlighted as follows:

- NDVI is suitable for any sensor.
- NDMI plays a key role in the final regression models and the validation orchard.
- EVI and AVI from MODIS
- EVI and GCI from Landsat-7
- OSAVI and ReCI from Landsat-8 and Sentinel-2
- ARVI, which is an enhancement to the NDVI, is the best for Sentinel-2

The soil-adjusted vegetation OSAVI gives better results than SAVI (see definitions in point **2.2.1**), probably because of the nature of the olive orchard, where the canopy area of olive trees only covers 27% of the whole orchard (see point **3** and **figure 10**), and the soil is well exposed.

7.3 Meteorological data

For climate parameters, water-based parameters such as the SPEI drought index (especially the SPEI 3 and 6 months), topsoil moisture, rainfall, specific humidity, and evapotranspiration clearly show better correlations than others.

On the other hand, those temperature-based parameters, such as monthly average temperature, soil temperature, etc., appear to be more negligible when assessing to forecast olive yields because water stress is more important than temperature (**Table 7**). However, there is an exception in the surface net thermal radiation. In the absence of extensive research that is beyond the scope of this study, a hypothesis could be that surface net thermal radiation measurements are not affected by wind chill conditions or evapotranspiration which on the other hand can alter readings of measuring instruments on ground temperatures.

Hence, surface net thermal radiation measurements could be a better indicator of the thermal stress over fertile flowers and how many fertile flowers could be burned and dried because of this thermal stress. Although, the olive tree can withstand with summer elevated temperatures quite well but the blossom out time is a sensitive timing and temperatures even around 30ºC seriously affect blooming in olive trees.

As mentioned in point **3.1.2** the authors F. Aguilera & Valenzuela (2009) demonstrated by statistical analysis that cumulative solar radiation is one of the meteorological parameters that most affect olive floral phenology, and hence yield production.

8 Conclusions

The several conclusions that can be drawn from the final results **table 9** and **figure 33**. Sentinel-2 data was in general highly correlated with olive yield, while relationships were weaker for MODIS, followed by Landsat-7 and climate datasets. Here, there is strong directly proportional relationship between resolution and correlation as the higher the resolution, the higher the square R. On the other hand, there is another two factors showing a clear trend in the statistical analyses. Both olive yield observations (length of the time series) and predictors show an inversely proportional weight, so fewer observations or predictors means, higher square R. As mentioned previously, this could be because fewer observations reduce uncertainty.

As mentioned above, the facts and conclusions drawn from **table 9** do not show a clear, constant and recurrent parameter or predictor in all orchards that can be used to predict olive oil production. This circumstance could be related to the climatological conditions of each year. For instance, with normal rainfall rates but extremely hot conditions, most probably the NDMI would not be as relevant as the surface net thermal radiation from the "ERA5-Land Monthly Aggregated" (ECMWF Climate Reanalysis). Ideally, the second step would be to identify the best combination of VI+ sensor to get the best results. However, this combination seems that would change every year depending on the meteorological conditions given each phenological year.

Therefore, due to the yearly climatic variability the approach is to highlight which are the most accurate sensors, best vegetation indices and most relevant climate parameters to forecast olive oil yield. Anyhow, overall, water-based parameters are the dominants giving an early warning to farmers in June, 4 months in advance of the harvest month. This fact underpins the regression model of this research against the results retrieved by Cubillas et al. (2022), where the olive yield data is considered as ground truth.

9 Limitations of this study and potential pathways in future studies

Therefore, these results answer positively the two questions proposed for this research study (see **Rationale 1.1**, and **Objectives and purpose of this M.Sc 1.3**). In summary, both questions could be merged into only one: "How accurately do vegetation indices and weather data extracted from satellite sensors forecast olive crop yield?" In the absence of further studies, it is reasonable to say that VI's and weather data works fairly well to assess olive yield predictions. However, to develop a better model several limitations encountered in this study should be overcome. The main limitations could be focused on these two aspects:

- Lack of access to ground truth data from farmers to get real production and any other data such as rainfall, or crop conditions.
- As shown in the conclusions, Sentinel-2 is the best sensor to increase accuracy, but Sentinel-2 data begins in 2015, so the length of this study could be considered a little bit limited on time.

Consequently, the potential pathways to improve and increase model accuracy must be focused on overcoming these limitations. Hence, for future studies, the access to farmers (the more, the better and distributed in different areas and altitudes) to get exact production, rainfall data and crop conditions through the year will be an excellent step to overcome the limitations described in this study. Given that Sentinel-2 is the best sensor but limited data since 2015, it would be valuable to expand the study to 2023, 2024 to validate the models, and also using newer generations of sensors. This could be the best way to find a constant and recurrent combination of VI and/or climate plus sensor parameters that could give accurate forecast independently on the meteorological conditions given each phenological year.

10 Bibliography

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 170191. https://doi.org/10.1038/sdata.2017.191

Abbasi, N., Nouri, H., Didan, K., Barreto-Muñoz, A., Borujeni, S. C., Opp, C., Nagler, P., Thenkabail, P. S., & Siebert, S. (2023). Mapping Vegetation Index-Derived Actual Evapotranspiration across Croplands Using the Google Earth Engine Platform. *Remote Sensing*, *15*(4), 1017. https://doi.org/10.3390/rs15041017

Abdul-Jabbar, T. S., Ziboon, A. T., & Albayati, M. M. (2023). Crop yield estimation using different remote sensing data: literature review. *IOP Conference Series: Earth and Environmental Science*, *1129*(1), 012004. https://doi.org/10.1088/1755-1315/1129/1/012004

Achmakh, L., Janati, A., Boullayali, A., ElHassani, L., & Bouziane, H. (2020). Forecasting olive (Olea europaea L.) production using aerobiological and meteorological variables in Tétouan (NW Morocco). *Aerobiologia*, *36*(4), 749–759. https://doi.org/10.1007/s10453-020-09665-5

Aguilera, F., & Ruiz-Valenzuela, L. (2013). Forecasting olive crop yields based on long-term aerobiological data series and bioclimatic conditions for the southern Iberian Peninsula. *Spanish Journal of Agricultural Research*, *12*(1), 215–224. https://doi.org/10.5424/sjar/2014121-4532

Aguilera, F., & Valenzuela, L. R. (2009). Study of the floral phenology of Olea europaea L. in Jaén province (SE Spain) and its relation with pollen emission. *Aerobiologia*, *25*(4), 217. https://doi.org/10.1007/s10453-009-9127-5

Aguilera, F., & Valenzuela, L. R. (2012). Microclimatic-induced fluctuations in the flower and pollen production rate of olive trees (Olea europaea L.). *Grana*, *51*(3), 228–239. https://doi.org/10.1080/00173134.2012.659203

Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, *25*(3), 711–716. https://doi.org/10.1016/j.ejrs.2022.04.006

Alimentación, S. de A., Pesca y. (2019). *Análisis de las plantaciones de olivar en España* (Encuesta sobre Superficies y Rendimientos de Cultivos en España (ESYRCE), p. 24). Secretaría General de Agricultura y Alimentación Servicio de Estudios y Estadísticas. https://cpage.mpr.gob.es

Amankulova, K., Farmonov, N., Mukhtorov, U., & Mucsi, L. (2023). Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology. *Geocarto International*, *38*(1), 2197509. https://doi.org/10.1080/10106049.2023.2197509

Anastasiou, E., Balafoutis, A. T., & Fountas, S. (2023). Trends in Remote Sensing Technologies in Olive Cultivation. *Smart Agricultural Technology*, *3*, 100103. https://doi.org/10.1016/j.atech.2022.100103

Andalucia, J. de. (2023). *Primera Estrategia Andaluza para el sector del olivar 2023- 2027*. https://www.juntadeandalucia.es/sites/default/files/2023-11/Doc_resumen_%2016112023.pdf

Beguería, S., Serrano, S. M. V., Reig-Gracia, F., & Garcés, B. L. (2023). SPEIbase v.2.8 [Dataset]. *DIGITAL.CSIC*.

Besnard, G., Casas, R. R. de, Christin, P.-A., & Vargas, P. (2009). Phylogenetics of Olea (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: Tertiary climatic shifts and lineage differentiation times. *Annals of Botany*, *104*(1), 143–160. https://doi.org/10.1093/aob/mcp105

Bharadiya, J. P., Tzenios, N. T., & Reddy, M. (2023). Forecasting of Crop Yield using Remote Sensing Data, Agrarian Factors and Machine Learning Approaches. *Journal of Engineering Research and Reports*, *24*(12), 29–44. https://doi.org/10.9734/jerr/2023/v24i12858

Bojanowski, J. S., Sikora, S., Musiał, J. P., Woźniak, E., Dąbrowska-Zielińska, K., Slesiński, P., Milewski, T., & Łączyński, A. (2022). Integration of Sentinel-3 and MODIS Vegetation Indices with ERA-5 Agro-Meteorological Indicators for Operational Crop Yield Forecasting. *Remote Sensing*, *14*(5), 1238. https://doi.org/10.3390/rs14051238

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, *173*, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Buters, J., Clot, B., Galán, C., Gehrig, R., Gilge, S., Hentges, F., O'Connor, D., Sikoparija, B., Skjoth, C., Tummon, F., Adams-Groom, B., Antunes, C. M., Bruffaerts, N., Çelenk, S., Crouzy, B., Guillaud, G., Hajkova, L., Seliger, A. K., Oliver, G., … Stjepanovic, B. (2024). Automatic detection of airborne pollen: an overview. *Aerobiologia*, *40*(1), 13–37. https://doi.org/10.1007/s10453-022-09750-x

Carreño-Conde, F., Sipols, A. E., Blas, C. S. de, & Mostaza-Colado, D. (2021). A Forecast Model Applied to Monitor Crops Dynamics Using Vegetation Indices (NDVI). *Applied Sciences*, *11*(4), 1859. https://doi.org/10.3390/app11041859

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, *113*(5), 893–903. https://doi.org/10.1016/j.rse.2009.01.007

Chen, B., Jin, Y., & Brown, P. (2019). An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, *156*, 108–120. https://doi.org/10.1016/j.isprsjprs.2019.08.006

Chen, R., Yin, G., Liu, G., Li, J., & Verger, A. (2020). Evaluation and Normalization of Topographic Effects on Vegetation Indices. *Remote Sensing*, *12*(14), 2290. https://doi.org/10.3390/rs12142290

Cubillas, J. J., Ramos, M. I., Jurado, J. M., & Feito, F. R. (2022). A Machine Learning Model for Early Prediction of Crop Yield, Nested in a Web Application in the Cloud: A Case Study in an Olive Grove in Southern Spain. *Agriculture*, *12*(9), 1345. https://doi.org/10.3390/agriculture12091345

Dawson, D. (2022, December 30). *Olive Council Forecasts Significant Production Decline*. Olive Oil Times. https://www.oliveoiltimes.com/world/olive-council-forecasts-significant-production-decline/115544

Debaeke, P., Attia, F., Champolivier, L., Dejoux, J.-F., Micheneau, A., Bitar, A. A., & Trépos, R. (2023). Forecasting sunflower grain yield using remote sensing data and statistical models. *European Journal of Agronomy*, *142*, 126677. https://doi.org/10.1016/j.eja.2022.126677

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

Dixon, D. J., Callow, J. N., Duncan, J. M. A., Setterfield, S. A., & Pauli, N. (2021). Satellite prediction of forest flowering phenology. *Remote Sensing of Environment*, *255*, 112197. https://doi.org/10.1016/j.rse.2020.112197

Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop Yield Assessment from Remote Sensing. *Photogrammetric Engineering & Remote Sensing*, *69*(6), 665–674. https://doi.org/10.14358/pers.69.6.665

ESRI. (2023, August 8). *Meta's Segment Anything Model (SAM) for segmenting objects in any imagery*. Meta's Segment Anything Model (SAM) for Segmenting Objects in Any Imagery. https://www.arcgis.com/home/item.html?id=9b67b441f29f4ce6810979f5f0667ebe

Ferencz, Cs., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai†, Gy., Timár, G., Molnár, G., Pásztor, SZ., Steinbach, P., Székely, B., Ferencz, O. E., & Ferencz-Árkos, I. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, *25*(20), 4113–4149. https://doi.org/10.1080/01431160410001698870

Fernandez-Mensaque, P. C., Minero, F. J. G., Morales, J., & Tomas, C. (1998). Forecasting olive (Olea europaea) crop production by monitoring airborne pollen. *Aerobiologia*, *14*(2–3), 185–190. https://doi.org/10.1007/bf02694204

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., García-Mozo, H., Vázquez, L., Ruiz, L., Guardia, C. D., & Domínguez-Vilches, E. (2008). Modeling Olive Crop Yield in Andalusia, Spain. *Agronomy Journal*, *100*(1), 98–104. https://doi.org/10.2134/agronj2006.0345

Galán, C., Vázquez, L., García-Mozo, H., & Domínguez, E. (2004). Forecasting olive (Olea europaea) crop yield based on pollen emission. *Field Crops Research*, *86*(1), 43–51. https://doi.org/10.1016/s0378-4290(03)00170-9

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/s0034-4257(96)00067-3

Gao, H., Zhang, X., Wang, X., & Zeng, Y. (2023). Phenology-Based Remote Sensing Assessment of Crop Water Productivity. *Water*, *15*(2), 329. https://doi.org/10.3390/w15020329

García-Mozo, H., Perez-Badía, R., & Galán, C. (2008). Aerobiological and meteorological factors' influence on olive (Olea europaea L.) crop yield in Castilla-La Mancha (Central Spain). *Aerobiologia*, *24*(1), 13–18. https://doi.org/10.1007/s10453-007-9075-x

Gayke, M. P. D., & Rokade, M. S. (2021). Crop yield prediction using weather data and NDVI time series. *International Research Journal of Modernization in Engineering Technology and Science*, *03*(09).

Gilabert, M. A., González-Piqueras, J., García-Haro, F. J., & Meliá, J. (2002). A generalized soil-adjusted vegetation index. *Remote Sensing of Environment*, *82*(2–3), 303–310. https://doi.org/10.1016/s0034-4257(02)00048-2

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, *33*(11). https://doi.org/10.1029/2006gl026457

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, *30*(5). https://doi.org/10.1029/2002gl016450

Gitelson, A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, *22*(3), 247–252. https://doi.org/10.1016/1011-1344(93)06963-4

Gonzales, D., Ibarra, N. H. de, & Anderson, K. (2022). Remote Sensing of Floral Resources for Pollinators – New Horizons From Satellites to Drones. *Frontiers in Ecology and Evolution*, *10*, 869751. https://doi.org/10.3389/fevo.2022.869751

Gracia-Romero, A., Rufo, R., Gómez-Candón, D., Soriano, J. M., Bellvert, J., Yannam, V. R. R., Gulino, D., & Lopes, M. S. (2023). Improving in-season wheat yield prediction using remote sensing and additional agronomic traits as predictors. *Frontiers in Plant Science*, *14*, 1063983. https://doi.org/10.3389/fpls.2023.1063983

Han, J., Zhang, Z., & Cao, J. (2020). Developing a New Method to Identify Flowering Dynamics of Rapeseed Using Landsat 8 and Sentinel-1/2. *Remote Sensing*, *13*(1), 105. https://doi.org/10.3390/rs13010105

Hatfield, J. L. (1983). Remote sensing estimators of potential and actual crop yield. *Remote Sensing of Environment*, *13*(4), 301–311. https://doi.org/10.1016/0034-4257(83)90032-9

Helman, D. (2018). Land surface phenology: What do we really 'see' from space? *Science of The Total Environment*, *618*, 665–673. https://doi.org/10.1016/j.scitotenv.2017.07.237

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-x

Huete, A. R. (2012). Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass*, *6*(9), 513–532. https://doi.org/10.1111/j.1749-8198.2012.00507.x

Idso, S. B., Jackson, R. D., & Reginato, R. J. (1977a). Remote sensing for agricultural water management and crop yield prediction. *Agricultural Water Management*, *1*(4), 299–310. https://doi.org/10.1016/0378-3774(77)90021-x

Idso, S. B., Jackson, R. D., & Reginato, R. J. (1977b). Remote-Sensing of Crop Yields. *Science*, *196*(4285), 19–25. https://doi.org/10.1126/science.196.4285.19

Ilyas, Q. M., Ahmad, M., & Mehmood, A. (2023). Automated Estimation of Crop Yield Using Artificial Intelligence and Remote Sensing Technologies. *Bioengineering*, *10*(2), 125. https://doi.org/10.3390/bioengineering10020125

India, D. of C., GJUST, Hisar, Singh, K., Sunila, India, D. of C., GJUST, Hisar, & Kumar, S. (2020). Crop Yield Prediction Techniques using Remote Sensing Data. *International Journal of Engineering and Advanced Technology*, *9*(3), 3683–3689. https://doi.org/10.35940/ijeat.c6217.029320

International Olive Council. (2014). https://www.internationaloliveoil.org/#

Jaafar, H. H., & Ahmad, F. A. (2015). Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. *International Journal of Remote Sensing*, *36*(18), 4570–4589. https://doi.org/10.1080/01431161.2015.1084434

Jacobsen, A., Broge, N. H., & Hansen, B. U. (1995). *Monitoring wheat fields and grasslands using spectral reflectance data*. *26*.

Jeong, S., Yeom, J., Ko, J., Chung, D., & Lee, S.-G. (2023). *Application cases of remote sensing-integrated crop model to simulate and predict crop yield with satellite images*. https://doi.org/10.5194/egusphere-egu23-10586

Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors*, *21*(4), 1406. https://doi.org/10.3390/s21041406

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Jiménez-Jiménez, S. I., Marcial-Pablo, M. de J., Ojeda-Bustamante, W., Sifuentes-Ibarra, E., Inzunza-Ibarra, M. A., & Sánchez-Cohen, I. (2022). VICAL: Global Calculator to Estimate Vegetation Indices for Agricultural Areas with Landsat and Sentinel-2 Data. *Agronomy*, *12*(7), 1518. https://doi.org/10.3390/agronomy12071518

Jurado, J. M., Ortega, L., Cubillas, J. J., & Feito, F. R. (2020). Multispectral Mapping on 3D Models and Multi-Temporal Monitoring for Individual Characterization of Olive Trees. *Remote Sensing*, *12*(7), 1106. https://doi.org/10.3390/rs12071106

Karimli, N., & Selbesoğlu, M. O. (2023). Remote Sensing-Based Yield Estimation of Winter Wheat Using Vegetation and Soil Indices in Jalilabad, Azerbaijan. *ISPRS International Journal of Geo-Information*, *12*(3), 124. https://doi.org/10.3390/ijgi12030124

Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(2), 261–270. https://doi.org/10.1109/36.134076

Kefi, M., Pham, T. D., Kashiwagi, K., & Yoshino, K. (2016). Identification of irrigated olive growing farms using remote sensing techniques. *Euro-Mediterranean Journal for Environmental Integration*, *1*(1), 3. https://doi.org/10.1007/s41207-016-0004-7

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., & Lichtenberger, J. (2018a). Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology*, *260*, 300–320. https://doi.org/10.1016/j.agrformet.2018.06.009

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., & Lichtenberger, J. (2018b). Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation

indices. *Agricultural and Forest Meteorology*, *260*, 300–320. https://doi.org/10.1016/j.agrformet.2018.06.009

Khan, A., Khan, U., Waleed, M., Khan, A., Kamal, T., Marwat, S. N. K., Maqsood, M., & Aadil, F. (2018). Remote Sensing: An Automated Methodology for Olive Tree Detection and Counting in Satellite Images. *IEEE Access*, *6*, 77816–77828. https://doi.org/10.1109/access.2018.2884199

Khanal, S., KC, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities. *Remote Sensing*, *12*(22), 3783. https://doi.org/10.3390/rs12223783

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, *1*(3), 155–159. https://doi.org/10.1016/s0034-4257(70)80021-9

Kurucu, Y., Esetlílí, T., Erden, H., Öztürk, G., Güven, A. İ., & Çamaşircioğlu, E. (2015). Digitalization of Olive Trees by Using Remote Sensing Techniques. *2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 121–124. https://doi.org/10.1109/agro-geoinformatics.2015.7248143

Landsat 8 (L8) Data Users Handbook. (2019). Department of the Interior U.S. Geological Survey.

Li, A., Liang, S., Wang, A., & Qin, J. (2007). Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques. *Photogrammetric Engineering & Remote Sensing*, *73*(10), 1149–1157. https://doi.org/10.14358/pers.73.10.1149

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008

Madueño, J. M. M., Surián, M. R., & Páez, M. D. Z. (2005, September 21). Diagnóstico de la Desertificación en Andalucía. *XI Congreso Nacional de Teledetección*. XI Congreso Nacional de Teledetección.

McCree, K. J. (1972). The action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agricultural Meteorology*, *9*, 191–216. https://doi.org/10.1016/0002-1571(71)90022-7

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D., & Verdin, J. P. (2017). A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data*, *4*(1), 170012. https://doi.org/10.1038/sdata.2017.12

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

Minero, F. J. G., Candau, P., Morales, J., & Tomas, C. (1998). Forecasting olive crop production based on ten consecutive years of monitoring airborne pollen in Andalusia (southern Spain). *Agriculture, Ecosystems & Environment*, *69*(3), 201–215. https://doi.org/10.1016/s0167-8809(98)00105-4

Modis Web site. (2023). https://modis.gsfc.nasa.gov/about/

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset

for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. *Remote Sensing*, *14*(9), 1990. https://doi.org/10.3390/rs14091990

Navrozidis, Ioannis, Alexandridis, T. K., Moshou, D., Pantazi, X. E., Tamouridou, A. A., Kozhukh, D., Castef, F., Lagopodi, A., Zartaloudis, Z., Mourelatos, S., & Santos, F. J. N. de. (2019). Olive Trees Stress Detection Using Sentinel-2 Images. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, *00*, 7220–7223. https://doi.org/10.1109/igarss.2019.8898076

Orlandi, F., Msallem, M., Bonofiglio, T., Dhiab, A. B., Sgromo, C., Romano, B., & Fornaciari, M. (2010). Relationship between olive flowering and latitude in two Mediterranean countries (Italy and Tunisia). *Theoretical and Applied Climatology*, *102*(3–4), 265–273. https://doi.org/10.1007/s00704-009-0239-1

Oteros, J., Orlandi, F., García-Mozo, H., Aguilera, F., Dhiab, A. B., Bonofiglio, T., Abichou, M., Ruiz-Valenzuela, L., Trigo, M. M. del, Guardia, C. D. de la, Domínguez-Vilches, E., Msallem, M., Fornaciari, M., & Galán, C. (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

Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, *31*(2), 221–230.

Piferrer, S. J., Losa, S. M., & Pérez, J. J. L. (2017). *Calendario de siembra, recolección y comercialización 2014-2016*. Ministerio de Agricultura, Pesca y Alimentación. https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/calendarios-siembras-recoleccion/

Post, W. M., Emanuel, W. R., Zinke, P. J., & Stangenberger, A. G. (1982). Soil carbon pools and world life zones. *Nature*, *298*(5870), 156–159. https://doi.org/10.1038/298156a0

Rallo, L. (1994). Evaluación agronómica y obtención de nuevas variedades de olivo. *Fruticultura Profesional*, *62*, 17–27.

Ribeiro, H., Abreu, I., & Cunha, M. (2017). Olive crop-yield forecasting based on airborne pollen in a region where the olive groves acreage and crop system changed drastically. *Aerobiologia*, *33*(4), 473–480. https://doi.org/10.1007/s10453-017-9483-5

Ribeiro, H., Cunha, M., & Abreu, I. (2009). A bioclimatic model for forecasting olive yield. *The Journal of Agricultural Science*, *147*(6), 647–656. https://doi.org/10.1017/s0021859609990256

Ronchetti, G., Manfron, G., Weissteiner, C. J., Seguini, L., Scacchiafichi, L. N., Panarello, L., & Baruth, B. (2023). Remote sensing crop group-specific indicators to support regional yield forecasting in Europe. *Computers and Electronics in Agriculture*, *205*, 107633. https://doi.org/10.1016/j.compag.2023.107633

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring Vegetation Systems in the Great Plains with ERTS (Earth Resources Technology Satellite). *Proceedings of 3rd Earth Resources Technology Satellite Symposium*, *Sp-351*, 309–317.

Roux, G. L. (2023, November 3). *Italy's olive growers lament poor harvests from extreme weather*. Phys.Org. https://phys.org/news/2023-11-italy-olive-growers-lament-poor.html

Sala, O. E., Chapin, F. S., III, Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M., Poff, N. L., Sykes, M. T., Walker, B. H., Walker, M., & Wall, D. H. (2000). Global Biodiversity Scenarios for the Year 2100. *Science*, *287*(5459), 1770–1774. https://doi.org/10.1126/science.287.5459.1770

Sanz-Cortes, F., Martinez-Calvo, J., Badenes, M. L., Bleiholder, H., Hack, H., Llacer, G., & Meier, U. (2002). Phenological growth stages of olive trees (Olea europaea). *Annals of Applied Biology*, *140*(2), 151–157. https://doi.org/10.1111/j.1744-7348.2002.tb00167.x

Sellers, P. J., Mintz, Y., Sud, Y. C., & Dalcher, A. (1986). A Simple Biosphere Model (SIB) for Use within General Circulation Models. *Journal of the Atmospheric Sciences*, *43*(6), 505–531. https://doi.org/10.1175/1520-0469(1986)043<0505:asbmfu>2.0.co;2

Sentinel-2 User Handbook. (2015). European Space Agency.

Solgi, S., Ahmadi, S. H., & Seidel, S. J. (2023). Remote sensing of canopy water status of the irrigated winter wheat fields and the paired anomaly analyses on the spectral vegetation indices and grain yields. *Agricultural Water Management*, *280*, 108226. https://doi.org/10.1016/j.agwat.2023.108226

Sousa, D., & Small, C. (2023). Which Vegetation Index? Benchmarking Multispectral Metrics to Hyperspectral Mixture Models in Diverse Cropland. *Remote Sensing*, *15*(4), 971. https://doi.org/10.3390/rs15040971

Steven, M. D. (1982). Optical Remote Sensing for Prediction of Crop Yields. *Light Measurement '81*, 137–142. https://doi.org/10.1117/12.959727

Taloor, A. K., Manhas, D. S., & Kothyari, G. C. (2021). Retrieval of land surface temperature, normalized difference moisture index, normalized difference water index of the Ravi basin using Landsat data. *Applied Computing and Geosciences*, *9*, 100051. https://doi.org/10.1016/j.acags.2020.100051

Torres, L. G., Peña-Barragán, J. M., López-Granados, F., Jurado-Expósito, M., & Fernández-Escobar, R. (2008). Automatic assessment of agro-environmental indicators from remotely sensed images of tree orchards and its evaluation using olive plantations. *Computers and Electronics in Agriculture*, *61*(2), 179–191. https://doi.org/10.1016/j.compag.2007.11.004

Tucker, C. J., Vanpraet, C. L., Sharman, M. J., & Ittersum, G. V. (1985). Satellite remote sensing of total herbaceous biomass production in the senegalese sahel: 1980–1984. *Remote Sensing of Environment*, *17*(3), 233–249. https://doi.org/10.1016/0034-4257(85)90097-5

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009jcli2909.1

Viola, F., Caracciolo, D., Pumo, D., & Noto, L. V. (2013). Olive Yield and Future Climate Forcings. *Procedia Environmental Sciences*, *19*, 132–138. https://doi.org/10.1016/j.proenv.2013.06.015

Wall, L., Larocque, D., & Léger, P. (2008). The early explanatory power of NDVI in crop yield modelling. *International Journal of Remote Sensing*, *29*(8), 2211–2225. https://doi.org/10.1080/01431160701395252

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2015). Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2. *Nature*, *538*(7626), 499–501. https://doi.org/10.1038/nature19772

Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, *14*(6), 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zarco-Tejada, P. J., Miller, J. R., Morales, A., Berjón, A., & Agüera, J. (2004). Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sensing of Environment*, *90*(4), 463–476. https://doi.org/10.1016/j.rse.2004.01.017

Zhang, L., Qiao, N., Baig, M. H. A., Huang, C., Lv, X., Sun, X., & Zhang, Z. (2019). Monitoring vegetation dynamics using the universal normalized vegetation index (UNVI): An optimized vegetation index-VIUPD. *Remote Sensing Letters*, *10*(7), 629–638. https://doi.org/10.1080/2150704x.2019.1597298

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*(3), 471–475. https://doi.org/10.1016/s0034-4257(02)00135-9

Zhu, X., Guo, R., Liu, T., & Xu, K. (2021). Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data. *Remote Sensing*, *13*(10), 2016. https://doi.org/10.3390/rs13102016

Zohary, D., & Spiegel-Roy, P. (1975). Beginnings of Fruit Growing in the Old World. *Science*, *187*(4174), 319–327. https://doi.org/10.1126/science.187.4174.319

11 ANNEXES

11.1 Equations of the vegetation indices applied in this study

Equations of the vegetation indices applied in this MSc thesis	
Vegetation index	**Equation**
2-band Enhanced Vegetation Index (EVI2)	2.5 * (NIR - RED) / (NIR + 2.4 * RED + 1)
Advanced Vegetation Index (AVI)	(NIR * (1 - RED) * (NIR - RED)) ^ 1 / 3
Atmospherically Resistant Vegetation Index (ARVI)	(NIR - 2 * RED + BLUE) / (NIR + 2 * RED - BLUE)
Enhanced Vegetation Index (EVI)	(2.5 * (NIR - RED)) / (NIR + 6 * RED - 7.5 * BLUE + 1)
Green Chlorophyll Index (GCI)	NIR / (GREEN - 1)
Green Normalized Difference Vegetation Index (GNDVI)	(NIR - GREEN) / (NIR + GREEN)
Normalized Difference Moisture Index (NDMI)	(NIR - SWIR1) / (NIR + SWIR1)
Normalized Difference Vegetation Index (NDVI)	(NIR - RED) / (NIR + RED)
Optimized Soil Adjusted Vegetation Index (OSAVI)	1.16 * (NIR - RED) / (NIR + RED + 0.16)
Red-Edge Chlorophyll Vegetation Index (RECL)	(NIR / RED) – 1
Soil Adjusted Vegetation Index (SAVI)	(NIR - RED) / (NIR + RED + 0.5) * (1 + 0.5)
Structure Intensive Pigment Vegetation Index (SIPI)	(NIR - BLUE) / (NIR - RED)

Table 10 Equations used to calculate the vegetation indices

11.2 The BBCH-scale

Growth stage	BBCH Code	Description
0: Bud development	0	Foliar buds at the apex of shoots grown the previous crop-year are completely closed, sharp-pointed, stemless and ochre-coloured.
	1	Foliar buds start to swell and open, showing the new foliar primordia.
	3	Foliar buds lengthen and separate from the base.
	7	External small leaves open, not completely separated, remaining joined by apices.
	9	External small leaves opening further with their tips inter crossing.
1: Leaf development	11	First leaves completely separated. Grey-greenish coloured.
	15	The leaves are more separated without reaching their final size. First leaves turn greenish on the upperside.
	19	Leaves get the typical variety size and shape.
3: Shoot development	31	Shoots reach 10% of final size.
	33	Shoots reach 30% of final size.
	37	Shoots reach 70% of final size.
5: Inflorescence emergence.	50	Inflorescence buds in leaf axiles are completely closed. They are sharp-pointed, stemless and ochre-coloured.
	51	Inflorescence buds start to swell on its stem.
	52	Inflorescence buds open. Flower cluster development starts.
	54	Flower cluster growing
	55	Flower cluster totally expanded. Floral buds start to open.
	57	The corolla, green-coloured, is longer than calyx.
	59	The corolla changes from green to white colour.
6: Flowering	60	First flowers open.
	61	Beginning of flowering: 10% of flowers open.
	65	Full flowering: at least 50% of flowers open.
	67	First petals falling.
	68	Majority of petals fallen or faded.
	69	End of flowering, fruit set, non-fertilized ovaries fallen.
7: Fruit development	71	Fruit size about 10% of final size.
	75	Fruit size about 50% of final size. Stone starts to lignificate (it shows cutting resistance).
	79	Fruit size about 90% of final size. Fruit suitable for picking green olives.
8: Maturity of fruit	80	Fruit deep green colour becomes light green, yellowish.
	81	Beginning of fruit colouring.
	85	Increasing of specific fruit colouring.
	89	Harvest maturity: fruits get the typical variety colour, remaining turgid, suitable for oil extraction.
9: Senescence	92	Overripe: fruits lose turgidity and start to fall.

Table 11 BBCH scale with the description for the olive trees phenology (Zadoks et al., 1974).

11.3 Background about the remote sensing platforms used in this study

In this study, the Sentinel-2, MODIS and Landsat-8 have been used as source of reflectance from the areas of interest. Below, there is a brief description of these three remote sensing platforms.

11.3.1 Sentinel-2

SENTINEL-2 is a European wide-swath (290 km), high-resolution, multi-spectral imaging mission (**Figure 35**). The SENTINEL-2 mission includes two identical two identical satellites, SENTINEL-2A and SENTINEL-2B.

Figure 35 Sentinel-2 satellite. No changes were made in this image licensed under the Creative Commons Attribution-Share Alike 2.0 France license. https://creativecommons.org/

These twin polar-orbiting satellites fly in the same orbit, phased at 180° to each other in a sun-synchronous orbit at a mean altitude of 786 km. This layout is to ensure the angle of sunlight upon the Earth's surface is consistently maintained.

This orbital configuration allows a high revisit frequency of 10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes (*Sentinel-2 User Handbook*, 2015).

The MultiSpectral Instrument (MSI) uses a push-broom concept. A push-broom sensor, also known as an along-track scanner, works by collecting rows of images across the orbital swath (in this case along 290 km) and employs the forward motion of the spacecraft along the path of the orbit to provide new rows for acquisition. The average period of observation over land and coastal areas is approximately 17 minutes and the maximum period of observation is 32 minutes (*Sentinel-2 User Handbook*, 2015).

The MSI works passively, by collecting sunlight reflected from the Earth. The light reflected up to the MSI instrument from the Earth and its atmosphere is collected by a three-mirror (M1, M2 and M3) telescope and is split via a beam-splitter and focused onto two separate focal plane assemblies within the instrument: one for the ten VNIR wavelengths and one for the three SWIR wavelengths.

The spectral separation of each band into individual wavelengths is accomplished by stripe filters mounted on top of the detectors (*Sentinel-2 User Handbook*, 2015). The optical design of the MSI telescope allows for a 290 km Field-Of-View (FOV).

The Sentinel-2 provides a wide range of images elaborated from Top-Of-Atmosphere radiances in sensor geometry to atmospherically corrected surface reflectance in cartographic geometry (Level-2C) derived from the associated Level-1C products which includes radiometric and geometric corrections including orthorectification and spatial registration on a global reference system with sub-pixel accuracy (*Sentinel-2 User Handbook*, 2015).Sentinel-2 products available for users are listed on **figure 36**.

Type	Code	Description
User Product	Level-1B	Top-Of-Atmosphere radiances in sensor geometry
	Level-1C	Top-of-atmosphere reflectances in cartographic geometry
	Level-2A	Atmospherically corrected Surface Reflectances in cartographic geometry

Figure 36 Sentinel-2 products.

SENTINEL-2 will carry an optical instrument, the Multi-Spectral Instrument (MSI), that samples 13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m of spatial resolution (Sentinel-2 User Handbook, 2015). The 1B product includes the refined geometrical which is used to generate the Level-1C product (*Sentinel-2 User Handbook*, 2015).

11.3.2 Landsat-7 and 8

The Landsat programme started with the launch of Landsat 1 in July 1972. Since them nine spacecrafts integrate the Landsat missions with the last one, Landsat 9 launched on September 2021. The Landsat satellites can be classified into three groups, based on their sensor and platform characteristics (Chander et al., 2009). The first group consists of Landsat-1, Landsat-2, and Landsat-3, each of which carries a MultiSpectral Scanner (MSS) sensor and a Return Beam Vidicon (RBV) camera.

The second group, comprising Landsat-4 and Landsat-5, carries the MSS and the Thematic Mapper I. The third group, Landsat 6, and Landsat-7 carry the Enhanced Thematic Mapper (ETM) and the Enhanced Thematic Mapper Plus (ETM+) respectively.

Landsat-7, and previous Landsat satellites, had detectors that swept back-and-forth, like a whisk broom, as they collected data across the image swath. After Landsat-7, the instruments on the last two Landsat missions, Landsat-8 and 9, have linear detector arrays, also called "push-broom" detectors, which collect data across the entire image swath at once. Landsat-8 and 9 carries two sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) being the instruments onboard Landsat 9 (the Operational Land Imager 2 (OLI-2) and the Thermal Infrared Sensor 2 (TIRS-2) an improved replicas of those currently collecting data onboard Landsat-8 (*Landsat 8 (L8) Data Users Handbook*, 2019).

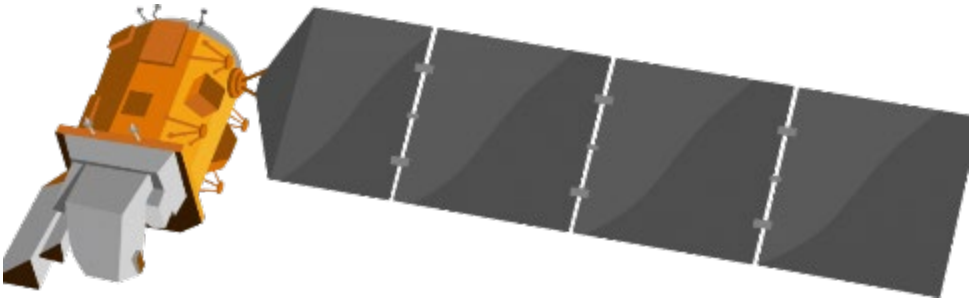


*Figure 37 Landsat 8. As this caption contain the copyright reference can be used **private, scholarly**, and **non-commercial** use for informational purposes. Copyright by the Freie Universität Berlin. Source: https://blogs.fu-berlin.de/reseda/landsat-8*

This study has used the data from the Landsat missions 7 and 8, which were launched in April 1999 and February 2013 respectively to an altitude of 705 km (**Figure 37**). These satellites are orbiting the Earth in a sun-synchronous, near-polar orbit that allows to complete one Earth orbit every 99 minutes (*Landsat 8 (L8) Data Users Handbook*, 2019) or more or less 14.5 orbits per day with a repeat coverage of 16 days. Landsat-7 carry the Enhanced Thematic Mapper Plus (ETM+) with a panchromatic band with 15m spatial resolution.

On the other hand, the Landsat-8 carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments. The OLI measures in the visible, near infrared, and

BAND	SPECTRAL	WAVELEN. [µm]	GEOM. [m]	SENSOR
1	aerosols	0.435 – 0.451	30	OLI
2	blue	0.452 – 0.512	30	OLI
3	green	0.533 – 0.590	30	OLI
4	red	0.636 – 0.673	30	OLI
5	NIR	0.851 – 0.879	30	OLI
6	SWIR-1	1.566 – 1.651	30	OLI
7	SWIR-2	2.107 – 2.294	30	OLI
8	pan	0.503 – 0.676	15	OLI
9	cirrus	1.363 – 1.384	30	OLI
10	TIR-1	10.600 – 11.190	100	TIRS
11	TIR-2	11.500 – 12.510	100	TIRS

*Figure 38 Landsat 8 Bands Designations. As this caption contain the copyright reference can be used **private, scholarly**, and **non-commercial** use for informational purposes. Copyright by the Freie Universität Berlin. Source: https://blogs.fu-berlin.de/reseda/landsat-8*

shortwave infrared portions (VNIR, NIR, and SWIR) of the spectrum. The TIRS measures land surface temperature in two thermal bands.

As the Sentinel-2, the Landsat-8 sensors use a push-broom concept. The images have 11 spectral bands (**Figure 38**).

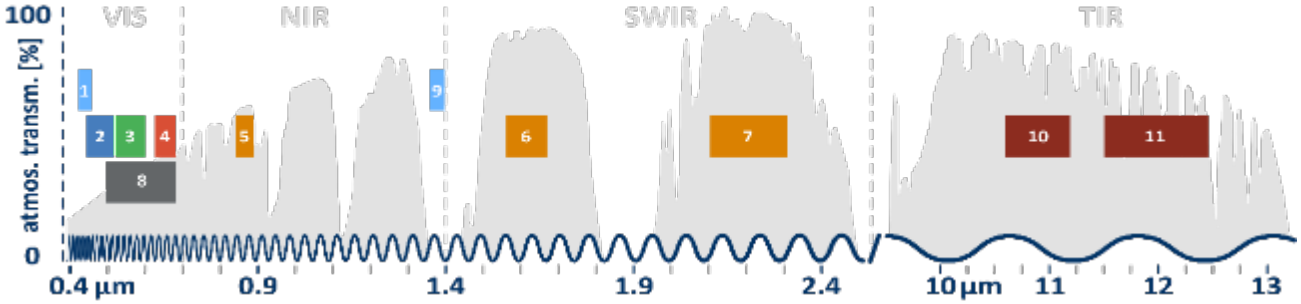


*Figure 39 Landsat 8 spectral bands. As this caption contain the copyright reference can be used **private, scholarly**, and **non-commercial** use for informational purposes. Copyright by the Freie Universität Berlin. Source: https://blogs.fu-berlin.de/reseda/landsat-8*

The OLI multispectral bands 1-7 and 9 with 30 meters resolution, the OLI panchromatic band 8 with 15 meters resolution and the TIRS bands 10-11 collected at 100 meters but resampled to 30 meters to match OLI bands (**Figure 38** and **Figure 39**).

Landsat-8 acquires over 500 scenes each day and offers two datasets with distinct levels of preprocessing being the datasets known as Level-1 without atmospheric correction and second one with surface reflectance and atmospheric correction named as Level-2.

11.3.3 Terra satellite and MODIS instrumentation

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a satellite-based sensor used for earth and climate measurements. There are two MODIS sensors in Earth orbit: one on board the Terra satellite, launched by NASA in December 1999; and one on board the Aqua satellite, launched in May 2002.

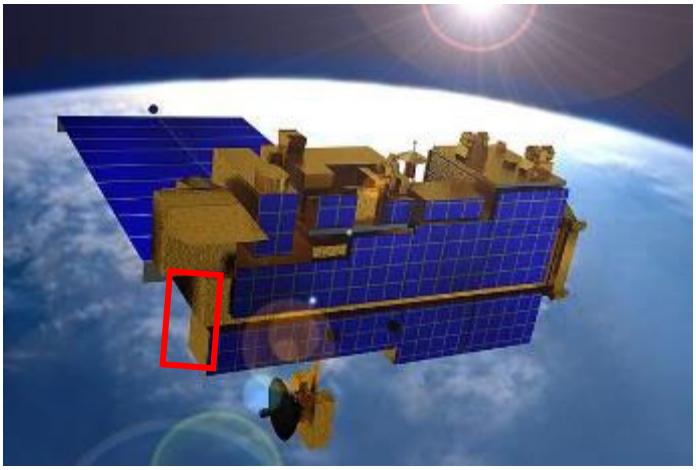


Figure 40 Terra spacecraft (NASA). See highlighted in the red box the position of the MODIS instrument. This image is public domain in the USA because it was solely created by NASA. NASA copyright policy states that "NASA material is not protected by copyright unless noted"

Terra was launched to an orbit at 705 km in a circular sun-synchronous polar orbit that takes it from north to south (on the daylight side of the Earth) every 99 minutes which means that the same region is acquired several times during daylight. The Terra satellite carries five instruments that observe Earth's atmosphere, ocean, land,

snow and ice, and energy budget (see highlighted in the red box in **figure 40**). The orbits of the Terra and Aqua satellites are designed in such a way that both satellites complete the Earth surface around 1 or 2 days.

MODIS (MODerate resolution Imaging Spectroradiometer) is a whiskbroom multi-spectral sensor (across track-scanner) that provides high radiometric sensitivity (12 bit) in 36 spectral bands ranging in wavelength from 0.4 µm to 14.4 µm with a swath of 2330 km (across track). Two bands are imaged at a nominal resolution of 250 m at nadir, with five bands at 500 m, and the remaining 29 bands at 1 km. a 2,330-km swath and provides global coverage every one to two days (*Modis Web Site*, 2023).

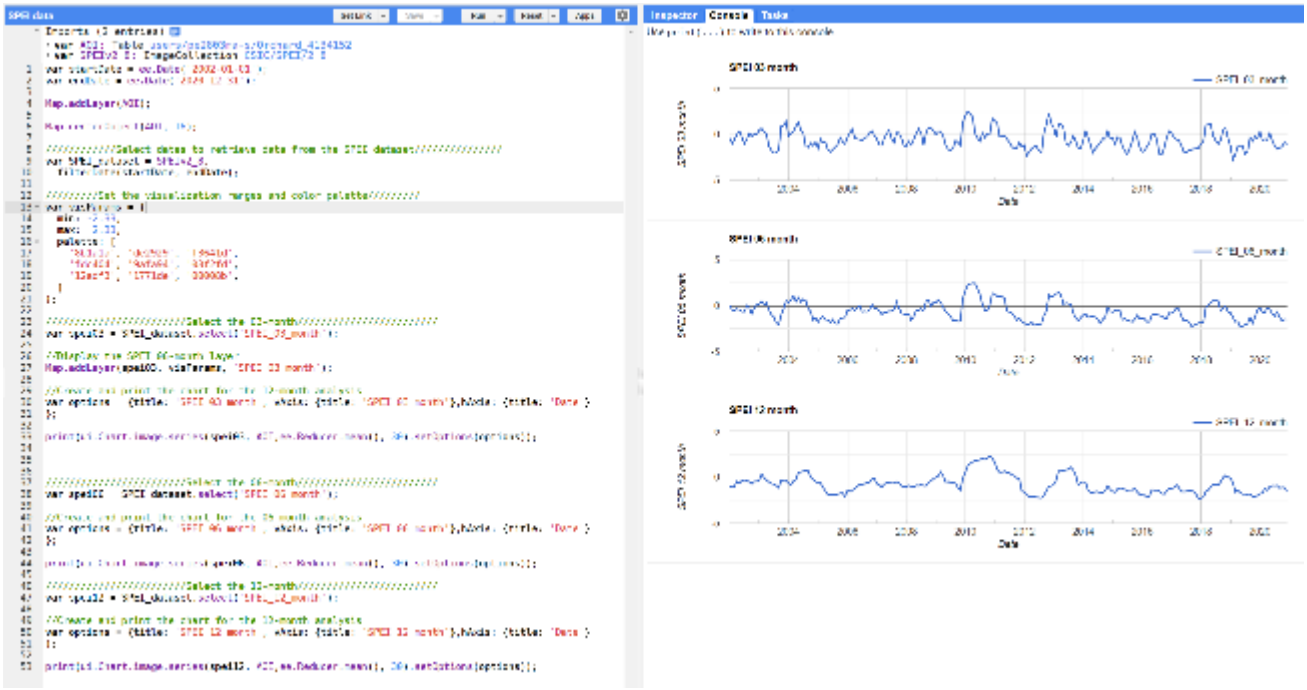
11.4 Google Earth Engine Scripts

11.4.1 Script to extract reflectance from MODIS, Landsat-7, Landsat-8 and Sentinel-2

Link URL to test and use the code below:
https://code.earthengine.google.com/69e5c0eab999fa023e68345278e9bfc6

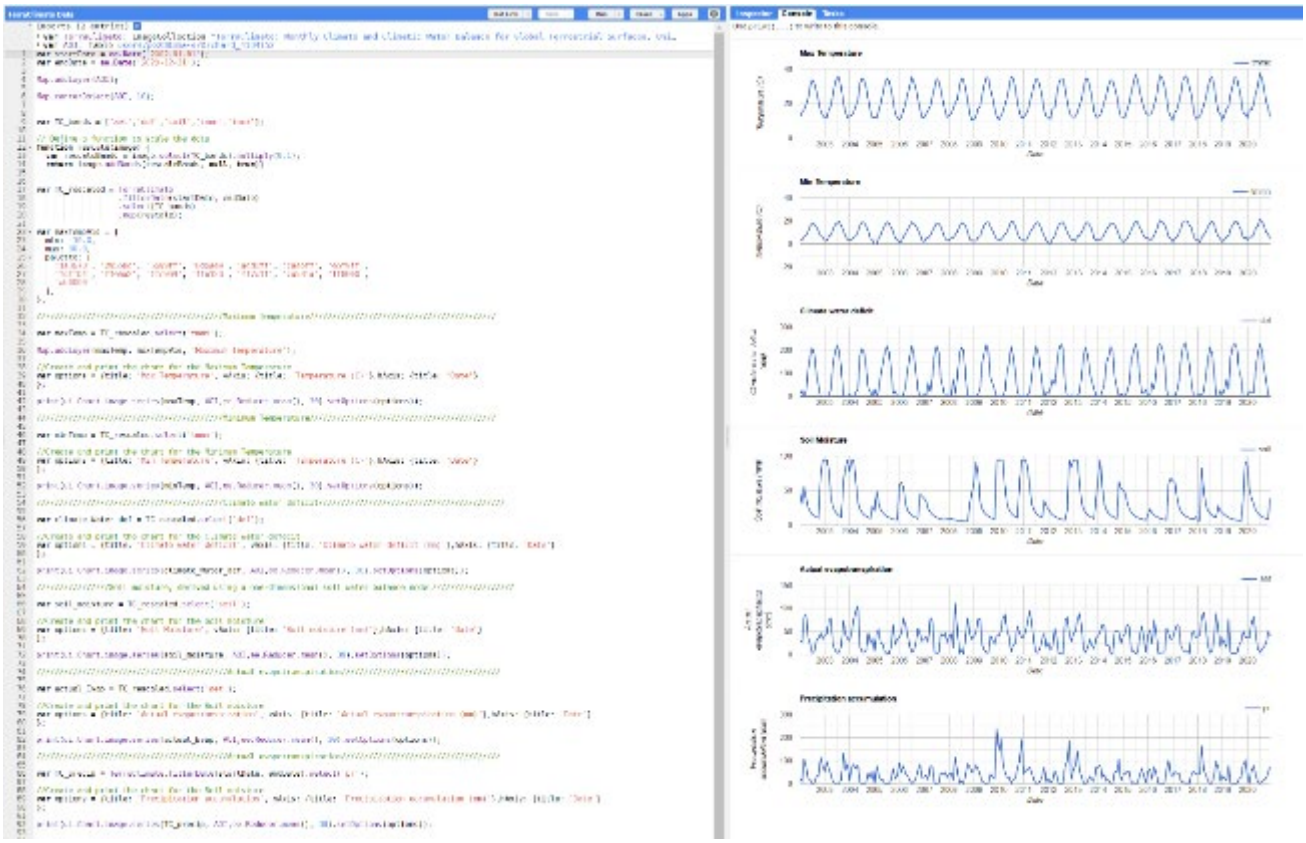
11.4.2 Script of the SPEI

Link URL to test and use the code below:
https://code.earthengine.google.com/df739e4c0b4c94b65cbc0689b1551c5e



11.4.3 Script to extract data from TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho

Link URL to test and use the code below:
https://code.earthengine.google.com/423687b2a6e71c2f18e672070cc8f944

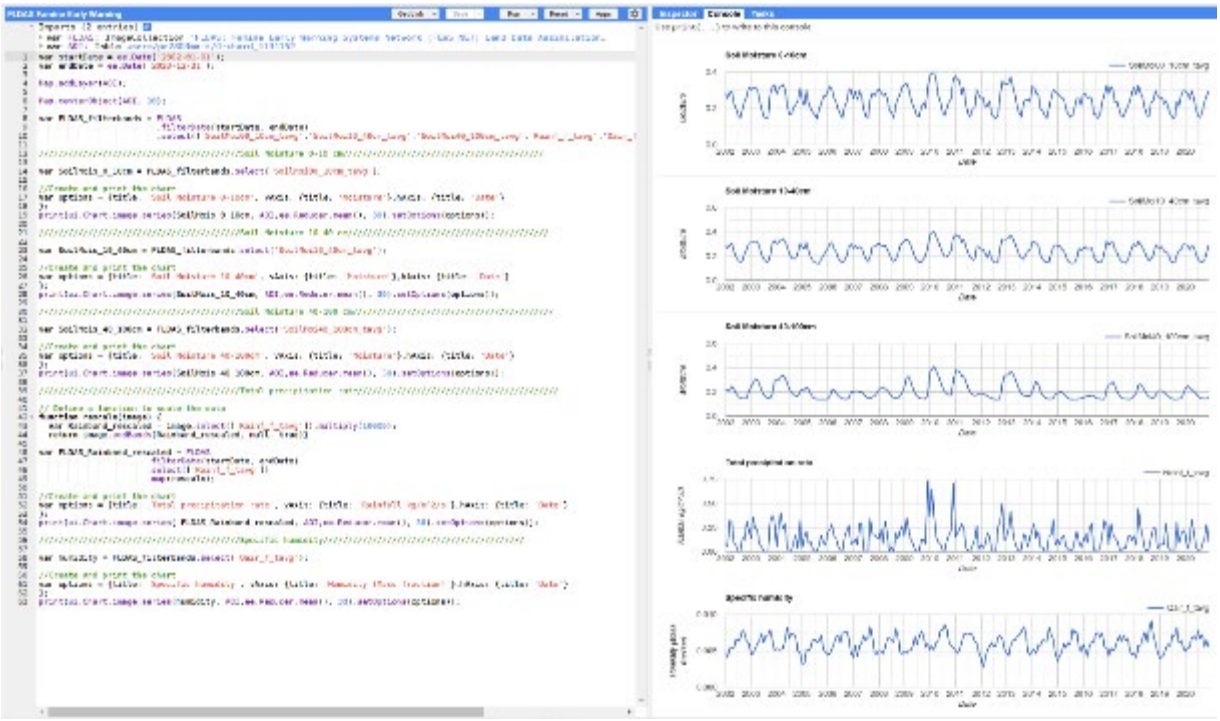


11.4.4 Script to extract data from MOD15A2H.061: Terra Leaf Area Index/FPAR 8-Day Global 500m

Link URL to test and use the code below:
https://code.earthengine.google.com/22c3b63a762ad2cf8973c8e513a2a34f

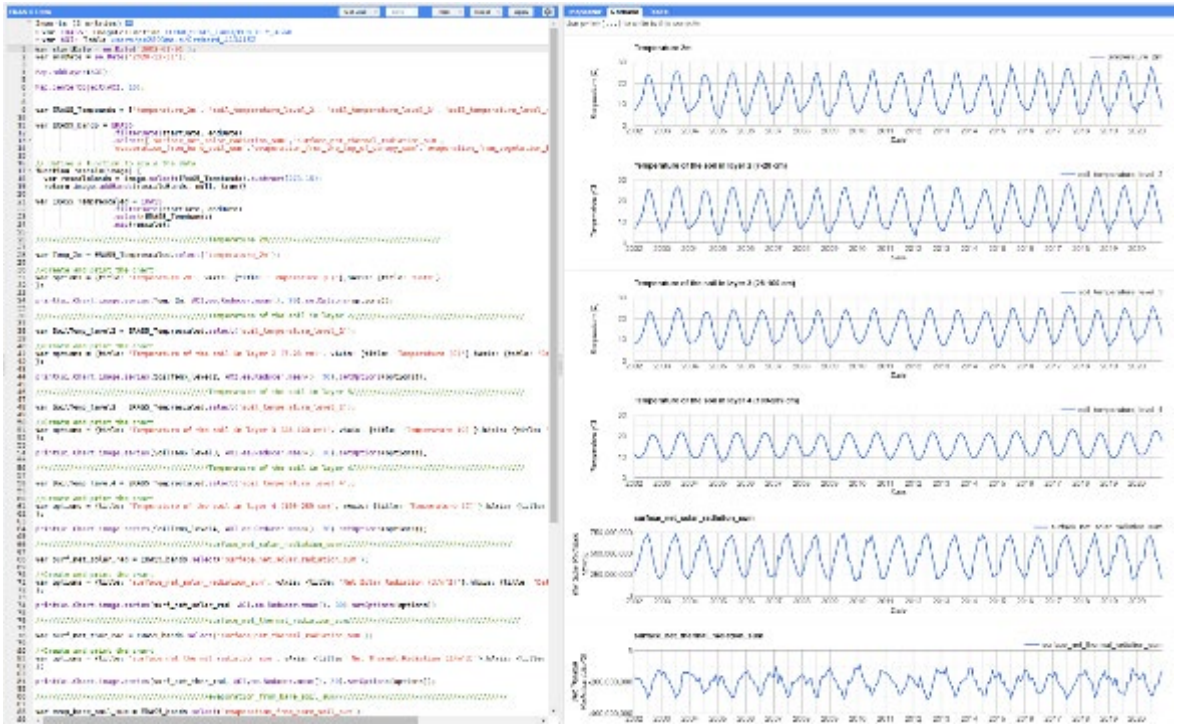
11.4.5 Script to extract data from FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System

Link URL to test and use the code below:
https://code.earthengine.google.com/7a1fb0f569bd1ee31b686c414809a781



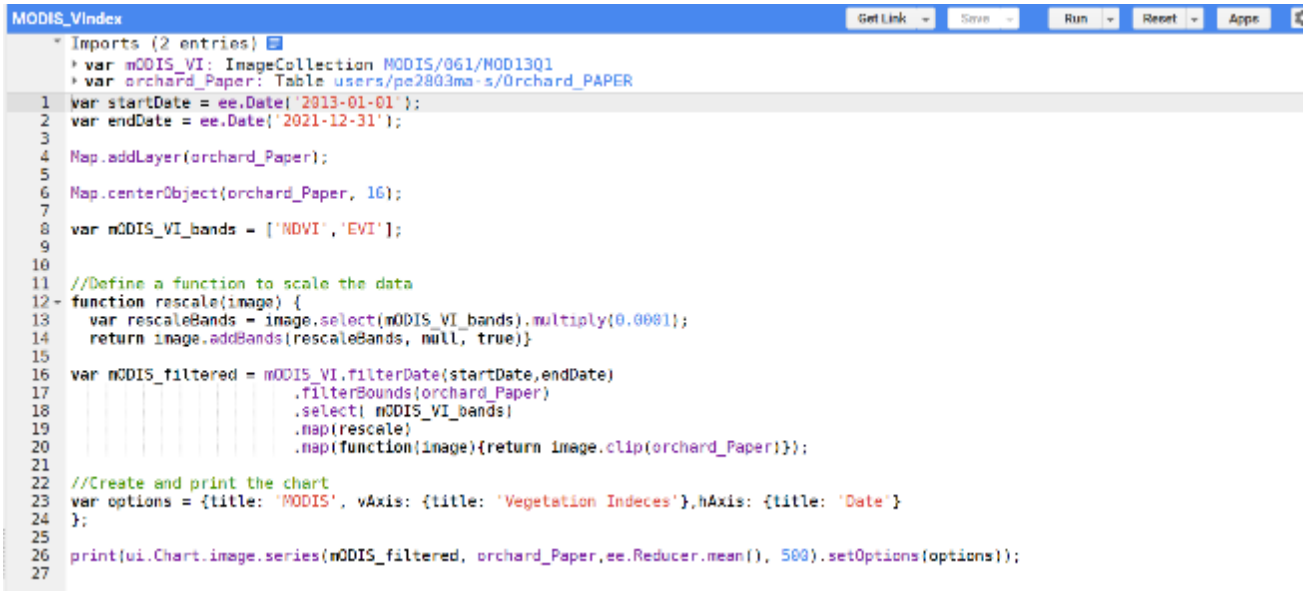
11.4.6 Script to extract data from ERA5-Land Monthly Aggregated - ECMWF Climate Reanalysis

Link URL to test and use the code below:
https://code.earthengine.google.com/f0a762b9c2cd0700c646fa10a8e29104



11.4.7 Script to extract data from MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m

Link URL to test and use the code below:
https://code.earthengine.google.com/d2b0209dfe1f28e5ea871fc7a6e61d92



11.5 Metadata of datasets used in this study

11.5.1 Temperature dataset

The metadata and specific references and links are as follow:

- The ID for the average monthly temperature is: 49cf3b82-89eb-46a9-82a1-669620695886
- The metadata reference is: https://portalrediam.cica.es/geonetwork/srv/api/records/49cf3b82-89eb-46a9-82a1-669620695886/formatters/xml
- For downloading the average monthly temperature, the two-step links are below. These links provide access to the average monthly temperature, in raster format, calculated from the data of the meteorological stations of the State Meteorological Agency, from 1951 onwards.
- The link to the website to get the download link is: https://portalrediam.cica.es/geonetwork/srv/spa/catalog.search#/metadata/49cf3b82-89eb-46a9-82a1-669620695886
- Next, the link to download the average monthly temperature dataset for the Andalucia region is: https://portalrediam.cica.es/descargas?path=%2F04_RECURSOS_NATURALES%2F03_CLIMA%2F02_CARACTERIZACION_CLIMATICA%2F02_TEMPERATURA%2F01_TEMPERATURA%2F03_MENSUAL%2FTEMP_MEDIA_MENSUAL

11.5.2 Precipitation dataset

The metadata and specific references and links are as follow:

- The ID for the average monthly precipitation is: 58dd0df3-17e9-4aac-ab76-d8e249765cb4
- The metadata reference is: https://portalrediam.cica.es/geonetwork/srv/api/records/58dd0df3-17e9-4aac-ab76-d8e249765cb4/formatters/xml
- For downloading the average monthly precipitation, the two-step links are below. These links provide access to the average monthly temperature, in raster format, calculated from the data of the meteorological stations of the State Meteorological Agency, from 1991 onwards.
- The link to the website to get the download link is: https://portalrediam.cica.es/geonetwork/srv/spa/catalog.search#/metadata/58dd0df3-17e9-4aac-ab76-d8e249765cb4
- Next, the link to download the average monthly temperature dataset for the Andalucia region is: https://portalrediam.cica.es/descargas?path=%2F04_RECURSOS_NATURALES%2F03_CLIMA%2F02_CARACTERIZACION_CLIMATICA%2F03_PRECIPITACION%2F01_PRECIPITACION%2F03_MENSUAL%2FPRECIP_MENSUAL

11.5.3 Regional of Orthophotos of the study area

The metadata and specific references and links are as follow:

- The ID for the average monthly temperature is: a9c4fb61-4b5a-4762-9cf0-3ae01539cb9a
- The metadata reference is: https://portalrediam.cica.es/geonetwork/srv/api/records/a9c4fb61-4b5a-4762-9cf0-3ae01539cb9a/formatters/xml
- For downloading the regional orthophotos of Andalusia the two-step links are below. These links provide access to the dataset in *ecw (Enhanced Compression Wavelet) format.
- The link to the website to get the download link is:

https://portalrediam.cica.es/geonetwork/srv/spa/catalog.search#/metadata/a9c4fb61-4b5a-4762-9cf0-3ae01539cb9a

- Next, the link to download the regional orthophotos of Andalusia is: https://www.juntadeandalucia.es/medioambiente/portal/web/guest/landing-page-servicio-ogc/-/asset_publisher/1qlWV3LW9vV6/content/rediam.-wms-ortofoto-digital-de-andaluc-c3-ada-a-c3-b1o-2020/20151

11.5.4 Digital elevation model

The metadata and specific references and links are as follow:

- The ID for the average monthly temperature is: c1b18ea9-6bc5-4359-9db8-d3655293c501
- The metadata reference is: https://portalrediam.cica.es/geonetwork/srv/api/records/c1b18ea9-6bc5-4359-9db8-d3655293c501/formatters/xml
- For downloading this DEM, the two-step links are below. These links provide access to the dataset in ASCII XYZ and TIFF formats.
- The link to the website to get the download link is: https://portalrediam.cica.es/geonetwork/srv/spa/catalog.search#/metadata/c1b18ea9-6bc5-4359-9db8-d3655293c501
- Next, the link to download the regional orthophotos of Andalusia is: https://portalrediam.cica.es/descargas?path=%2F01_CARACTERIZACION_TERRITORIO%2F07_BASES_REF_ELEV%2F01_ELEVACIONES%2F01_PROYECTOS_REGIONALES%2F2013_AND_5m_V206_el

11.5.5 Land Use dataset

Metadata, specific references and website to download the land use dataset in shapefile format is:

https://www.juntadeandalucia.es/organismos/agriculturapescaaguaydesarrollorural/servicios/sigpac/visor/paginas/sigpac-descarga-informacion-geografica-shapes-provincias.html

It has the character of a public registry of administrative profile and contains updated information on the plots of land with defined agricultural uses or exploitations. This dataset is downloaded from the Regional Government of Andalusia but in another website portal, out of the REDIAM catalogue.

11.5.6 Olive yield Information in the study area

Metadata, specific references and website to download the land use dataset in shapefile format is:

https://www.mapa.gob.es/en/estadistica/temas/estadisticas-agrarias/agricultura/esyrce/

11.5.7 The Standardised Precipitation-Evapotranspiration Index

Metadata and SPEI database are available under the CC-BY 4.0 license. The link of the Earth Engine Data Catalog is:
https://developers.google.com/earth-engine/datasets/catalog/CSIC_SPEI_2_8

The code written to extract the data is shown in the annexes **11.4.2**. For this study the time scales extracted are the 2, 3, 6 and 12 months. The link to go directly to GEE and run the code is:
https://code.earthengine.google.com/df739e4c0b4c94b65cbc0689b1551c5e

11.5.8 TerraClimate

Metadata and the link of the Earth Engine Data Catalog is:
https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE#description

The code written to extract the data is shown in the annexes **11.4.3**. For this study the bands were rescales accordingly to the scale used by the Earth Engine Data Catalog. The link to go directly to GEE and run the code is:
https://code.earthengine.google.com/423687b2a6e71c2f18e672070cc8f944

11.5.9 MOD15A2H.061: Terra Leaf Area Index/FPAR 8-Day Global 500m

The code written to extract the data is shown in the annexes **11.4.4**. For this study the bands were rescales accordingly to the scale used by the Earth Engine Data Catalog.

The link to go directly to GEE and run the code is:
https://code.earthengine.google.com/22c3b63a762ad2cf8973c8e513a2a34f

11.5.10 MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m

Metadata and the link to go directly to GEE and run the code is:
https://code.earthengine.google.com/d2b0209dfe1f28e5ea871fc7a6e61d92

11.5.11 FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System

The code written to extract the data is shown in the annexes **11.4.5**. The link to go directly to GEE and run the code is:
https://code.earthengine.google.com/7a1fb0f569bd1ee31b686c414809a781

11.5.12 ERA5-Land Monthly Aggregated - ECMWF Climate Reanalysis

Metadata and the code written to extract the data is shown in the annexes **11.4.6**. The link to go directly to GEE and run the code is:
https://code.earthengine.google.com/f0a762b9c2cd0700c646fa10a8e29104

11.6 Anex 3 – Multilinear Regression Tables

11.6.1 All olive Jaen's orchards

Adjusted R Square from OVERALL lowest p-values with α <1% in All olive Jaen's orchards									
Multiple R	0.834362589	0.852664358	0.887665186	0.899092328	0.910593968	0.919703943	0.920648777	0.919802016	0.919938224
R Square	0.69616093	0.727036507	0.787949482	0.808367014	0.829181375	0.845855342	0.84759417	0.846035749	0.846286337
Adjusted R Square	**0.678288043**	**0.69291607**	**0.745539378**	**0.753614732**	**0.763481904**	**0.768783013**	**0.750608642**	**0.748058499**	**0.723315406**
Standard Error	438503.7155	428418.505	389986.8548	383748.8123	375986.1638	371748.7739	386082.7306	388051.6484	406660.6637
Observations	19	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8	9
X Predictors									
	MODIS NDMI - April	MODIS NDMI - April MODIS NDMI - March	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April SPEI 6 months - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April SPEI 6 months - April MODIS EVI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April SPEI 6 months - April NDVI from MODIS - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April SPEI 6 months - April NDVI from MODIS - April EVI from MODIS - April

Adjusted R Square from MODIS lowest p-values with α <1% All olive Jaen's orchards								
Multiple R	0.834362589	0.852664358	0.887665186	0.899092328	0.910593968	0.91325823	0.913312968	0.914253756
R Square	0.69616093	0.727036507	0.787949482	0.808367014	0.829181375	0.834040595	0.834140577	0.835859931
Adjusted R Square	**0.678288043**	**0.69291607**	**0.745539378**	**0.753614732**	**0.763481904**	**0.751060892**	**0.728593671**	**0.704547876**
Standard Error	438503.7155	428418.505	389986.8548	383748.8123	375986.1638	385732.5081	402763.0879	420226.3054
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
X Predictors								
	MODIS NDMI - April	MODIS NDMI - April MODIS NDMI - March	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April MODIS EVI - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April MODIS EVI - April NDVI from MODIS - April	MODIS NDMI - April MODIS NDMI - March MODIS NDMI - January MODIS SIPI - April MODIS ARVI - April MODIS EVI - April NDVI from MODIS - April EVI from MODIS - April

Adjusted R Square from LandSat 7 lowest p-values with 1% > α <5% All olive Jaen's orchards								
Multiple R	0.558017737	0.649492889	0.713776514	0.785935258	0.835824105	0.854901483	0.872141299	0.874556012
R Square	0.311383795	0.421841013	0.509476911	0.61769423	0.698601934	0.730856545	0.760630445	0.764848217
Adjusted R Square	**0.27087696**	**0.349571139**	**0.411372293**	**0.50846401**	**0.582679601**	**0.596284818**	**0.608304364**	**0.576726791**
Standard Error	660146.139	623504.4799	593143.843	542022.528	499429.8357	491221.331	483853.6738	502979.2475
Observations	19	19	19	19	19	19	19	19
STEPS	1	2	3	4	5	6	7	8
X Predictors								
	LandSat7 SIPI - June	LandSat7 SIPI - June LandSat7 NDVI - June	LandSat7 SIPI - June LandSat7 NDVI - June LandSat7 ReCI - Novembe	LandSat7 SIPI - June LandSat7 NDVI - June andSat7 ReCI - Novembe LandSat7 NDVI - Novembe	LandSat7 SIPI - June LandSat7 NDVI - June andSat7 ReCI - Novembe andSat7 NDVI - Novembe LandSat7 ARVI - April	LandSat7 SIPI - June LandSat7 NDVI - June andSat7 ReCI - Novembe andSat7 NDVI - Novembe LandSat7 ARVI - April LandSat7 EVI - Novembe	LandSat7 SIPI - June LandSat7 NDVI - June andSat7 ReCI - Novembe andSat7 NDVI - Novembe LandSat7 ARVI - April LandSat7 EVI - November LandSat7 ARVI - Novembe	LandSat7 SIPI - June LandSat7 NDVI - June LandSat7 ReCI - November LandSat7 NDVI - November LandSat7 ARVI - April LandSat7 EVI - November LandSat7 ARVI - November LandSat7 SIPI - April

Adjusted R Square from LandSat 8 lowest p-values with 1% > α <5% All olive Jaen's orchards						
Multiple R	0.914158089	0.934781099	0.95765208	0.965318767	0.965479544	0.986215875
R Square	0.835685011	0.873815702	0.917097506	0.931840322	0.93215075	0.972621752
Adjusted R Square	**0.80829918**	**0.823341983**	**0.854920635**	**0.84096075**	**0.762527625**	**0.808352265**
Standard Error	367293.902	352588.7145	319524.7878	334544.4976	408797.5461	367243.0438
Observations	8	8	8	8	8	8
STEPS	1	2	3	4	5	6
X Predictors						
	LandSat8 GNDVI - July	LandSat8 GNDVI - July LandSat8 AVI - June	LandSat8 GNDVI - July LandSat8 AVI - June LandSat8 EVI - December	LandSat8 GNDVI - July LandSat8 AVI - June LandSat8 EVI - December LandSat8 ARVI - July	LandSat8 GNDVI - July LandSat8 AVI - June LandSat8 EVI - December LandSat8 ARVI - July LandSat8 SAVI - July	LandSat8 GNDVI - July LandSat8 AVI - June andSat8 EVI - Decembe LandSat8 ARVI - July LandSat8 SAVI - July LandSat8 EVI2 - July

Brown coloured text are hyperlinks to easily browse throughout the document

Adjusted R Square from SENTINEL-2 with p-values 1% > α <5% All olive Jaen's orchards			
Multiple R	0.971736598	0.986260029	0.993905833
R Square	0.944272016	0.972708844	0.987848804
Adjusted R Square	**0.925696022**	**0.945417688**	**0.951395218**
Standard Error	172235.3893	147619.0206	139301.5208
Observations	5	5	5
STEPS	1	2	3
X Predictors			
	Sentinel2 ReCI - March	Sentinel2 ReCI - March Sentinel2 OSAVI - March	Sentinel2 ReCI - March Sentinel2 OSAVI - March Sentinel2 ARVI - August

Adjusted R Square from lowest p-values α <1% of CLIMATE DATASETS All olive Jaen's orchards					
Multiple R	0.705078516	0.778659185	0.822677775	0.836842525	0.84151316
R Square	0.497135714	0.606310127	0.676798721	0.700305412	0.708144399
Adjusted R Square	**0.467555462**	**0.557098893**	**0.612158465**	**0.614678386**	**0.595892245**
Standard Error	564126.972	514509.0988	481467.3426	479900.6759	491460.1047
Observations	19	19	19	19	19
STEPS	1	2	3	4	5
X Predictors					
	SPEI 6 months - April	SPEI 6 months - April Actual Evapotranspiration (TerraCual	SPEI 6 months - April Evapotranspiration (TerraClimate) - Janu Soil Moisture 0-10 cm (FLDAS) - August	SPEI 6 months - April Actual Evapotranspiration (TerraClimate) - January Soil Moisture 0-10 cm (FLDAS) - August SPEI 3 months - March	SPEI 6 months - April Actual Evapotranspiration (TerraClimate) - January Soil Moisture 0-10 cm (FLDAS) - August SPEI 3 months - March SPEI 12 months - May

11.6.2 Orchard 4134152

Adjusted R Square from OVERALL lowest p-values with α <1% Orchard 4134152				
Multiple R	0.627921786	0.894875123	0.969950217	0.999816838
R Square	0.394285769	0.800801486	0.940803424	0.99963371
Adjusted R Square	**0.242857212**	**0.668002477**	**0.85200856**	**0.998168548**
Standard Error	903.2719543	598.1320209	399.3447032	44.42499444
Observations	6	6	6	6
STEPS	1	2	3	4
X Predictors				
	LandSat7 NDMI - March	LandSat7 NDMI - March LandSat8 NDMI - February	LandSat7 NDMI - March LandSat8 NDMI - February Sentinel2 SIPI - July	LandSat7 NDMI - March LandSat8 NDMI - February LandSat7 NDVI - May Sentinel2 SIPI - July

Adjusted R Square from OVERALL lowest p-values of MODIS with 1%> α <5% Orchard 4134152					
Multiple R	0.617890974	0.74539046	0.803287845	0.824478275	0.872826263
R Square	0.381789256	0.555606938	0.645271362	0.679764426	0.761825685
Adjusted R Square	**0.330271694**	**0.474808199**	**0.538852771**	**0.537437504**	**0.612966737**
Standard Error	1138.399829	1008.101969	944.637748	946.0861902	865.4054781
Observations	14	14	14	14	14
STEPS	1	2	3	4	5
X Predictors					
	MODIS NDMI - May	MODIS NDMI - May MODIS GCI - July	MODIS NDMI - May MODIS GCI - July MODIS NDMI - April	MODIS NDMI - May MODIS GCI - July MODIS NDMI - April MODIS AVI - July	MODIS NDMI - May MODIS GCI - July MODIS NDMI - April MODIS AVI - July MODIS NDMI - June

Adjusted R Square from lowest p-values of Landsat 8 with 1%> α <5% Orchard 4134152

Multiple R	0.903409592	0.931844287	0.988906786	0.999063574	0.999360786
R Square	0.816148891	0.868333776	0.977936631	0.998128024	0.998721981
Adjusted R Square	**0.779378669**	**0.802500664**	**0.955873263**	**0.994384073**	**0.992331884**
Standard Error	657.9292287	622.4984155	294.2431354	104.9702314	122.6590966
Observations	7	7	7	7	7
STEPS	1	2	3	4	5
X Predictors					
	LandSat8 NDMI - February	LandSat8 NDMI - February LandSat8 NDVI - January	LandSat8 NDMI - February LandSat8 NDVI - January LandSat8 GCI - February	LandSat8 NDMI - February LandSat8 NDVI - January LandSat8 GCI - February LandSat8 GCI - January	LandSat8 NDMI - February LandSat8 NDVI - January LandSat8 GCI - February LandSat8 GCI - January LandSat8 AVI - September

Adjusted R Square from lowest p-values of CLIMATE DATASETS with 1%> α <5% Orchard 4134152

Multiple R	0.653417786	0.754145356	0.831945025	0.859844401	0.860677089	0.878490775	0.904962218
R Square	0.426954803	0.568735218	0.692132524	0.739332394	0.740765051	0.771746042	0.818956616
Adjusted R Square	**0.379201036**	**0.49032344**	**0.599772281**	**0.623480124**	**0.578743208**	**0.576099793**	**0.607739334**
Standard Error	1096.026327	993.0996413	880.0332397	853.5705995	902.8569642	905.6852792	871.2301076
Observations	14	14	14	14	14	14	14
STEPS	1	2	3	4	5	6	7
X Predictors							
	Temperature Monthly Average (Raster datare	Monthly Average (Raster dataure	Monthly Average (Raster data)e	Monthly Average (Raster data	Monthly Average (Raster da	Temperature Monthly Average (Raster data) - October	Temperature Monthly Average (Raster data) - October
	Actual Evapotranspiration (TerraClimate) -	potranspiration (TerraClimate) -	hotranspiration (TerraClimate) -	transpiration (TerraClimate) -	transpiration (TerraClimate)	Actual Evapotranspiration (TerraClimate) - November	Actual Evapotranspiration (TerraClimate) - November
	Evaporation from the top of canopy (ERAS) n	from the top of canopy (ERA5	from the top of canopy (ERA			Evaporation from the top of canopy (ERA5) - October	Evaporation from the top of canopy (ERAS) - October
	Climate Water Deficit (TerraClimate) -	Water Deficit (TerraClimate				Climate Water Deficit (TerraClimate) - April	Climate Water Deficit (TerraClimate) - April
	Maximum Monthly Temperature (TerraClin					Maximum Monthly Temperature (TerraClimate) - October	Maximum Monthly Temperature (TerraClimate) - October
						Monthly Rainfall (mm) (TerraClimate) - November	Monthly Rainfall (mm) (TerraClimate) - November
							Precipitation Monthly Average (Raster data) - November

Adjusted R Square from lowest p-values of CLIMATE DATASETS with 1%> α <5% Orchard 4134152

Multiple R	0.653417786	0.754145356	0.831945025	0.859844401	0.904249225	0.904551686	0.904962218
R Square	0.426954803	0.568735218	0.692132524	0.739332394	0.81766666	0.818213753	0.818956616
Adjusted R Square	**0.379201036**	**0.49032344**	**0.599772281**	**0.623480124**	**0.703708323**	**0.66239697**	**0.607739334**
Standard Error	1096.026327	993.0996413	880.0332397	853.5705995	757.1906116	808.2554879	871.2301076
Observations	14	14	14	14	14	14	14
STEPS	1	2	3	4	5	6	7
X Predictors							
	Temperature Monthly Average (Raster datare	Monthly Average (Raster dataure	Monthly Average (Raster data)e	Monthly Average (Raster data	Monthly Average (Raster da	Temperature Monthly Average (Raster data) - October	Temperature Monthly Average (Raster data) - October
	Actual Evapotranspiration (TerraClimate) -	potranspiration (TerraClimate) -	hotranspiration (TerraClimate) -	transpiration (TerraClimate)		Actual Evapotranspiration (TerraClimate) - November	Actual Evapotranspiration (TerraClimate) - November
	Evaporation from the top of canopy (ERAS) n	from the top of canopy (ERA5	from the top of canopy (ERA			Evaporation from the top of canopy (ERA5) - October	Evaporation from the top of canopy (ERAS) - October
	Climate Water Deficit (TerraClimate) -	Water Deficit (TerraClimate				Climate Water Deficit (TerraClimate) - April	Climate Water Deficit (TerraClimate) - April
	Precipitation Monthly Average (Raster data					Precipitation Monthly Average (Raster data) - November	Precipitation Monthly Average (Raster data) - November
						Monthly Rainfall (mm) (TerraClimate) - November	Monthly Rainfall (mm) (TerraClimate) - November
							Maximum Monthly Temperature (TerraClimate) - October

Adjusted R Square from lowest p-values of Landsat 7 with 1%> α <5% Orchard 4134152

Multiple R	0.741195129	0.796890133	0.878826384	0.899897635	0.918390607	0.931730335	0.932675436
R Square	0.549370219	0.635033884	0.772335814	0.809815753	0.843441307	0.868121418	0.869883468
Adjusted R Square	**0.511817737**	**0.568676409**	**0.704036558**	**0.725289421**	**0.745592124**	**0.755082633**	**0.718080848**
Standard Error	971.9333681	913.581088	756.7710834	729.093433	701.6342329	688.4228639	738.5974169
Observations	14	14	14	14	14	14	14
STEPS	1	2	3	4	5	6	7
X Predictors							
	LandSat7 NDMI - March	LandSat7 NDMI - March LandSat7 NDVI - May	LandSat7 NDMI - March LandSat7 NDVI - May LandSat7 GCI - June	LandSat7 NDMI - March LandSat7 NDVI - May LandSat7 GCI - June LandSat7 AVI - February	LandSat7 NDMI - March LandSat7 NDVI - May LandSat7 GCI - June LandSat7 AVI - February LandSat7 AVI - January	LandSat7 NDMI - March LandSat7 NDVI - May LandSat7 GCI - June LandSat7 AVI - February LandSat7 AVI - January LandSat7 EVI - January	LandSat7 NDMI - March LandSat7 NDVI - May LandSat7 GCI - June LandSat7 AVI - February LandSat7 AVI - January LandSat7 EVI - January LandSat7 ReCI - November

Adjusted R Square from lowest p-values of Sentinel 2 with 1%> α <5% Orchard 4134152

Multiple R	0.605930204	0.97708124	0.998822966	0.999998949
R Square	0.367151412	0.954687749	0.997647318	0.999997897
Adjusted R Square	**0.208939265**	**0.924479582**	**0.994118295**	**0.999989485**
Standard Error	923.2823749	285.2737578	79.61249869	3.366097819
Observations	6	6	6	6
STEPS	1	2	3	4
X Predictors				
	Sentinel2 SIPI - July	Sentinel2 SIPI - July Sentinel2 ARVI - July	Sentinel2 SIPI - July Sentinel2 ARVI - July Sentinel2 NDVI - June	Sentinel2 SIPI - July Sentinel2 ARVI - July Sentinel2 NDVI - June Sentinel2 ReCI - July

Brown coloured text are hyperlinks to easily browse throughout the document

11.6.3 Orchard 5034172

Adjusted R Square from OVERALL lowest p-values with 1%> α <5% Orchard 5034172			
Multiple R	0.94313345	0.996623094	0.999632673
R Square	0.889500704	0.993257592	0.999265481
Adjusted R Square	**0.852667605**	**0.986515183**	**0.997061925**
Standard Error	335.2828917	101.4342895	47.34713568
Observations	5	5	5
STEPS	1	2	3
X Predictors			
	Evaporation from bare soil at the top of land surface (ERoration from bare soil at the top of land surface (ERA5) - Nove	Evaporation from bare soil at the top of land surface (ERA5) - November
	Specific humidity (FLDAS) - January	Specific humidity (FLDAS) - January
		Soil temperature (28-100 cm) (ERA5) - May

Adjusted R Square from lowest p-values with 1%> α <5% of METEO DATASETS Orchard 5034172						
Multiple R	0.619223329	0.726828379	0.789961977	0.821128503	0.832688066	0.833178568
R Square	0.383437531	0.528279493	0.624039925	0.674252019	0.693369415	0.694186526
Adjusted R Square	**0.339397355**	**0.455707107**	**0.530049907**	**0.555798208**	**0.540654123**	**0.426599736**
Standard Error	531.0823175	482.0676845	447.9376402	435.4936641	443.144183	494.7896792
Observations	16	16	16	16	16	16
STEPS	1	2	3	4	5	6
X Predictors						
	Evaporation from bare soil at the top of l Evaporation from bare soil at the top of land surface (ERA5) -	Evaporation from bare soil at the top of land surface (ERA5) - November	Evaporation from bare soil a Evaporation from bare soil at the top of land surface (ERA5) - November			
	Specific humidity (FLDAS) - January	Specific humidity (FLDAS) - January	Specific humidity (FLDAS) - J Specific humidity (FLDAS) - January			
		Soil temperature (7-28 cm) (ERA5) - April	Soil temperature (7-28 cm) Soil temperature (7-28 cm) (ERA5) - April			
			Evaporation from vegetatior Evaporation from vegetation transpiration (ERA5) - February			
			SPEI 1 month - August			

Adjusted R Square from LandSat 8 lowest p-values with 1%> α <5% Orchard 5034172			
Multiple R	0.820290948	0.832181176	0.834772509
R Square	0.67287724	0.69252551	0.696845141
Adjusted R Square	**0.607452688**	**0.538788265**	**0.393690282**
Standard Error	504.4351056	546.7760375	626.911972
Observations	7	7	7
STEPS	1	2	3
X Predictors			
	LandSat8 NDMI - April	LandSat8 NDMI - April	LandSat8 NDMI - April
		LandSat8 GCI - November	LandSat8 GCI - November
			LandSat8 NDVI - November

Adjusted R Square from Sentinel 2 lowest p-values with 1%> α <5% Orchard 5034172		
Multiple R	0.922604301	0.977457894
R Square	0.851198697	0.955423935
Adjusted R Square	**0.801598263**	**0.91084787**
Standard Error	389.0765036	260.8123365
Observations	5	5
STEPS	1	2
X Predictors		
	Sentinel2 NDVI - April	Sentinel2 NDVI - April
		Sentinel2 GNDVI - November

Brown coloured text are hyperlinks to easily browse throughout the document

11.6.4 Orchard 4434212

Adjusted R Square from OVERALL lowest p-values with 1%> α <5% Orchard 4434212							
Multiple R	0.839786473	0.916392164	0.941194484	0.980437162	0.987990919	0.999134955	0.999138653
R Square	0.70524132	0.839774599	0.885847057	0.961257028	0.976126055	0.998270659	0.998278047
Adjusted R Square	**0.663132937**	**0.786366132**	**0.817355291**	**0.922514056**	**0.936336147**	**0.993082635**	**0.986224377**
Standard Error	1926.642972	1534.288462	1418.650826	924.0243861	837.5648161	276.0846822	389.6077527
Observations	9	9	9	9	9	9	9
STEPS	1	2	3	4	5	6	7
X Predictors							
	Specific humidity (FLDAS) - July	Specific humidity (FLDAS) - July SPEI 3 months - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October MODIS GCI - November	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October MODIS GCI - November SPEI 1 month - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October MODIS GCI - November SPEI 1 month - September Climate Water Deficit (TerraClimate) - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October MODIS GCI - November SPEI 1 month - September Climate Water Deficit (TerraClimate) - September LandSat7 EVI - April

Adjusted R Square from METEO DATASETS lowest p-values with 1%> α <5% Orchard 4434212							
Multiple R	0.839786473	0.916392164	0.941194484	0.971311768	0.971365318	0.983362743	0.995818181
R Square	0.70524132	0.839774599	0.885847057	0.943446551	0.943550581	0.967002284	0.991653849
Adjusted R Square	**0.663132937**	**0.786366132**	**0.817355291**	**0.886893103**	**0.849468215**	**0.868009135**	**0.93320793**
Standard Error	1926.642972	1534.288462	1418.650826	1116.391476	1287.911653	1205.99055	857.7487186
Observations	9	9	9	9	9	9	9
STEPS	1	2	3	4	5	6	7
X Predictors							
	Specific humidity (FLDAS) - July	Specific humidity (FLDAS) - July SPEI 3 months - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October Rainfall rate (kg/m^2/s) (FLDAS) - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October Rainfall rate (kg/m^2/s) (FLDAS) - September Actual Evapotranspiration (TerraClimate) - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October Rainfall rate (kg/m^2/s) (FLDAS) - September Actual Evapotranspiration (TerraClimate) - September Monthly Rainfall (mm) (TerraClimate) - September	Specific humidity (FLDAS) - July SPEI 3 months - September Soil Moisture 10-40 cm (FLDAS) - October Rainfall rate (kg/m^2/s) (FLDAS) - September Actual Evapotranspiration (TerraClimate) - September Monthly Rainfall (mm) (TerraClimate) - September Climate Water Deficit (TerraClimate) - September

Adjusted R Square from LandSat 7 lowest p-values with 1%> α <5% Orchard 4434212							
Multiple R	0.810869454	0.909314885	0.979274237	0.981722241	0.981910849	0.99167383	0.992194025
R Square	0.657509271	0.826853561	0.958978031	0.963778559	0.964148915	0.983416984	0.984448984
Adjusted R Square	**0.608582024**	**0.769138081**	**0.93436485**	**0.927557117**	**0.904397107**	**0.933667937**	**0.875591871**
Standard Error	2076.788886	1594.953837	850.4332043	893.4491291	1026.378356	854.9362283	1170.836754
Observations	9	9	9	9	9	9	9
STEPS	1	2	3	4	5	6	7
X Predictors							
	LandSat7 EVI - April	LandSat7 EVI - April LandSat7 SIPI - June	LandSat7 EVI - April LandSat7 SIPI - June LandSat7 GCI - June	LandSat7 EVI - April LandSat7 SIPI - June LandSat7 GCI - June LandSat7 AVI - November	LandSat7 EVI - April LandSat7 SIPI - June LandSat7 GCI - June LandSat7 AVI - November LandSat7 EVI2 - April	LandSat7 EVI - April LandSat7 SIPI - June LandSat7 GCI - June LandSat7 AVI - November LandSat7 EVI2 - April LandSat7 SAVI - April	LandSat7 EVI - April LandSat7 SIPI - June LandSat7 GCI - June andSat7 AVI - Novembe LandSat7 EVI2 - April LandSat7 SAVI - April LandSat7 AVI - April

Adjusted R Square from MODIS lowest p-values with 1%> α <5% Orchard 4434212				
Multiple R	0.739151874	0.959992225	0.963724789	0.969075047
R Square	0.546345492	0.921585073	0.92876547	0.939106448
Adjusted R Square	**0.481537706**	**0.895446764**	**0.886024751**	**0.878212895**
Standard Error	2390.179714	1073.348386	1120.668698	1158.437534
Observations	9	9	9	9
STEPS	1	2	3	4
X Predictors				
	MODIS GCI - Novembe	MODIS GCI - November EVI from MODIS - May	MODIS GCI - November EVI from MODIS - May MODIS NDVI - May	MODIS GCI - November EVI from MODIS - May MODIS NDVI - May MODIS AVI - May

Adjusted R Square from Sentinel 2 lowest p-values with 1%> α <5% Orchard 4434212		
Multiple R	0.966106537	0.967166218
R Square	0.93336184	0.935410494
Adjusted R Square	**0.91114912**	**0.870820988**
Standard Error	1003.023798	1209.417781
Observations	5	5
STEPS	1	2
X Predictors		
	Sentinel2 SIPI - April	Sentinel2 SIPI - April Sentinel2 NDMI - December

Brown coloured text are hyperlinks to easily browse throughout the document

11.6.5 Orchard from Cubillas et al (2022)

OVERALL lowest p-values with α <1% Orchard from Cubillas et al (2022)				
Multiple R	0.911668945	0.950546826	0.96768668	0.998304916
R Square	0.831140266	0.903539268	0.936417511	0.996612704
Adjusted R Square	**0.802996977**	**0.864954976**	**0.888730644**	**0.99209631**
Standard Error	1882.221961	1558.381143	1414.562623	377.0069277
Observations	8	8	8	8
STEPS	1	2	3	4
X Predictors				
	Leaf Index Area (MOD15A2H.061) - June	Leaf Index Area (MOD15A2H.061) - June Evaporation from vegetation transpiration (ERA5) - February	Leaf Index Area (MOD15A2H.061) - June Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June	Leaf Index Area (MOD15A2H.061) - June Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June LandSat7 SIPI - December

Adjusted R Square from lowest p-values of Meteo datasets with 1%> α <5% Orchard from Cubillas et al (2022)					
Multiple R	0.862993359	0.946547122	0.963027896	0.964200726	0.96551723
R Square	0.744757537	0.895951455	0.927422728	0.92968304	0.932223522
Adjusted R Square	**0.702217126**	**0.854332037**	**0.872989774**	**0.835927094**	**0.762782328**
Standard Error	2314.111334	1618.513832	1511.310626	1717.721797	2065.417778
Observations	8	8	8	8	8
STEPS	1	2	3	4	5
X Predictors					
	Evaporation from vegetation transpiration (ERA5) - February	Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June	Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June SPEI 12 months - March	Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June SPEI 12 months - March Soil Moisture (TerraClimate) - December	Evaporation from vegetation transpiration (ERA5) - February SPEI 3 months - June SPEI 12 months - March Soil Moisture (TerraClimate) - December Precipitation Monthly Average (Raster data) - July

Adjusted R Square from lowest p-values of Landsat 7 with 1%> α <5% Orchard from Cubillas et al (2022)					
Multiple R	0.845486027	0.940316046	0.952778961	0.953113943	0.96722286
R Square	0.714846623	0.884194267	0.907787749	0.908426188	0.935520061
Adjusted R Square	**0.66732106**	**0.837871974**	**0.838628561**	**0.786327771**	**0.774320214**
Standard Error	2445.94702	1707.510721	1703.521921	1960.236319	2014.562387
Observations	8	8	8	8	8
STEPS	1	2	3	4	5
X Predictors					
	LandSat7 SIPI - December	LandSat7 SIPI - December LandSat7 EVI - June	LandSat7 SIPI - December LandSat7 EVI - June LandSat7 SAVI - April	LandSat7 SIPI - December LandSat7 EVI - June LandSat7 SAVI - April LandSat7 EVI - April	LandSat7 SIPI - December LandSat7 EVI - June LandSat7 SAVI - April LandSat7 EVI - April LandSat7 GNDVI - February

Brown coloured text are hyperlinks to easily browse throughout the document

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. Anthony Lawther: The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).

2. Rickard Hansen: Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).

3. Emil Bayramov: Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).

4. Rafael Villarreal Pacheco: Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).

5. Siri Oestreich Waage: a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).

6. Edgar Pimiento: Shallow landslide susceptibility – Modelling and validation (2010).

7. Martina Schäfer: Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).

8. August Pieter van Waarden-Nagel: Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).

9. Samira Muhammad: Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).

10. Fredros Oketch Okumu: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).

11. Svajunas Plunge: Advanced decision support methods for solving diffuse water pollution problems (2011).

12. Jonathan Higgins: Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).

13. Mårten Karlberg: Mobile Map Client API: Design and Implementation for Android (2011).

14. Jeanette McBride: Mapping Chicago area urban tree canopy using color infrared imagery (2011).

15. Andrew Farina: Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).

16. David Kanyari: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. Laura V. Drews: Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).

18. Qaisar Nadeem: Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).

19. Ahmed Mohamed El Saeid Mustafa: Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).

20. Daniel Patrick Taylor: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).

21. Angeleta Oveta Wilson: A Participatory GIS approach to unearthing Manchester's Cultural Heritage 'gold mine' (2013).

22. Ola Svensson: Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).

23. Monika Ogden: Land use impact on water quality in two river systems in South Africa (2013).

24. Stefan Rova: A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).

25. Yann Buhot: Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).

26. Christina Fotiou: Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).

27. Inese Linuza: Accuracy Assessment in Glacier Change Analysis (2014).

28. Agnieszka Griffin: Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).

29. Brynja Guðmundsdóttir: Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).

30. Oleksandr Nekrasov: Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).

31. Sarah Tressel: Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).

32. Caroline Gevaert: Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).

33. Salem Jamal-Uddeen: Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).

34. Samanah Seyedi-Shandiz: Schematic representation of geographical railway network at the Swedish Transport Administration (2014).

35. Kazi Masel Ullah: Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).

36. Alexia Chang-Wailing Spitteler: Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. Alessandro De Martino: Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).

38. Alireza Mollasalehi: GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).

39. Negin A. Sanati: Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).

40. Karen McIntyre: Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).

41. Kees van Duijvendijk: Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).

42. Sebastian Andersson Hylander: Evaluation of cultural ecosystem services using GIS (2015).

43. Deborah Bowyer: Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).

44. Stefan Arvidsson: Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).

45. Damián Giménez Cruz: GIS-based optimal localisation of beekeeping in rural Kenya (2016).

46. Alejandra Narváez Vallejo: Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).

47. Anna Lundgren: Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).

48. Oluwatomi Esther Adejoro: Does location also matter? A spatial analysis of social achievements of young South Australians (2016).

49. Hristo Dobrev Tomov: Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).

50. Vincent Muller: Impact of Security Context on Mobile Clinic Activities A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).

51. Gezahagn Negash Seboka: Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).

52. Holly Buhler: Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).

53. Lars Ole Grottenberg: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).

54. Sean Grant: The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).

55. Irshad Jamal: Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).

56. Fulgencio Sanmartín: Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. Nezha Acil: Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).

58. Julia Hjalmarsson: A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).

59. Mathewos Tamiru Amato: Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).

60. Karim Alaa El Din Mohamed Soliman El Attar: Bicycling Suitability in Downtown, Cairo, Egypt (2016).

61. Gilbert Akol Echelai: Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).

62. Terje Slinning: Analytic comparison of multibeam echo soundings (2016).

63. Gréta Hlín Sveinsdóttir: GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).

64. Jonas Sjögren: Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).

65. Nadine Raska: 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).

66. Panagiotis Symeonidis: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).

67. Michaela Bobeck: A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).

68. Raghdaa Eissa: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).

69. Zahra Khais Shahid: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).

70. Mirza Amir Liaquat Baig: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).

71. Joakim Jörwall: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).

72. Elin Haettner: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).

73. Harry Eriksson: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).

74. Daniel Gardevärn: PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).

75. Kim Friberg: Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).

76. Viktor Svanerud: Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).

77. Lisa-Gaye Greene: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).

78. Katarina Jemec Parker: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).

79. Angela Kabiru: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).

80. Kristean Björkmann: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).

81. Williams Erhunmonmen Ojo: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).

82. Daniel Assefa: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).

83. Adela Nistora: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).

84. Marc Seliger: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).

85. Luka Jovičić: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).

86. Christina Kourdounouli: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).

87. Jeremy Azzopardi: Effect of distance measures and feature representations on distance-based accessibility measures (2018).

88. Patrick Kabatha: An open source web GIS tool for analysis and visualization of elephant GPS telemetry data, alongside environmental and anthropogenic variables (2018).

89. Richard Alphonce Giliba: Effects of Climate Change on Potential Geographical Distribution of Prunus africana (African cherry) in the Eastern Arc Mountain Forests of Tanzania (2018).

90. Eiður Kristinn Eiðsson: Transformation and linking of authoritative multi-scale geodata for the Semantic Web: A case study of Swedish national building data sets (2018).

91. Niamh Harty: HOP!: a PGIS and citizen science approach to monitoring the condition of upland paths (2018).

92. José Estuardo Jara Alvear: Solar photovoltaic potential to complement hydropower in Ecuador: A GIS-based framework of analysis (2018).

93. Brendan O'Neill: Multicriteria Site Suitability for Algal Biofuel Production Facilities (2018).

94. Roman Spataru: Spatial-temporal GIS analysis in public health – a case study of polio disease (2018).

95. Alicja Miodońska: Assessing evolution of ice caps in Suðurland, Iceland, in years 1986 - 2014, using multispectral satellite imagery (2019).

96. Dennis Lindell Schettini: A Spatial Analysis of Homicide Crime's Distribution and Association with Deprivation in Stockholm Between 2010-2017 (2019).

97. Damiano Vesentini: The Po Delta Biosphere Reserve: Management challenges and priorities deriving from anthropogenic pressure and sea level rise (2019).

98. Emilie Arnesten: Impacts of future sea level rise and high water on roads, railways and environmental objects: a GIS analysis of the potential effects of increasing sea levels and highest projected high water in Scania, Sweden (2019).

99. Syed Muhammad Amir Raza: Comparison of geospatial support in RDF stores: Evaluation for ICOS Carbon Portal metadata (2019).

100. Hemin Tofiq: Investigating the accuracy of Digital Elevation Models from UAV images in areas with low contrast: A sandy beach as a case study (2019).

101. Evangelos Vafeiadis: Exploring the distribution of accessibility by public transport using spatial analysis. A case study for retail concentrations and public hospitals in Athens (2019).

102. Milan Sekulic: Multi-Criteria GIS modelling for optimal alignment of roadway by-passes in the Tlokweng Planning Area, Botswana (2019).

103. Ingrid Piirisaar: A multi-criteria GIS analysis for siting of utility-scale photovoltaic solar plants in county Kilkenny, Ireland (2019).

104. Nigel Fox: Plant phenology and climate change: possible effect on the onset of various wild plant species' first flowering day in the UK (2019).

105. Gunnar Hesch: Linking conflict events and cropland development in Afghanistan, 2001 to 2011, using MODIS land cover data and Uppsala Conflict Data Programme (2019).

106. Elijah Njoku: Analysis of spatial-temporal pattern of Land Surface Temperature (LST) due to NDVI and elevation in Ilorin, Nigeria (2019).

107. Katalin Bunyevácz: Development of a GIS methodology to evaluate informal urban green areas for inclusion in a community governance program (2019).

108. Paul dos Santos: Automating synthetic trip data generation for an agent-based simulation of urban mobility (2019).

109. Robert O' Dwyer: Land cover changes in Southern Sweden from the mid-Holocene to present day: Insights for ecosystem service assessments (2019).

110. Daniel Klingmyr: Global scale patterns and trends in tropospheric NO2 concentrations (2019).

111. Marwa Farouk Elkabbany: Sea Level Rise Vulnerability Assessment for Abu Dhabi, United Arab Emirates (2019).

112. Jip Jan van Zoonen: Aspects of Error Quantification and Evaluation in Digital Elevation Models for Glacier Surfaces (2020).

113. Georgios Efthymiou: The use of bicycles in a mid-sized city – benefits and obstacles identified using a questionnaire and GIS (2020).

114. Haruna Olayiwola Jimoh: Assessment of Urban Sprawl in MOWE/IBAFO Axis of Ogun State using GIS Capabilities (2020).

115. Nikolaos Barmpas Zachariadis: Development of an iOS, Augmented Reality for disaster management (2020).

116. Ida Storm: ICOS Atmospheric Stations: Spatial Characterization of CO2 Footprint Areas and Evaluating the Uncertainties of Modelled CO2 Concentrations (2020).

117. Alon Zuta: Evaluation of water stress mapping methods in vineyards using airborne thermal imaging (2020).

118. Marcus Eriksson: Evaluating structural landscape development in the municipality Upplands-Bro, using landscape metrics indices (2020).

119. Ane Rahbek Vierø: Connectivity for Cyclists? A Network Analysis of Copenhagen's Bike Lanes (2020).

120. Cecilia Baggini: Changes in habitat suitability for three declining Anatidae species in saltmarshes on the Mersey estuary, North-West England (2020).

121. Bakrad Balabanian: Transportation and Its Effect on Student Performance (2020).

122. Ali Al Farid: Knowledge and Data Driven Approaches for Hydrocarbon Microseepage Characterizations: An Application of Satellite Remote Sensing (2020).

123. Bartlomiej Kolodziejczyk: Distribution Modelling of Gene Drive-Modified Mosquitoes and Their Effects on Wild Populations (2020).

124. Alexis Cazorla: Decreasing organic nitrogen concentrations in European water bodies - links to organic carbon trends and land cover (2020).

125. Kharid Mwakoba: Remote sensing analysis of land cover/use conditions of community-based wildlife conservation areas in Tanzania (2021).

126. Chinatsu Endo: Remote Sensing Based Pre-Season Yellow Rust Early Warning in Oromia, Ethiopia (2021).

127. Berit Mohr: Using remote sensing and land abandonment as a proxy for long-term human out-migration. A Case Study: Al-Hassakeh Governorate, Syria (2021).

128. Kanchana Nirmali Bandaranayake: Considering future precipitation in delineation locations for water storage systems - Case study Sri Lanka (2021).

129. Emma Bylund: Dynamics of net primary production and food availability in the aftermath of the 2004 and 2007 desert locust outbreaks in Niger and Yemen (2021).

130. Shawn Pace: Urban infrastructure inundation risk from permanent sea-level rise scenarios in London (UK), Bangkok (Thailand) and Mumbai (India): A comparative analysis (2021).

131. Oskar Evert Johansson: The hydrodynamic impacts of Estuarine Oyster reefs, and the application of drone technology to this study (2021).

132. Pritam Kumarsingh: A Case Study to develop and test GIS/SDSS methods to assess the production capacity of a Cocoa Site in Trinidad and Tobago (2021).

133. Muhammad Imran Khan: Property Tax Mapping and Assessment using GIS (2021).

134. Domna Kanari: Mining geosocial data from Flickr to explore tourism patterns: The case study of Athens (2021).

135. Mona Tykesson Klubien: Livestock-MRSA in Danish pig farms (2021).

136. Ove Njøten: Comparing radar satellites. Use of Sentinel-1 leads to an increase in oil spill alerts in Norwegian waters (2021).

137. Panagiotis Patrinos: Change of heating fuel consumption patterns produced by the economic crisis in Greece (2021).

138. Lukasz Langowski: Assessing the suitability of using Sentinel-1A SAR multi-temporal imagery to detect fallow periods between rice crops (2021).

139. Jonas Tillman: Perception accuracy and user acceptance of legend designs for opacity data mapping in GIS (2022).

140. Gabriela Olekszyk: ALS (Airborne LIDAR) accuracy: Can potential low data quality of ground points be modelled/detected? Case study of 2016 LIDAR capture over Auckland, New Zealand (2022).

141. Luke Aspland: Weights of Evidence Predictive Modelling in Archaeology (2022).

142. Luís Fareleira Gomes: The influence of climate, population density, tree species and land cover on fire pattern in mainland Portugal (2022).

143. Andreas Eriksson: Mapping Fire Salamander (Salamandra salamandra) Habitat Suitability in Baden-Württemberg with Multi-Temporal Sentinel-1 and Sentinel-2 Imagery (2022).

144. Lisbet Hougaard Baklid: Geographical expansion rate of a brown bear population in Fennoscandia and the factors explaining the directional variations (2022).

145. Victoria Persson: Mussels in deep water with climate change: Spatial distribution of mussel (Mytilus galloprovincialis) growth offshore in the French Mediterranean with respect to climate change scenario RCP 8.5 Long Term and Integrated Multi-Trophic Aquaculture (IMTA) using Dynamic Energy Budget (DEB) modelling (2022).

146. Benjamin Bernard Fabien Gérard Borgeais: Implementing a multi-criteria GIS analysis and predictive modelling to locate Upper Palaeolithic decorated caves in the Périgord noir, France (2022).

147. Bernat Dorado-Guerrero: Assessing the impact of post-fire restoration interventions using spectral vegetation indices: A case study in El Bruc, Spain (2022).

148. Ignatius Gabriel Aloysius Maria Perera: The Influence of Natural Radon Occurrence on the Severity of the COVID-19 Pandemic in Germany: A Spatial Analysis (2022).

149. Mark Overton: An Analysis of Spatially-enabled Mobile Decision Support Systems in a Collaborative Decision-Making Environment (2022).

150. Viggo Lunde: Analysing methods for visualizing time-series datasets in open-source web mapping (2022).

151. Johan Viscarra Hansson: Distribution Analysis of Impatiens glandulifera in Kronoberg County and a Pest Risk Map for Alvesta Municipality (2022).

152. Vincenzo Poppiti: GIS and Tourism: Developing strategies for new touristic flows after the Covid-19 pandemic (2022).

153. Henrik Hagelin: Wildfire growth modelling in Sweden - A suitability assessment of available data (2023).

154. Gabriel Romeo Ferriols Pavico: Where there is road, there is fire (influence): An exploratory study on the influence of roads in the spatial patterns of Swedish wildfires of 2018 (2023).

155. Colin Robert Potter: Using a GIS to enable an economic, land use and energy output comparison between small wind powered turbines and large-scale wind farms: the case of Oslo, Norway (2023).

156. Krystyna Muszel: Impact of Sea Surface Temperature and Salinity on Phytoplankton blooms phenology in the North Sea (2023).

157. Tobias Rydlinge: Urban tree canopy mapping - an open source deep learning approach (2023).

158. Albert Wellendorf: Multi-scale Bark Beetle Predictions Using Machine Learning (2023).

159. Manolis Papadakis: Use of Satellite Remote Sensing for Detecting Archaeological Features: An Example from Ancient Corinth, Greece (2023).

160. Konstantinos Sourlamtas: Developing a Geographical Information System for a water and sewer network, for monitoring, identification and leak repair - Case study: Municipal Water Company of Naoussa, Greece (2023).

161. Xiaoming Wang: Identification of restoration hotspots in landscape-scale green infrastructure planning based on model-predicted connectivity forest (2023).

162. Sarah Sienaert: Usability of Sentinel-1 C-band VV and VH SAR data for the detection of flooded oil palm (2023).

163. Katarina Ekeroot: Uncovering the spatial relationships between Covid-19 vaccine coverage and local politics in Sweden (2023).

164. Nikolaos Kouskoulis: Exploring patterns in risk factors for bark beetle attack during outbreaks triggered by drought stress with harvester data on attacked trees: A case study in Southeastern Sweden (2023).

165. Jonas Almén: Geographic polarization and clustering of partisan voting: A local-level analysis of Stockholm Municipality (2023).

166. Sara Sharon Jones: Tree species impact on Forest Fire Spread Susceptibility in Sweden (2023).

167. Takura Matswetu: Towards a Geographic Information Systems and Data-Driven Integration Management. Studying holistic integration through spatial accessibility of services in Tampere, Finland. (2023).

168. Duncan Jones: Investigating the influence of the tidal regime on harbour porpoise Phocoena phocoena distribution in Mount's Bay, Cornwall (2023).

169. Jason Craig Joubert: A comparison of remote sensed semi-arid grassland vegetation anomalies detected using MODIS and Sentinel-3, with anomalies in ground-based eddy covariance flux measurements (2023).

170. Anastasia Sarelli: Land cover classification using machine-learning techniques applied to fused multi-modal satellite imagery and time series data (2024).

171. Athanasios Senteles: Integrating Local Knowledge into the Spatial Analysis of Wind Power: The case study of Northern Tzoumerka, Greece (2024).

172. Rebecca Borg: Using GIS and satellite data to assess access of green area for children living in growing cities (2024).

173. Panagiotis–Dimitrios Tsachageas: Multicriteria Evaluation in Real Estate Land-use Suitability Analysis: The case of Volos, Greece (2024).

174. Hugo Nilsson: Inferring lane-level topology of signalised intersections from aerial imagery and OpenStreetMap using deep learning (2024).

175. Pavlos Alexantonakis: Estimating lake water volume fluctuations using Sentinel-2 and ICESat-2 remote sensing data (2024).

176. Karl-Martin Wigen: Physical barriers and where to find them (2024).

177. Martin Storsnes: Temporal RX-algorithm performance on Sentinel-2 images (2024).

178. Saulė Gabrielė Petraitytė: The Relation Between Covid-19 Vaccination and Voting Trends in Lithuania: A Spatial Analysis (2024).

179. Pedro Martinez Duran: Olive yield forecasting from remote sensing and climate datasets in the Jaen province (Spain) (2024).