



Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Master Thesis

Normalized Difference Vegetation Index derived metrics and Spatial analysis for Area Wide Pest Management of the Olive Fruit Fly in Italy

Submitted by

Student LACAGNINA Giacomo, BSc

in the framework of the Master programme

Mountain Forestry

in partial fulfilment of the requirements for the academic degree

Diplom-Ingenieur

Vienna, January 2021

Supervisor:

Univ.Prof. Dr.rer.nat. Clement Atzberger

Institute of Geomatics

Department for Space, Landscape and Infrastructure (RALI)

1. Affidavit

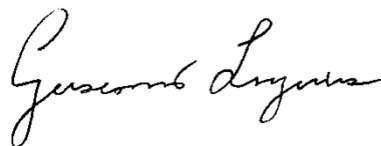
I hereby declare that I have authored this master thesis independently, and that I have not used any assistance other than that which is permitted. The work contained herein is my own except where explicitly stated otherwise. All ideas taken in wording or in basic content from unpublished sources or from published literature are duly identified and cited, and the precise references included.

I further declare that this master thesis has not been submitted, in whole or in part, in the same or a similar form, to any other educational institution as part of the requirements for an academic degree.

I hereby confirm that I am familiar with the standards of Scientific Integrity and with the guidelines of Good Scientific Practice, and that this work fully complies with these standards and guidelines.

Vienna, 02/02/2022

Giacomo LACAGNINA (*manu propria*)

A handwritten signature in black ink, reading "Giacomo Lacagnina". The signature is written in a cursive style with a large initial 'G' and 'L'.

2. Preface

The present research was conducted as part of an internship at the “Comunità Montana del Sebino Bresciano” in the framework of the “Erasmus + Traineeship” program, over the period from March 2021 to October 2021. The research questions were constructed in collaboration the Forest and Agriculture Office’s technicians with the intention to fit the study to local needs. The host institution provided technical and financial support for the collection of trapping data.

3. Acknowledgements

I would like to express my gratitude to my supervisor, Professor Clement Atzberger and to my co-supervisor Dr. Francesco Vuolo, who guided and supported me throughout this research.

I would like to acknowledge the staff of the Office of Agriculture and Forestry of the Comunità Montana of Sebino Bresciano for the technical support and for sharing their precious expertise and knowledge about the territory. I extend my special thanks to all the farmers who gave us the permits to inspect their olive groves for the collection of trapping data during the whole season. I would like to say thanks to my family for their support and to my Friends in Vienna who have been a great source of motivation in the last year.

4. Table of content

1.	Affidavit	i
2.	Preface.....	ii
3.	Acknowledgements	iii
4.	Table of content	iv
5.	Abstract	v
1.	Introduction.....	1
1.1.	The Olive Fruit Fly	2
1.2.	Remote Sensing for Area Wide Pest Management	4
1.3.	Modelling pest risk using climatic and geographic data	7
1.4.	Aim of the study and research questions	10
2.	Materials and methods	11
2.1.	Study Area: Comunità Montana del Sebino Bresciano	11
2.2.	The olive oil sector and classification of the olive groves	12
2.3.	Sampling and data Collection	13
2.4.	NDVI metrics as predictors of the OFF-population density	15
2.5.	Spatial analysis of the OFF-population density in relation to geographic factors	20
3.	Results and discussion	23
3.1.	NDVI metrics as predictors of the OFF-population density	24
3.2.	Spatial analysis of the OFF-population density in relation to geographic factors	28
4.	Conclusions.....	36
5.	References	38
6.	List of abbreviations	43
7.	List of tables / figures	44
8.	Appendix A: Time-series of trapping data during the monitoring campaign 2021	47
9.	Appendix B: Scripts.....	50
1.	B.1 GEE Script: Extraction of the NDVI median values for the sample groves	50
2.	B.2 R-Script: Multi-layer perceptron LOOCV and creation of the OFF risk output raster	52
10.	Appendix C: NDVI curves for the sample olive groves	55

5. Abstract

Area Wide Pest Management (AWPM) consists in the development of pest management tactics over large spatial areas and represents a potential solution to reduce costs and environmental impacts of pest infestation (Giles et al., 2008) thanks to the more targeted use of pesticides. The Olive Fruit Fly (OFF) is a major pest of Olive trees in Italy however, the intensity of damages varies according to the location of the olive groves, and the associated climatic factors. Relating the OFF distribution and dynamics with climatic factors is essential to design and implement successful AWPM strategies. This research integrated the use of trapping data, satellite image data and mathematical models to predict the OFF infestation in an area of Northern Italy. (I) Normalized Difference Vegetation Index derived metrics, from Copernicus Sentinel-2 image data, and (II) topographic variables (i.e. elevation, aspect and distance from the coast), were used as a proxy of temperature, to predict the OFF infestation. The results shown that (I) NDVI data from Sentinel-2 satellite acquisitions do not meet the spatial requirements for the extraction of seasonality parameters for the olive groves in the study area. This is probably due to the small size, the irregular shape, and the presence of non-vegetation pixels in the neighbourhoods of the parcels. Nevertheless, the proposed methodology might provide useful outcomes to pest managers, and could be tested in areas with extensive olive plantations. The spatial analysis (II) was conducted through a Neural Network regression and the model topology selected by Cross-Validation. Although the resulting model will require an external data set to be tested, this study provided a baseline model to predict the OFF infestation and produced a map of the OFF distribution in the study area. Information on the olive load and management practices should be considered in further studies to include other factors influencing the OFF-population dynamics.

1. Introduction

In a world facing an increasing food demand associated with limited cropland availability, the reduction of yield losses caused by crop pests is of great importance to achieve sustainable food production. In fact, excessive pest populations can lead to widespread short term crop failures and food insecurity (Yan et al., 2015). Several different strategies have been implemented to reduce pest-related risks such as the use of pesticides, resistance cultivars and a number of management strategies, including Area Wide Pest Management.

Area Wide Pest Management (AWPM) consists in the development and application of pest management tactics over a large spatial area, and it has been recognized as a potential solution to reduce costs and environmental impacts of pest infestation for both agriculture and forest sector (Giles et al., 2008). Nevertheless, if insect monitoring networks are not integrated with information of population ecology and environmental factors influencing the pest population, AWPM can be unsustainably costly and might lead to inconsistent results.

The spatial variation of insect populations depends on environmental factors, such as topography and climate, and biotic interactions including anthropogenic factors. Relating pest distribution with such factors, and understanding their spatial dynamics, is an essential part to design and implement successful pest management strategies (Kounatidis et al., 2008; Petacchi et al., 2015).

The olive tree *Olea europea* subsp. *sativa* L. is a typical Mediterranean culture which developed in areas characterized by mild climate and suffer from very low temperature in winter.

It is a traditional cultivation in Italy, which is well-known for the production for both table-olives and olive oil. Extensive cultivations can be found in the southern and central regions, but there is also a significant presence of this species in the area surrounding the subalpine lakes in the North of the Country. In these areas, the olive trees take advantage of the mitigation effect exerted by the water mass on the local climate (Rolfi, 2003).

The Olive Fruit Fly (OFF), *Bractocera oleae* Gmelin, is a major pest of Olive trees in Southern Europe as well as other olive-growing regions all over the world (García-Chapeton et al., 2020).

Even though the distribution of the OFF interest all olive-cultivated areas in Italy, the intensity of damages varies according to the location, the elevation of the olive groves, and to the associated climatic factors (Delrio & Lentini, 2016).

Especially temperature, have a strong influence on the OFF population on the mortality rate, length of the reproductive period and duration of the development cycle. It also has an indirect effect influencing the phenology of the Olive trees (Delrio & Lentini, 2016).

In the next section, we will provide an overview of the biology and life cycle of the pest and the most common adopted strategies for its management and control.

The following part of this chapter is a review of different approaches to study pest-related risk and pest population dynamics in relation to climatic factors over large areas. Pros and cons of the different methodology are presented, and their relevance to the present research is eventually discussed.

1.1. The Olive Fruit Fly

The Olive Fruit Fly is an Insect pertaining to the order of *Diptera* and the family of *Tephritidae*. The Pupae generally overwinter in the soil and the first generation of adults emerges in spring (march-may). At this time of the year, there is no fruits available on the olive trees and the adults will have to wait until the seed begins to harden to break the reproductive-dormancy and lay their eggs into the olives. This phenological phase is reached in Northern Italy around mid-July (Petacchi et al., 2015).

The biology of the insect is therefore related to olive fruit age and availability (Gutierrez et al., 2009) and both are linked to local weather conditions. The development Temperature threshold at have been reported by Rice (2000) and are summarized in (Table 1).

Table 1 Temperature thresholds for the development of the Olive Fruit Fly (Rice, 2000)

Stage	Lower	Upper
	°C	°C
Egg	6-8	35-38
Larvae	4-8	35
Pupae	5-9	30
Adult	4	39

Adults Females prefer to lay single eggs in untouched fruit, but multiple attacks can happen in case of fruit scarcity (Gutierrez et al., 2009), and can lay up to 200-250 eggs in their life cycle (Ferrari et al., 2006). The eggs hatch after few days and the larvae begins feeding on the olive pulp. The larval stages (three larval instars) last around 20 days and then pupation occurs. After one week the pupae turns into adult and emerge from the fruits. In summer months the whole cycle can be even faster and, from egg to adult, takes around 3 weeks to complete (Ferrari et al., 2006).

The number of generations/years vary according to the climatic areas. It can reach up to 6-7 generation in warmer regions while only 2-3 generations are found in relatively colder ones (such is the case of this study area). The second generation is the one causing the most severe damages to the olive production (Delrio & Lentini, 2016).

The damage caused by the OFF to olive cultivations interest the fruits, causing the partial destruction of the pulp and early fruit fallen. Moreover, the insect causes an alteration of the pulps which facilitate the entry of secondary infections by bacteria and fungi that rot the fruit (García-Chapeton et al., 2020). Rotting eventually results in acidification and a decrease of oil quality and loss of economic value.

Monitoring the OFF-adult's population is normally done using chromotropic traps activated with sexual pheromones and food-bates. Trap measurements are used as one of the indicators for the OFF infestation and so, as an information support tool, for the application of pesticides, which are active on the adult's population (Kalamatianos et al., 2017) and preventive treatments. One OFF per trap per inspection date (i.e. 5 days) and five OFFs per trap per inspection date were reported as intervention thresholds by Castrignano et al. (2012) for table olives and olive oil production, respectively. An economic threshold of 7 olive fruit flies per trap in summer and 5 OFFs in autumn is instead given by Kalamatianos et al. (2017)

However it is worth noting that the intervention threshold for the application of larval pesticides have to be set based on the active infestation rate (I-II larval stages), normally over a sample of 100 olives. Haniotakis (2005), reported an economic threshold based on preimaginal infestation rate of 8-10 % of fruit infested for oil production while the acceptable infestation for table olive was set as zero.

The threshold can differ from region to region and varies according to the nature of the treatment to be applied and to the final use of the fruits (table olives or olive oil).

The chemical control of the OFF is done using pesticides such as *Dimethoate*, *Fosmet* and *Imidacloprid*, which have a lethal effect on the larvae. According to local farmers and experts, the Dimethoate was the most utilized product in the area due to its high efficacy and low solubility in fats. However, the use of this pesticide has been banned by COMMISSION IMPLEMENTING REGULATION (EU) 2019/1090 of 26 June 2019 Concerning the Non-Renewal of Approval of the Active Substance Dimethoate, because it is considered a threaten to the user's health as well as for non-target mammals and arthropods including honey bees.

Even though alternative methods, such as mass-trapping, bait sprays and preventive treatments using organic products are already available, they still have some practical limitations and are not completely widespread among farmers. The banning of the dimethoate put more pressure on both technicians and producers to find effective and more sustainable solutions.

In addition to that, the efficacy of all preventive methods depends largely on the size of the treated area (Belcari, 2019). The best results are obtained when the treatment is applied on large extensions. These facts support the need for innovative solutions and consistent information on the OFF-population dynamics to support decision making by Area Wide Pest Managers and farmers.

1.2. Remote Sensing for Area Wide Pest Management

The term Remote Sensing (RS) refers to measurements of the radiation (reflected or emitted) from the Earth's surface, and it is based on the interaction between the electromagnetic radiation and soil or plant materials.

Among many applications, RS data allows with a minimum amount of ground sampling, the identification and assessment of crop stress and pest damage. Thanks to the wide coverage, repeatability, and comparatively low cost, with respect to field data collection, RS data represent a potential time and cost-effective alternative to provide information about insect outbreaks (Haghighian et al., 2020).

The detection of pest infestation by remote sensing techniques is based on detecting changes in the green parts of the plants. Vegetation indices (VIs) are indicators that describes the characteristics of vegetation (e.g. vegetation health, density, water content) for each pixel of an image and their main purpose is to enhance the information of the reflectance data, by extracting the variability due to vegetation characteristics.

Prabhakar et al.(2012), cited 24 different vegetation indices which have been used to detect different kinds of biotic stress, caused by arthropods, termites, fungi, or bacteria.

The Normalized Difference Vegetation Index (NDVI) is a VI describing the “greenness” of the vegetation and it is the most widely used for monitoring conditions of plants and vegetation worldwide (Acharya & Thapa, 2015).

The NDVI can be computed using Equation 1:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

where the RED and NIR are the surface reflection for red (665 nm) and near infrared (842 nm) bands, respectively. NDVI values range from +1 to -1, where negative values correspond to an absence of vegetation. The formula is based on the fact that chlorophyll absorbs RED radiations while the mesophyll leaf structure scatters NIR.

The NDVI was used in several studies to monitor biotic stresses and insect breakouts both in the agriculture and forest sectors (Olsson et al., 2016; Prabhakar et al., 2012).

With high spatial resolution and short revisit frequency, Copernicus Sentinel-2 satellites are a good source of images for deriving NDVI over large areas (Haghighian et al., 2020).

Pettorelli et al. (2005), exhaustively described the characteristics of NDVI annual time-series for vegetated areas and how to derive NDVI indices, which are linked to the vegetation phenology and/or productivity.

The yearly NDVI time-series of a vegetated area typically shows an increase in correspondence with the start of growing season, the NDVI peaks, establishing a plateau, and then eventually decreases when the vegetation senescence starts (Figure 1).

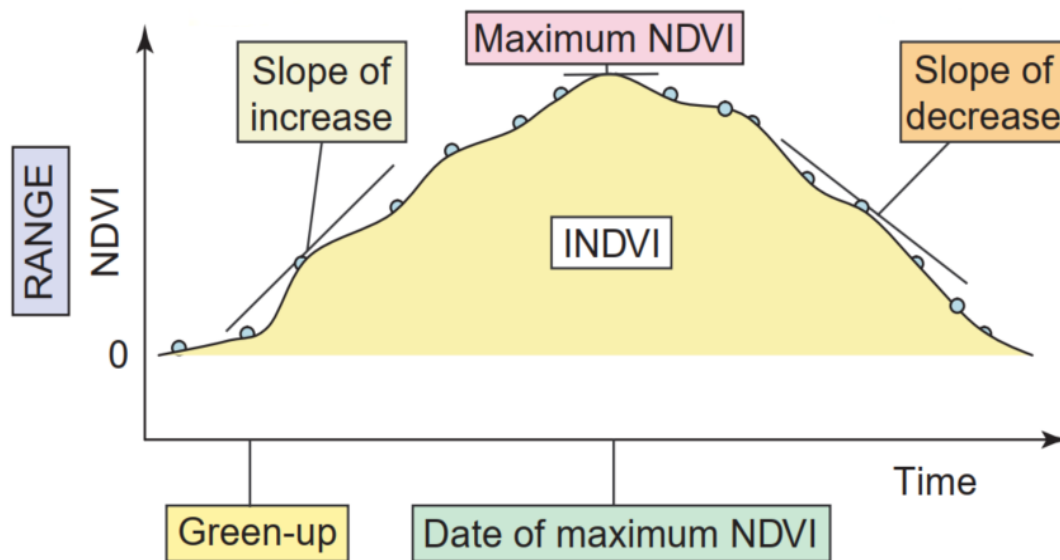


Figure 1 - Smoothed annual NDVI curve for vegetation. The NDVI values start rising since the beginning of the growing season, it reaches a maximum and then decreases in respondents to the start of the vegetation senescence. Adapted from Pettorelli et.al (2005)

Because of the residual noise in the NDVI data sets, mainly due to atmospheric conditions and clouds, the time-series needs to be gap-filled and smoothed through the application of different filtering algorithms. Shao et al. (2016), reported three different smoothers which can be applied in the TIMESAT software package namely i) adaptive Savitzky-Golay, ii) asymmetric Gaussian and iii) double-logistic function.

The Savitzky-Golay uses quadratic polynomial function to fit all points within a moving window of a given time-series dataset, and then the value of the central point is replaced by the fitted value.

The adaptive Savitzky-Golay reassign the weight for each point to favour points located above the initial polynomial fit, assuming that noises from clouds normally reduce the original NDVI value.

The asymmetric Gaussian algorithm relies mainly on five parameters to fit time-series data (i.e. Time of the min, or Max NDVI, width and flatness of both the right side and the left side of the function).

Eventually, the double logistic function estimates four parameters to determine the left inflection point, the right inflection point, and rates of changes at two inflection points.

An empirically based comparison of six selected NDVI time series smoothing techniques revealed the general superiority of the double logistic and asymmetric Gaussian function methods in terms of overall noise reduction and data integrity (Hird & Mcdermid, 2009).

According to the TIMESAT software manual (Eklundh & Jönsson, 2017), the asymmetric Gaussians or double logistic functions may be the better choice, compared to the adaptive Savitzky-Golay, when dealing with noisy time series.

After smoothing, a variety of NDVI-derived indices, or land surface phenology (LSP) metrics, can be derived from the curve.

Some LSP metrics are related to the overall vegetation productivity, such as the Integrated NDVI (INDVI) or the annual maximum NDVI value. They also include dates of the beginning (and the end) of the growing season, the timing of the annual maximum NDVI and the rate of increase and decrease of the NDVI. A list of the most important metrics, their features and biological meaning are presented in (Table 2) and graphically displayed in (Figure 1).

Table 2 - Examples of NDVI derived indices and their ecological meaning. Adapted from Pettorelli et.al (2005)

NDVI derived Index	Definition	Biological meaning
INDVI	Sum of positive NDVI over a given period	Annual production of vegetation
Annual maximum	Max value over a year	Annual production of vegetation
Time of the max	Time of Max	Timing max availability of vegetation
Beginning/End of the growing season	Dates estimated from moving average procedure	Start of the green-up/senescence

NDVI derived Indices were mainly used to assess the effects of climatic conditions on biomass and plants phenology. However, they were also applied to study trophic interactions (Pettorelli et al., 2005), to provide information for grassland management in protected areas(Weber et al., 2018) and for the prediction of crop phenology (Badr et al., 2015).

Due to their link with climatic and weather conditions, NDVI indices could provide additional information to predict the Olive Fruit Fly population for Area Wide Pest Management.

However, the correlation between the above-mentioned indices and the Olive Fruit Fly population density has not been studied yet.

The integration of Satellite data for the monitoring of this pest represents an opportunity to increase the effectiveness of area-wide monitoring programs while improving their time and cost efficiency.

1.3. Modelling pest risk using climatic and geographic data

Mathematical modelling is a widely adopted strategy to improve the knowledge on pests' population dynamics and achieve the early prediction of pest damages (Yan et al., 2015).

In this section, an overview of different modelling approaches, namely temperature-driven models, geostatistics, correlation analysis and Artificial Neural Networks, are presented. The main features, including different purposes, advantages, and disadvantages of each method and their relevance to this work is eventually discussed.

Temperature-driven models have been largely used to predict target events linked to the insect's phenology such as the occurrence of a development stage for certain pests.

Climate has a strong effect on the distribution and abundance of insects and so the mathematical description of such a relation has been of great interest among entomologists (Damos & Savopoulou-Soultani, 2012). Being temperature a factor of great influence, most of the models that describe insect development are temperature driven. Knowing this kind of relationship is a prerequisite for the prediction of timing and phenology of insects.

The application of temperature-driven model to study pest population dynamics can be found in several studies (Blum et al., 2015; Gutierrez et al., 2009; Petacchi et al., 2015).

Population modelling has been usually driven by temperature indices (e.g. Cumulated Degree Days-CDD) derived from weather-station data, which are then interpolated in order to obtain meaningful dataset, representative of large geographic areas. In case of limited availability of weather stations data, Land Surface Temperature (LST) data from satellites have also been used (Blum et al., 2015), even though their application so far was limited to homogeneous flat areas due to the coarse spatial resolution of LST maps (Malbêteau et al., 2017).

The application of temperature-driven models can become problematic when weather station's data are not available, and the rough topography of the area does not allow the use of LST data.

Geostatistical methods represent one category in the broader field of Spatial Interpolation Methods. They are a family of generalized least-squares regression algorithms, also known with the name of "Kriging" (Li & Heap, 2014).

Geostatistics can be a powerful tool to study spatial variation of insect populations and to support Integrated Pest Management, from field to regional level, it can be used to analyse spatial heterogeneity and optimize the use of chemicals, as well as to produce risk assessment maps (Sciarretta & Trematerra, 2014). Risk maps are important tools for land managers and other stakeholders to understand the distribution of the pest population in the interested area and support decision making (García-Chapeton et al., 2020).

The application of multi-variate geostatistics for the study of the OFF-population dynamics was given by Castrignanò et al., (2012), with a focus on a large olive growing area in Greece.

Data from a regional network of monitoring traps (700 McPhail traps) has been analysed by principal component analysis. Co-Kriging and factor Kriging, including elevation as independent variable. The model outputs were used to produce thematic maps and to define different priorities to monitoring zone for each of the selected periods. They found that OFF population density in summer was positively influenced by elevation, while the relation became less clear in October (patchy OFF distribution on the map). These results confirmed the findings of a previous research conducted in the same region (Kounatidis et al., 2008).

The geostatistics approach requires a large sample size to perform a meaningful variogram analysis (Sciarretta & Trematerra, 2014). As a rule-of-thumb, at least 50-100 samples are necessary to achieve a stable variogram; or at least 100, to produce a reliable estimation of the variogram (Li & Heap, 2014). Such sampling effort requires a considerable investment of both time and economic resources especially when it comes to large geographic areas.

Multiple regression is a linear regression method which models the relation between a set of independent variables, or predictors, and a target variable. It has been adopted for the evaluation of pest risks for different pests and crop types such as the cowpeas pests (Karungi et al., 2000) and the paddy stem borer in rice fields (Yang et al., 2008).

It is the simplest method for pest prediction and has the advantage that the statistical significance of every single selected predictor can be identified, and the results are easy to interpret.

The generalized approach consists in relating the target variable, to climatic and, or topographic factors (predictors)

Petacchi et al. (2015) used a temperature-driven model, kriging, and multiple regression to predict the expected day of emergence of overwintering adults of *Bractocera oleae*, in a mountainous region in Italy. First a CDD model was calibrated and validated using data collected in 3 experimental sites and historical data from extension service. Once validated, the model was run on 78 weather stations placed in the region. Spatial fits of model output were carried out using both ordinary Kriging and a regression correlation approach, the second one showing better results. The better accuracy was explained by the fact that the regression approach is independent from input data location and accounted for the rough topography of the site.

Elevation, (Euclidean) distance from the sea and aspect ($\cos \alpha$) were used as predictors to include all geographic factors influencing temperature. Elevation was found to be the best predictor however, the determination coefficient improved by increasing the number of independent variables confirming that other topographic factors have an influence on temperature.

Artificial Neural Networks are also a popular model approach for the prediction of crop pest risk (Yan et al., 2015). They are a family of techniques of artificial intelligence which were inspired by the computational mechanism of the human brain. The brain performs complex, non-linear, parallel computation and can self-organize and build its own “experience” (Mas & Flores, 2008). ANN’s present several advantages over conventional methods as they can account for any non-linear complex relationship between the predictors and the dependent variables and so do not require any assumptions about the used data set.

Yan et al. (2015) tested the use of Multiple Regression and Neural Network (NN) to predict the crop risk related to two different pests. The results shown that NN improved the predictions accuracy compared to the traditional approach however, the authors also stressed some disadvantages of NN models particularly:

- The importance of the single predictors cannot be identified as NN performs like a black box.
- The setup of the model parameters (e.g. model topology) is subjective.
- The model can run into overfitting issues in non-linear data training.

An application of NNs to the study the occurrence of the Olive Fruit Fly was developed by Kalamatianos et al. (2017), who tested different machine learning algorithms, for the prediction of future trap measurements. The model training and cross-validation was performed using two years trapping data from a small network of 16 traps, while the outputs of a CDD model, temperature data from sensors and the previous trapping measure (N flies in the previous visit) were used as predictors of the target variable. The NN experiment was set up as a classification problem (three different classes based on the intervention threshold), using a feed forward NN with one hidden layer and an increasing number of hidden nodes (1-15). The results shown that NN precision (% of correct classification) was comparable to other machine learning techniques and that the model performance increase when the CDD model outputs were included in the predictors. In another work (Kalamatianos et al., 2019) the same authors used weather sensors and trapping data, through a combination of Clustering algorithms and Neural Network based classification, to identify and localise microclimatic areas which have an influence on the OFF-life cycle.

The application of a variety of mathematical models to study the OFF-population dynamics in relation with climatic and topographic factors is well documented.

For the present research geostatistic methods were not considered due to resource constraints because of the high sampling effort they require to produce meaningful variograms (Sciarretta & Trematerra, 2014).

On the other hand, CDD models needs temperature data from sensors, or weather stations (Kalamatianos et al., 2017; Petacchi et al., 2015), which were not available at a sufficient spatial density

in the study area of this research. The alternative use of LST data was also excluded because their spatial resolution does not meet the requirements in relation to the rough topography of the region (Malbêteau et al., 2017).

In mountainous areas, topographic factors such as elevation, aspect, and the distance from large water bodies, play an important role in defining local climatic conditions and can be used as a proxy of temperature. The application of Neural Networks models to study the climatic influence on the *B. oleae* population has been a topic of interests in recent research works (Kalamatianos et al., 2017, 2019), but the relation between the OFF infestation and topographic factors using Neural Network regression has not been investigated yet. This is worth to be tested because it might provide a comparatively low-effort methodology, which could be easily replicated by local institutions without any major changes in their pest-monitoring protocols.

1.4. Aim of the study and research questions

The present research aimed at testing whether the use of NDVI derived metrics and spatial analysis can improve the efficiency and the effectiveness of AWPM programs of the Olive Fruit Fly in the Mountain Community of Sebino Bresciano (Italy).

A first exploratory investigation wanted to answer the question whether NDVI metrics from Sentinel-2 satellite data, could be used to enhance the information on the OFF-population density in the study area.

It is worth noting that, the Olive Fruit fly does not directly damage the tree vegetation. In fact, this work did not look for a correlation between the NDVI metrics and damages to the tree canopy but, the LSP indices were used as proxy of temperature to predict the pest population density.

The second part of this work tested the correlation between the main topographic factors influencing temperature, and the OFF population in the study area by a Neural Network regression model.

The aim of the spatial analysis was to produce an OFF risk map of the study area as a support tool for local pest monitoring and management programs.

2. Materials and methods

2.1. Study Area: Comunità Montana del Sebino Bresciano

The area of the present research corresponds to the administrative borders of the Comunità Montana del Sebino Bresciano) (Mountain Community of Brescian Sebino) (CMS), located in the Region of Lombardia in Northern Italy. The CMS is a local public body formed by nine municipalities counting approximately 35 000 inhabitants. The purposes of the institution are the valorisation of alpine areas, the support and coordination of the administrative tasks, the implementation of programs and initiatives in the agriculture, craftsmanship, culture and tourism sectors (Comunità Montana del Sebino Bresciano, 2021).

The CMS Office of agriculture and forest, provides technical support to the olive oil producers through different activities and initiatives. The idea of the present research and its design was developed in close collaboration with the technicians of the “Forest and agriculture Office” to meet their needs to improve AWPM of the OFF in the CMS area.

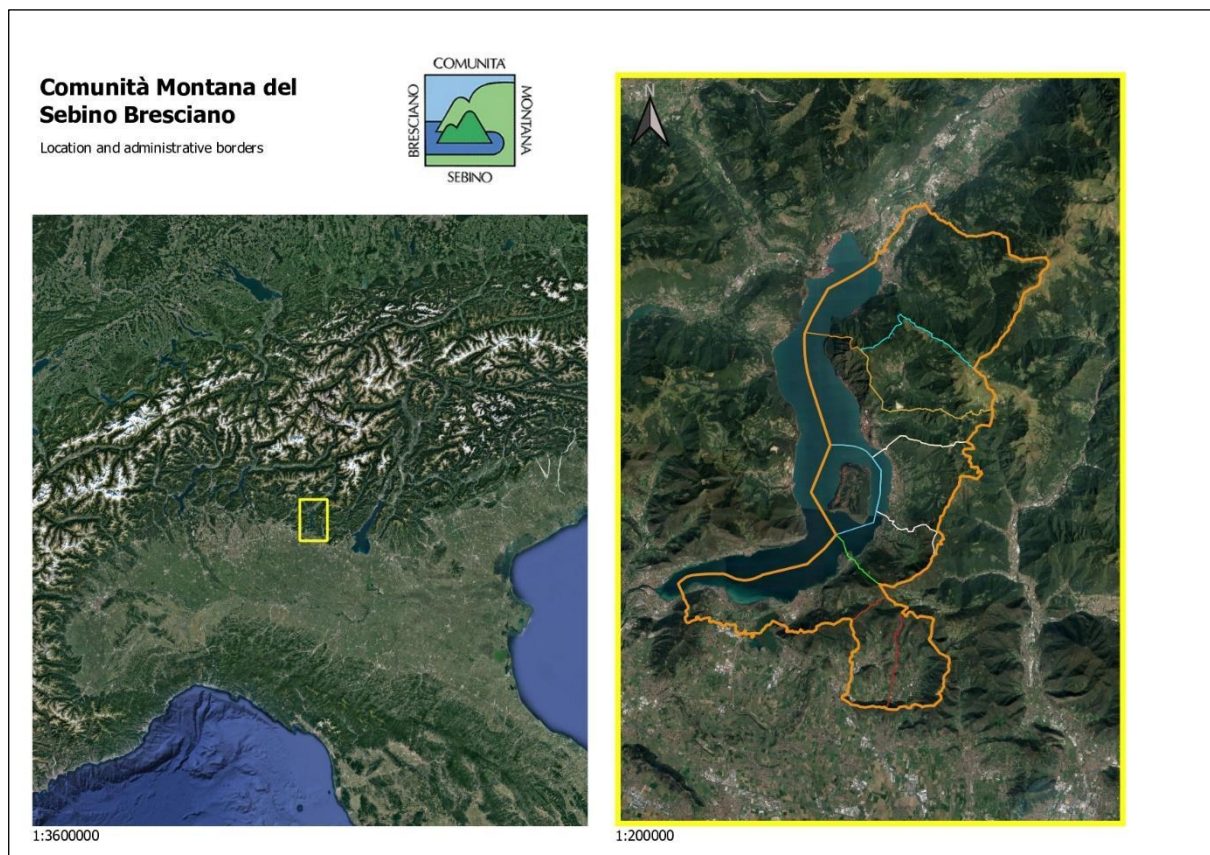


Figure 2 - Geolocation and the administrative borders of the study area. It is possible to distinguish the 9 municipalities composing the Comunità Montana del Sebino Bresciano.

The CMS cover a surface of 17826 ha of which almost 3 500 ha occupied by the Sebino lake, 14 000 ha considered as mountainous area and the remaining classified as hilly area (Figure 2).

The highest peak in the area is the Monte Guglielmo with an altitude of 1948 m the island of Monte Isola, in central sector of the Sebino Lake, reaches an elevation of 600 m.

The main element influencing the climate, especially along the shore and the hilly areas in the surrounding, is the huge mass of water represented by the lake.

The lake mitigation effect makes the climate to have a temperate to hot-temperate characteristics up to 400 m elevation. Between 400 and 1000 m the climatic conditions get more and more Atlantic, the precipitation increase, and winter mean temperature decrease. Above 1500 m we encounter colder and longer winters, together with short summer characterized by heavy rainfall (Bontempi, 2016).

2.2. The olive oil sector and classification of the olive groves

The mitigation effect caused by the lake, creates a microclimate which allows the olive trees to grow well in the CMS, close to the extreme northern border for the cultivation for this species. Anyway, in the recent years the olive cultivation has spread towards other areas which are not directly influenced by lake such as the hills south of the Sebino lake.

The traditional cultivars, which are still the most widespread in the area, are namely the Frantoio, Sbresa, Casaliva, Leccino and Pendolino (Rolfi, 2003). All those varieties are mainly used to produce olive oil, while the production of table olives is rare. The elevation range of the cultivation goes from the lake level (187 m) up to around 600 m. Above this threshold low temperatures, become a limiting factor for the development of the species.

The classification of olive groves was done manually using the software QGIS version 3.14 (QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>), by visual interpretation of orthophotos taken in the period from June to September 2015 (50x50cm). The orthophotos are available for public online consultation from the geodatabase of Region Lombardia (Agea, 2015). The classification was done using a nominal scale of 1:2000 and a minimum of 10 olives trees has been set as the lowest threshold for an olive grove to be considered. A total area of 243,8 has was classified as olive cultivation in the whole CMS. It is worth to say that field surveys for the correction of doubtful parcels are still in progress and so the area might vary slightly after this verification. The olive groves classification map is displayed in (Figure 3).

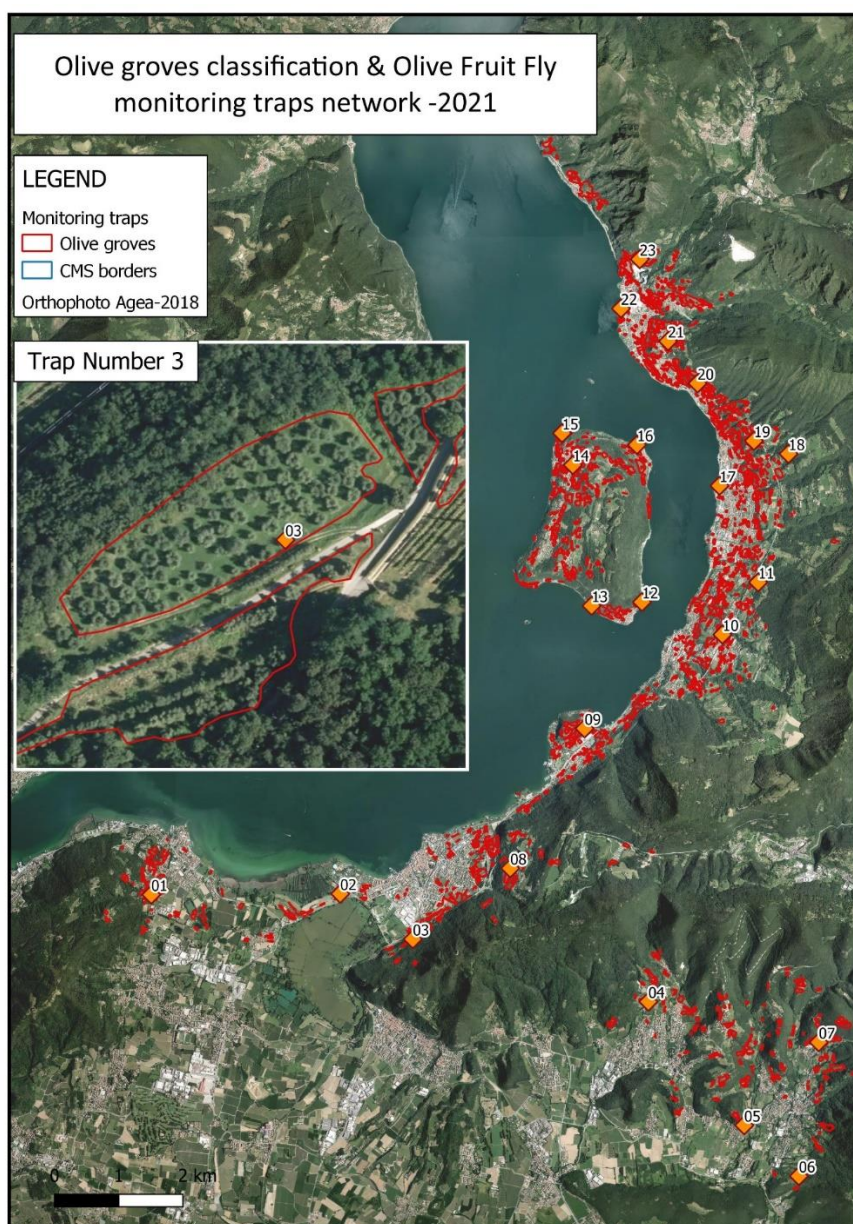


Figure 3 – Map showing the Classified olive groves in the CMS area and the location of the OFF-monitoring traps used in this research. A zoom-in to trap number 3 is shown to provide an example of a sample olive grove.

2.3. Sampling and data Collection

The CMS office of Agriculture since 2020 started a monitoring program of the OFF population in the area, setting a network of 9 traps in the CMS area (1 per each municipality) which have been monitored weekly starting from 30/07/2020 until the 17/09/2020. For the season 2021 the number of trapping stations was increased to 23 (1 trap/point). The location of the traps was selected to be representative of the different geographic conditions influencing temperature.

The sample size was defined in accordance with budget and time constraints while the location of the monitoring stations was limited by the willingness of the producers to give the permission to install the trap and to access their fields.

Monitoring was done using “WING TRAPS” which are star-shaped sticky chromotropic traps where the yellow colour attracts the target insects especially *Diptera tephritidae* (Figure 4a, 4b).

The traps were triggered by both a sexual pheromone attracting males, and ammonia baits (attracting both males and females).

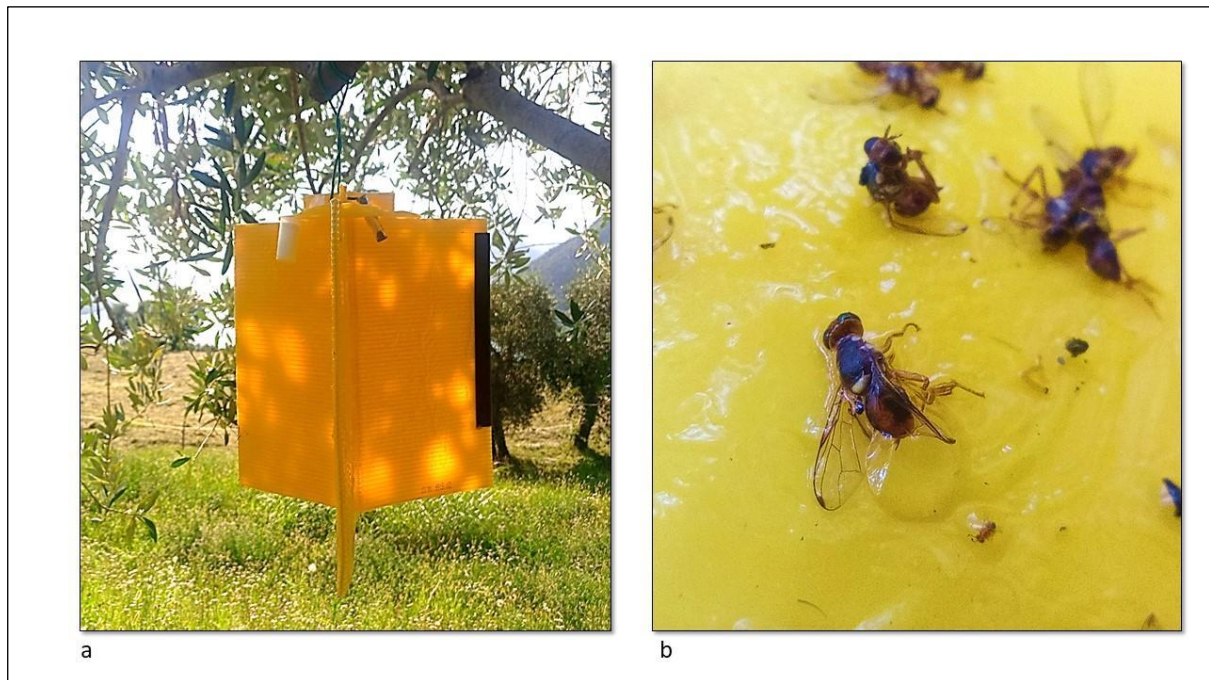


Figure 4 -a) One of the wing traps used for the monitoring of the OFF population. It is possible to see the ammonia and the pheromones baits in the upper part of the trap. b) Males OFF adults captured by a trap.

The traps were put in place between the 07/07/2021 and 09/07/2021. The number of adult's flies captured was monitored every two weeks for the first month, from 07/07/2021 to 06/08/2021, and then weekly in the period from 07/08/2021 to 30/09/2021.

The pheromones and the ammonia bait were replaced every 4 weeks or, for the ammonia baits, every time the content looked empty according to the instructions given by the producer (Isagro, 2020). The OFF-monitoring network is displayed in (Figure 3)

2.4. NDVI metrics as predictors of the OFF-population density

Drusch et al. (2012) provided an extensive description of Copernicus Sentinel-2 satellites, technical features, products, and possible applications. An overview of the satellite's main characteristics is presented in the following paragraphs including a description of the different products.

Sentinel-2 is a constellation of two multispectral high-resolution satellites, orbiting at a mean altitude of 786 km, providing optical observations over the Earth's surface from -56° to 84° latitude.

The mission includes two identical satellites operating simultaneously at the opposite side of the orbit, making it possible to halve the revisit frequency down to 5 days/visits at the equator.

Multi spectral images are provided to the users in forms products of different Levels, according to the different processing operations implemented. Level 0 and Level 1 provide raw compressed and uncompressed data, respectively. Level 1B data are radiometrically corrected radiances while the Level 1C product provides geo-coded top of atmosphere (TOA) reflectance, a sub-pixel multi-spectral and multi-date registration. A cloud and land/water mask are associated with the product. The L1C unitary product is a tile of 100×100 km². An additional atmospheric correction and enhanced cloud masks are applied to Level 2A products which represent the highest level of processed images provided to the users. In this work only L2A products were used as base images to perform any further analysis.

Sentinel-2 satellites are equipped with a Multi Spectral Instrument (MSI) with 13 spectral bands (Table 3) spanning from the visible and the near infrared to the short-wave infrared.

Table 3 - Sentinel 2 Bands and Spatial resolution (ESA, 2015)

Sentinel-2 Bands	Central Wavelength (nm)	Resolution (m)
Band 1 - Coastal aerosol	443	60
Band 2 - Blue	490	10
Band 3 - Green	560	10
Band 4 - Red	665	10
Band 5 - Vegetation Red Edge	705	20
Band 6 - Vegetation Red Edge	740	20
Band 7 - Vegetation Red Edge	783	20
Band 8 - NIR	842	10
Band 8A - Vegetation Red Edge	865	20
Band 9 - Water vapor	945	60
Band 10 - SWIR - Cirrus	1375	60
Band 11 - SWIR	1610	20
Band 12 - SWIR	2190	20

The spatial resolution varies from 10 m, for the visible and n-IR bands, to 60 m depending on the spectral band while the swath is of 290 km.

Even though the Sentinel-2 mission does not provide ready to use NDVI products, the index can be easily computed, considering the respective band's wavelengths, by the Equation 1:

$$NDVI = \frac{NIR - RED}{NIR + RED} = \frac{B8 - B4}{B8 + B4} \quad (2)$$

NDVI time-series can be extracted from Sentinel-2 imagery by using different software's and web platforms. The planned methods for the extraction and further analysis of the NDVI derived indices consists of 4 steps:

1. Creation of a Shapefile containing the polygons of 22 sample olive groves using QGIS;
2. Extraction of the NDVI-timeseries for the sample polygons in Google Earth Engine (GEE);
3. Extraction of the seasonality parameters from the time-series by TIMESAT;
4. Neural Network regression using NDVI metrics and data from the monitoring traps in RStudio;

The four steps are described in detail in the following paragraphs while a schematic representation of the planned procedure is shown in Figure 5.

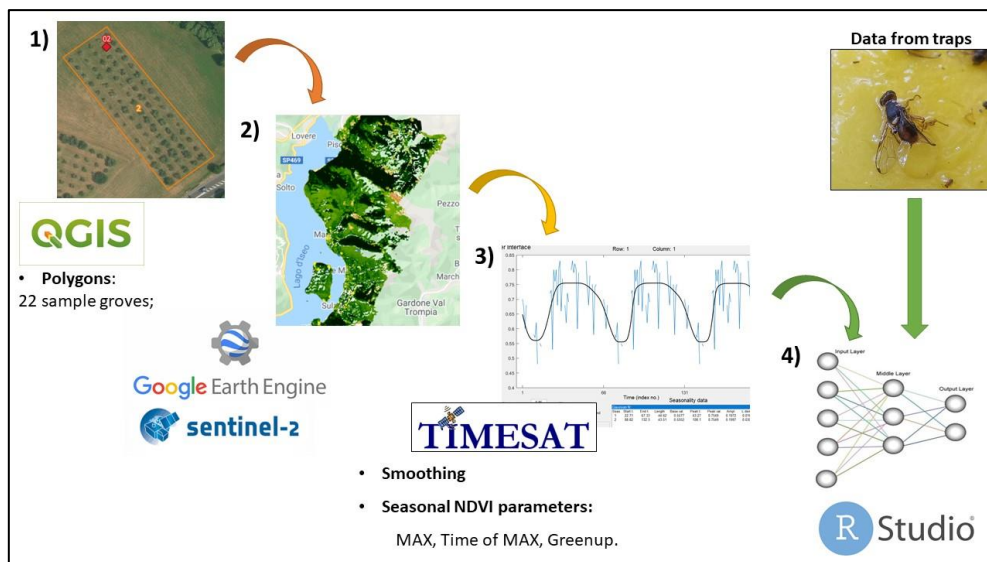


Figure 5- Schematic representation of the four steps for the planned methodology. Extraction of the NDVI time series for the sample groves (1-2) and seasonality parameters (3). Neural Network regression including OFF trapping data (4).

In this research the median NDVI time-series, from Sentinel-2 images, for the 22 sample olive groves was obtained using the web platform GEE.

GEE is a platform for scientific analysis and visualization of geospatial datasets, it hosts satellite imagery and stores it in a public data archive. The platform provides an application programming interface (API), an associated interactive development environment (IDE) and other tools to enable the analysis of large datasets and the visualization of the results (Gorelick et al., 2017).

The data catalogue is continuously updated and houses a large repository of publicly available geospatial datasets, including observations from a variety of satellites (e.g. Landsat, Sentinel-1 and Sentinel-2) in both optical and non-optical wavelengths. The images from the catalogue are grouped into “collections” which can be easily sorted and filtered to meet specific spatial, temporal or other criteria (e.g. Date range and Cloud cover) (Gorelick et al., 2017).

A shapefile containing the polygons of the 22 sampling groves was created using the software QGIS and then uploaded in the GEE platform. A Sentinel-2 image collection from 01-01-2021 to 01-10-2021 was uploaded and cloud-masking function was applied to exclude Cloudy pixels from the analysis and reduce the related-noise.

The NDVI band was added to the Collection and, because the olive groves were larger than the pixel size of the images (10x10 m), the median NDVIs were computed to reduce them to a single NDVI value per each sample grove. The resulting time series was exported in a “.CSV” format.

The same procedure described in the previous paragraph was used to obtain the median NDVI time-series for the period 01-10-2019 to the 31-12-2019 which was lately joint to the data from 2021 to create an artificial time-series of one year. The reason for that was to create an ASCII readable format for the TIMESAT software as described in the following section.

The TIMESAT program is primarily designed for the analysis of time-series of vegetation indices satellite data and the extraction of seasonality parameters.

A complete description of the 3.3 version of TIMESAT, used in this research, can be found in the software manual (Eklundh & Jönsson, 2017), including the main technical features and applications.

The software offers three different smoothers namely the adaptive Savitzky-Golay filtering method and methods based on upper envelope weighted fits to asymmetric Gaussian and double logistic model functions

Data organized as ASCII files or images (two-dimensional spatial arrays) can both be handled by the system. To be read in TIMESAT the first line of the ASCII file must contain the information about the number of years spanned by the time-series, the number of data values per year, and the number of time-series in the file. The data structure of the ASCII file is displayed in Figure 6.

n year	n values/year	n ts	
y1	y2	...	yN
y1	y2	...	yN
y1	y2	...	yN
...
y1	y2	...	yN

} n ts

Figure 6 - Data structure of ASCII files. The first line includes the number of years, the number of data per year and the number of time series (nts). The time series are then given line by line (y)(Eklundh & Jönsson, 2017).

In order to extract the seasonality parameters for the 2021 in TIMESAT at the time of this research (October 2021) it was necessary to integrate them with data from another year (2019) to complete the one-year time-series to be readable by the program Figure 7. This artifact would have no influence on the seasonality parameters which relates to the left side of the curve (max NDVI, time of the max, start of the growing season).

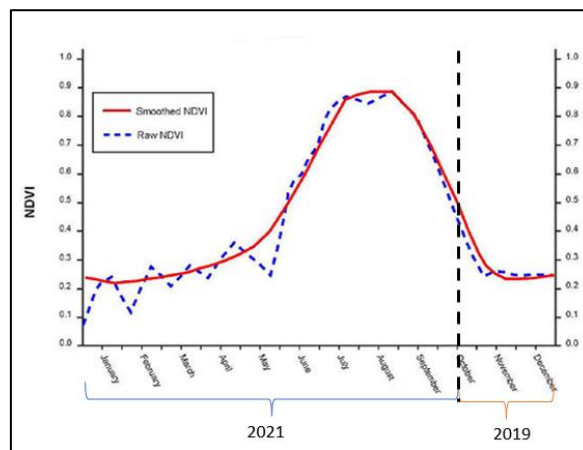


Figure 7 – A hypothetical schematic representation of the 1 year artificial time-series.

The steps of the processing in TIMESAT are summarized below:

- a. Read time-series $y_1; y_2; \dots y_N$.
- b. Fit a smooth function.
- c. Use fitted function to extract seasonality parameters.
- d. Write seasonality parameters to file

Following the extraction of the aforementioned parameters, the planned step was to build a Neural Network model using the extracted NDVI metrics as predictors of the OFF-population density.

The Neural Network regression can be performed through a so-called multi-layer perceptron.

The basic idea and mathematics of NN multi-layer perceptron is briefly presented by Yan et al. (2015), while Mas & Flores (2008) provided a deep dive into Neural Networks models and their application to remote sensing data.

A multi-layer perceptron (MLP) is composed of an input layer with input neurons (i.e., predictors), one or more hidden layers of computation neurons and an output layer (i.e., target variable) (Figure 8)

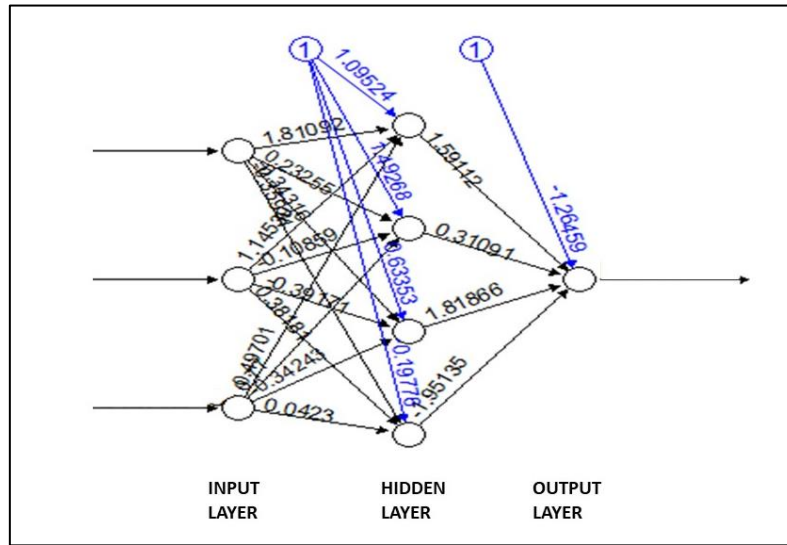


Figure 8 –Multi-layer perceptron Neural Network with one hidden layer. Black numbers represent the weights while blue numbers are the so-called Bias representing the intercept.

The hidden layers can detect and learn the relationship between the predictors and the dependent variable, no matter if such a relation is linear and non-linear.

The model training process occurs in two phases: *feedforward* phase and *backpropagation* phase.

In the *feedforward* phase, a training input vector is first given to the input layer and then propagated forward until the output vector is generated in the output layer. The neurons in one layer are fully interconnected with all the neurons in the next layer. The strength of connection between two neurons from adjacent layers is named *weight*.

In the *backpropagation* phase, the basic idea is to minimize the error through iterative backward propagation of error signals. The connection weights are adjusted recurrently until the cost function is minimized (i.e., one consecutive step with no decrease in error).

The cost function is defined as (Equation 2):

$$E = \frac{1}{2} \sum_{k=1}^l (y_{d,k} - y_k)^2 \quad (3)$$

where $y_{d,k}$ and y_k are the desired and actual output of neuron k in an output layer, respectively and l is the total number of neurons in the output layer.

When the value of the cost function reaches a pre-defined threshold value, a network is considered to have converged.

2.5. Spatial analysis of the OFF-population density in relation to geographic factors

This part of the research aimed at testing the correlation between the OFF-population density and the climatic conditions in the area of the CMS. Due to the lack of weather stations at a sufficient spatial resolution and the limitation for the use of LST data from satellite in an area with rough topography, the main geographic factors influencing temperature were used as independent variables for the analysis. A Neural Network regression was selected as the most suitable model to account for any non-linear complex relationships among the selected variables and the pest population density.

Elevation, Aspect (namely the direction, based on the azimuth, to which a surface slope faces) and the distance from the lake coastline, were selected as predictors of the OFF infestation, and their values for each of the sample stations were computed using open-source software QGIS version 3.14.16.

Elevation data were taken from the Digital Terrain Model (DTM) 5x5 m of the CMS area (Geoportale Regione Lombardia, 2015)

The spatial distribution of the aspect was calculated using the “Aspect” geoprocessing tool, starting from the aforementioned DTM. The raster of the $\cos(\text{aspect})$ was obtained using the raster calculator. The cosine of the aspect was computed to obtain a linear variable having values of 1 for North-facing slopes and -1 for South-facing slopes.

The distance from the coastline was calculated using the GRASS “r.grow distance” tool, available in QGIS, starting from the coastline shapefile from the regional geodatabase, setting spatial resolution of 100x100 m. Figure 9 shows the spatial distribution of the three dependent variables in the area of the research.

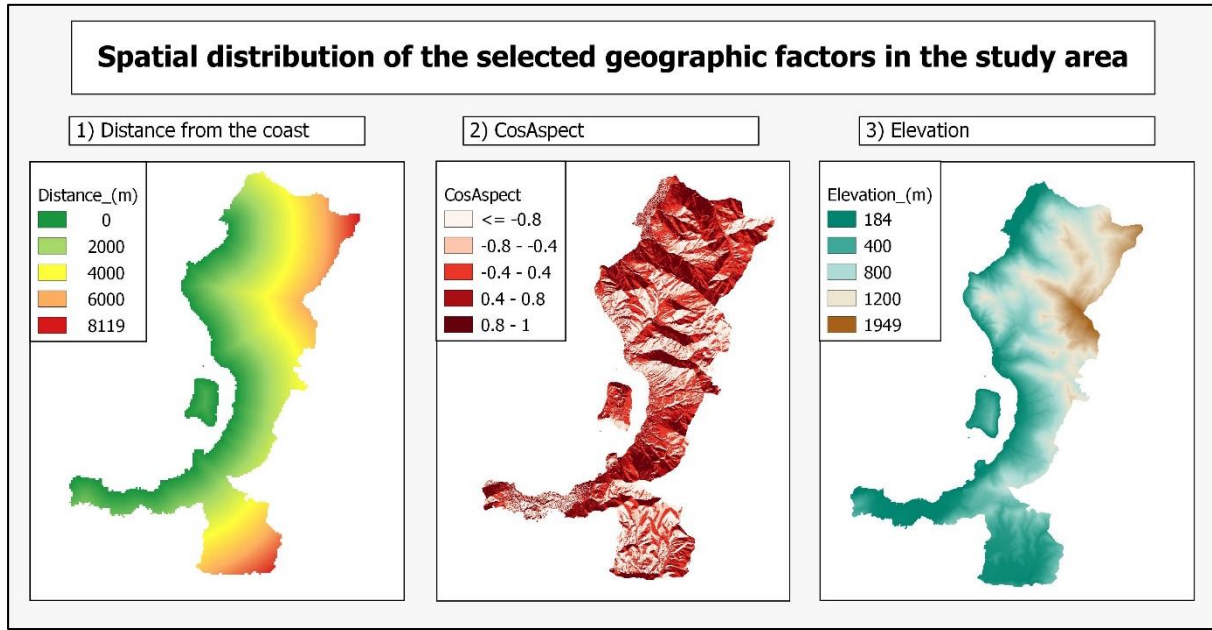


Figure 9 - Spatial distribution of the predictors in the CMS: Distance from the lake (1), Cos of the Aspect (2) and Elevation (3).

The Neural Network used to perform the spatial analysis is a multi-layer perceptron (described in section 2.4). Before starting the training of the Neural Network data of both the dependent variable and the predictors were normalized into a range between [0,1]. Data normalization had the goal to avoid weights overestimation (as the predictors have different units of measurement) and to improve the computation efficiency (Jayalakshmi & Santhakumaran, 2011).

The normalization was performed through a min-max equations as the following (Equation 3):

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

A Leave One Out Cross-Validation (LOOCV) was selected as the method for the selection of the most suitable architecture based on the lowest mean error from the cross-validation. LOOCV is a special case of k-fold cross validation, in which the number of folds equals the sample size, it is a sensible model selection criterion as it provides an almost unbiased estimate of the true generalisation ability of the model (Cawley & Talbot, 2004).

In this research the model selection was limited to a simple network having only one hidden layer, following the principle that this is sufficient to model any piecewise continuous function (Hornik et al., 1989). Ten different architectures with an increasing number of hidden neurons, from 1 to 10, were used in the LOOCV to test the effect of an increasing complexity on the model accuracy.

Models with more than 10 Neurons were not considered because it would increase the possibility to run into overfitting problems (Yan et al., 2015).

After the validation phase the model topology showing the lowest error was selected and used to draw a map of the predicted OFF density in the whole study area.

This creation of the aforementioned raster map was performed using the **raster** and the **rgdal** packages in RStudio. The three Input rasters (i.e. Elevation, Cos Aspect and Distance) were first resampled to get the same spatial resolution (100x100m) and then given as Input to the Neural Network model to generate an Output raster showing the predicted values for the Target variable.

3. Results and discussion

To get a first overview of the temporal and spatial distribution of the target pest in the study area, the trapping data for the 23 sample stations were plotted against the time of the monitoring period.

The average trapping load line was added to the plot to enable a quick visual comparison between the trapping load for single station and the average values (Figure 10) (Appendix A).

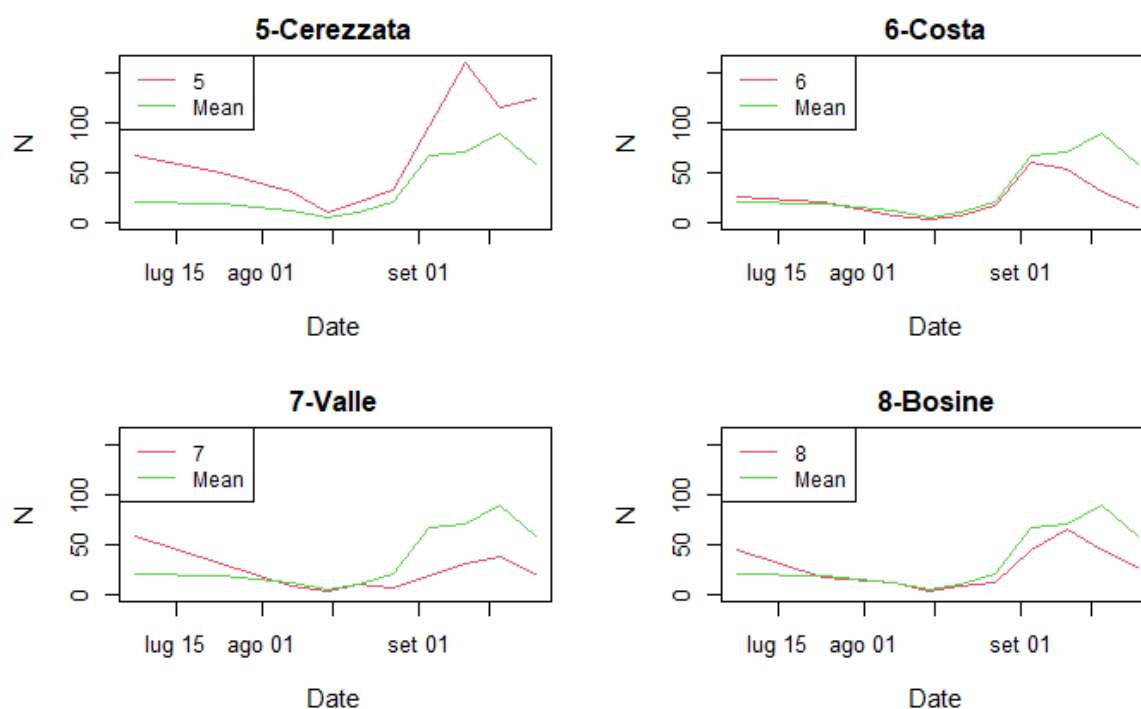


Figure 10 – Line plots showing the number of adults OFF per each monitoring visit (N) from 07-2021 to 10-2021, for four sample locations (red lines). The Green line shows the average trapping load among the 23 sample stations.

From the plots it was possible to notice that the absolute N of captures varied widely from one station to the other while the temporal trend was similar among the different locations.

Especially an increase in the number of captures was recorded in the month of September for all traps in accordance with the expected period of the second generation of the insect.

Being the OFF second generation the one causing most of the damages to the olive production (Delrio & Lentini, 2016), and September the month in which the intervention threshold is generally exceeded (Guidotti et al., 2003), the mean trapping load for the month of September was selected as Target variable to perform all further analysis.

3.1. NDVI metrics as predictors of the OFF-population density

The Google Earth Engine (GEE) platform was used to extract the median NDVI time series from Sentinel-2 images for the 22 sample olive groves. The GEE script used to these operations is attached in the Appendix B.1.

The results were exported in a “.csv” file format and pre-processed in Microsoft Excel to obtain a TIMESAT readable ASCII format:

- A regular time interval (5-7 days) between each NDVI value was set.
- NDVI values outside the [0,1] range were excluded.
- Data from 2021 were joint to the NDVI values from 2019 to obtain the artificial one-year NDVI time series.
- The artificial time-series was then multiplied to simulate three years of data.
- A first row including the TIMESAT required parameters was added (Table 4).

Table 4 - NDVI median values for the 22 sample grove for the month of January 2021. The first row includes the number of years (3), number of observations per year (66) and the number of time-series (22). -9999 values are missing values or values outside the [0-1] range which were excluded from the analysis.

3	66	22		
0.703794	0.67392	0.603283	0.676514	-9999
0.768182	0.76607	0.641635	0.76373	-9999
0.672913	0.992883	0.352641	0.994048	-9999
0.707544	0.672194	0.510167	0.717647	-9999
0.724568	0.742323	0.558532	0.741124	-9999
0.687853	0.669697	-9999	0.679579	-9999
0.781503	0.773573	0.715742	0.763098	-9999
0.632	0.669031	0.571299	0.6648	-9999
0.724844	0.688199	0.623637	0.750679	-9999
0.614887	0.726084	-9999	0.804671	-9999
0.576239	0.791667	0.649378	0.719417	-9999
0.649617	0.695798	0.573748	0.683673	-9999
0.665008	0.647704	0.545455	0.688401	-9999
0.648248	0.643532	0.560976	0.628541	-9999
0.712655	0.711103	0.636364	0.722346	-9999
0.700913	0.73185	0.57927	0.72005	-9999
0.638982	0.786262	0.44357	0.830308	-9999
0.762306	0.784413	0.61003	0.781825	-9999
0.71507	0.710454	0.539809	0.675456	-9999
0.781761	0.78892	0.595665	0.785276	-9999
0.672831	0.677115	-9999	0.662709	-9999
0.721107	0.708025	-9999	0.749356	-9999

The resulting file was then used in TIMESAT to produce the NDVI time series and then extract the seasonality parameters.

Because the raw NDVI data appeared to be still very noisy, the double logistic function was selected to smooth the curve. The polygons of all the 22 sample groves and the respective smoothed NDVI curves are displayed in Appendix C, while only the first six samples are shown in figure 11 as a support to the discussion.

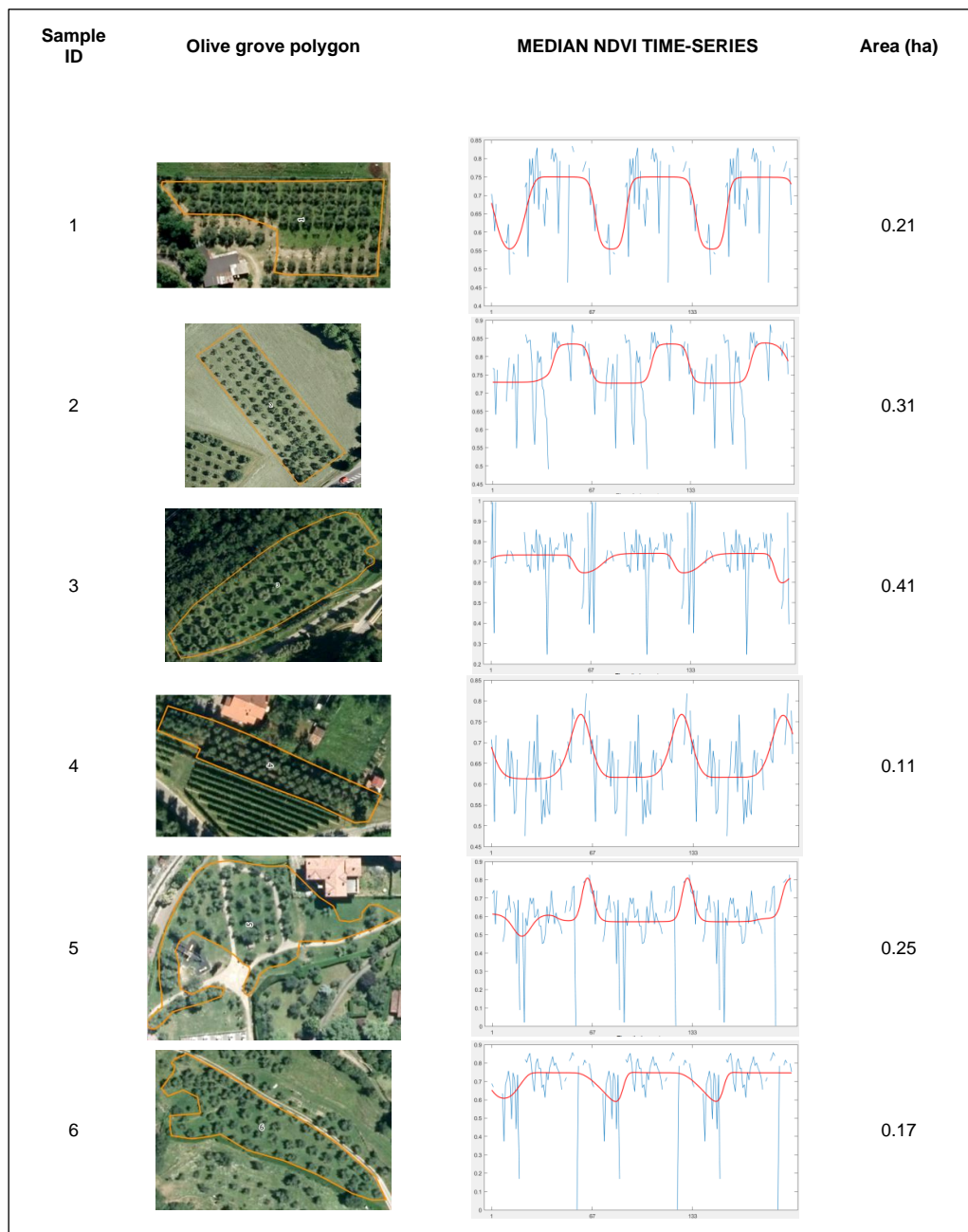


Figure 11 - Sample groves 1-6 and respective NDVI curves from TIMESAT. The red line showing the double logistic smoothed curve.

The smoothed NDVI curves shown an irregular shape which in some cases did not even follow the typical NDVI seasonal trend for vegetation (e.g. Sample grove n 5-8-10-21-22).

Comparing the graphs of the different groves, they had very different shapes (e.g. Sample groves n 11-10) making it impossible to detect any typical NDVI trend for olive tree plantations.

Due to the aforementioned anomalies, detected through a visual assessment, it was not possible to extract meaningful and reliable seasonality parameters for the sample groves. This is why the relation between NDVI derived metrics and the OFF infestation could not be tested as it was planned at the beginning of the present work.

From the visual analysis, one could observe that the application of the cloud mask and the exclusion of negative NDVI values, did not reduce the signal-noise sufficiently, and many low NDVI values were still presents, disturbing the signal (e.g. Sample groves n 8-17-21).

If the noise caused by atmospheric conditions is a common problem regardless of the size of the field, the geometric positional accuracy of the image data influences the reliability of time series especially for small size agricultural parcels and for fields with an irregular geometry (Vajsová et al., 2020).

In a study focusing on the ability of Sentinel-2 images to monitor vine growth across the year, Devaux et al. (2019) stress that the 10 m spatial resolution of S2 images represents a strong limitation in the case of small fields or fields with complex boundaries.

Among the 22 sample olive groves used for this research, the 91% had a size below the 0.5 ha while 68% were under 0.3 ha. This suggest that the size of the olive groves might have been a factor negatively influencing the reliability of the NDVI data from Sentinel-2 images.



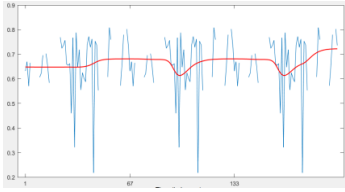

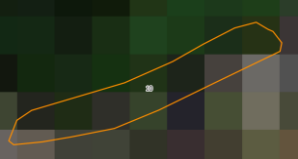
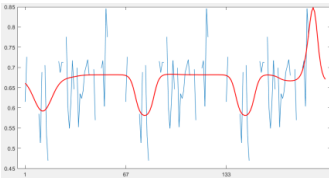


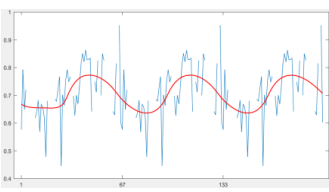


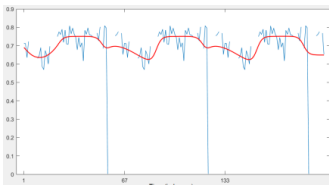


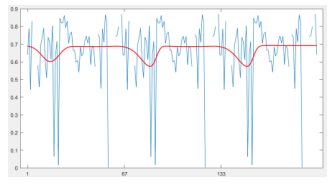


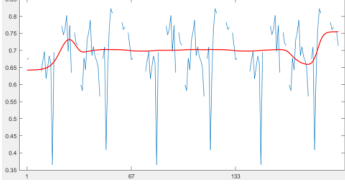
In addition to that, the irregular topography of the study area makes it very common to have olive plantations characterized by an irregular shape and a narrow-elongated geometry (e.g. sample groves n 4-10- 13-19) which might have a negative influence on the geometric positional accuracy of the fields.

A third factor of disturbance, which could be detected from the visual assessment of the orthophotos of the area, might be the presence of different crops (e.g. sample groves n 4-2-10), buildings (e.g. n 11-8-5) or water (e.g. n 16-19) in the surroundings of the sample parcels.

Especially in case non-vegetation pixels in the surroundings of the parcels, such as urban areas or water which are characterized by low or negative NDVI values (Viana et al., 2019), are included in the sample polygon, they can have a strong effect on the resulting median value of the olive grove.

Vajsová et al. (2020) proposed a possible approach to reduce the noise caused by heterogeneous pixels on the borders of parcels by introducing a 5m negative buffer, to isolate the clean pixels inside the parcels. The downside of this method is that very small blocks might disappear after the application of the negative buffer. Table 5 shows examples of possible factors of disturbances.

Table 5 - View of a selection of sample groves from Orthophoto and S2 image, and relative NDVI curves.

ID	SAMPLE OLIVE GROVES (Orthophoto)	SAMPLE OLIVE GROVES (Sentinel-2-True colors composite)	MEDIAN NDVI TIME-SERIES	AREA (ha)	POSSIBLE ERROR SOURCES
8				0.18	<ul style="list-style-type: none"> • Small size • Proximity to non-vegetation pixels • Irregular geometry
10				0.07	
11				0.5	
15				0.23	
17				0.46	
21				0.35	

From a comparison between the size of the sample olive groves and the size of the total olive groves obtained from the classification of the study area, it seems that the selected sample is quite representative of the global situation in the region (Table 6). Considering only the polygons with an area equal or greater than 0.05 ha, one could observe that the mean surface for the sample groves and the total is very similar, being 0,27 ha and 0,31 ha respectively. Moreover, the percentage of the total olive plantations under 0.3 ha in the area is 79%, compared to the 68% of the sample fields. Even though the accuracy of the classification over the whole area was not tested, these data suggests that the general context of the olive sector is characterized by a great number of small olive orchards.

Table 6 -Comparison of the sample olive groves with the total.

Olive groves	N (≥ 0.05 ha)	Mean area (ha)	% (≤ 0.3 ha)
Sample	22	0.27	68
Total	1060	0.31	79

These findings lead to the conclusion that Sentinel-2 data are possibly not suitable to extract NDVI time-series for such small and irregular plots while the exclusion of small olive groves would lead to a misrepresentation of the global context in the area.

Nevertheless, the proposed methodology could find an interesting application and should be tested in area characterized by a more regular topography and extensive olive plantations.

3.2. Spatial analysis of the OFF-population density in relation to geographic factors

Table 7 shows the descriptive statistics of the data set used for the analysis namely the mean trapping load for the month of September from the 23 locations (i.e. Target Variable) and the respective geographic factors (i.e. Elevation, Cos aspect and Distance from the coast).

Table 7 - Descriptive statistics of the data set used for the spatial analysis

Mean trapping load September (N of flies)	Elevation (m)	Distance from the coast (m)	Cos aspect
Min. : 16.00	Min. :187.0	Min. : 10	Min. :-0.9900
1st Qu.: 39.00	1st Qu.:190.5	1st Qu.: 100	1st Qu.: -0.6250

Median : 69.00	Median :278.0	Median : 500	Median : 0.1800
Mean : 71.96	Mean :273.6	Mean :1321	Mean : 0.1135
3rd Qu.:102.50	3rd Qu.:322.5	3rd Qu.: 898	3rd Qu.: 0.9150
Max. :150.00	Max. :482.0	Max. :7149	Max. : 1.0000

It is worth to observe that the target variable shown a great variability, ranging from 16-150 adults flies, and that even the minimum trapping load value would exceed the intervention threshold found in literature (Castrignano et al., 2012; Kalamatianos et al., 2017). These first results showing the high OFF infestation reached in the area during the examined period. The target variable was then plotted against the predictors to get a first impression on the existing correlation between them (Figure 12).

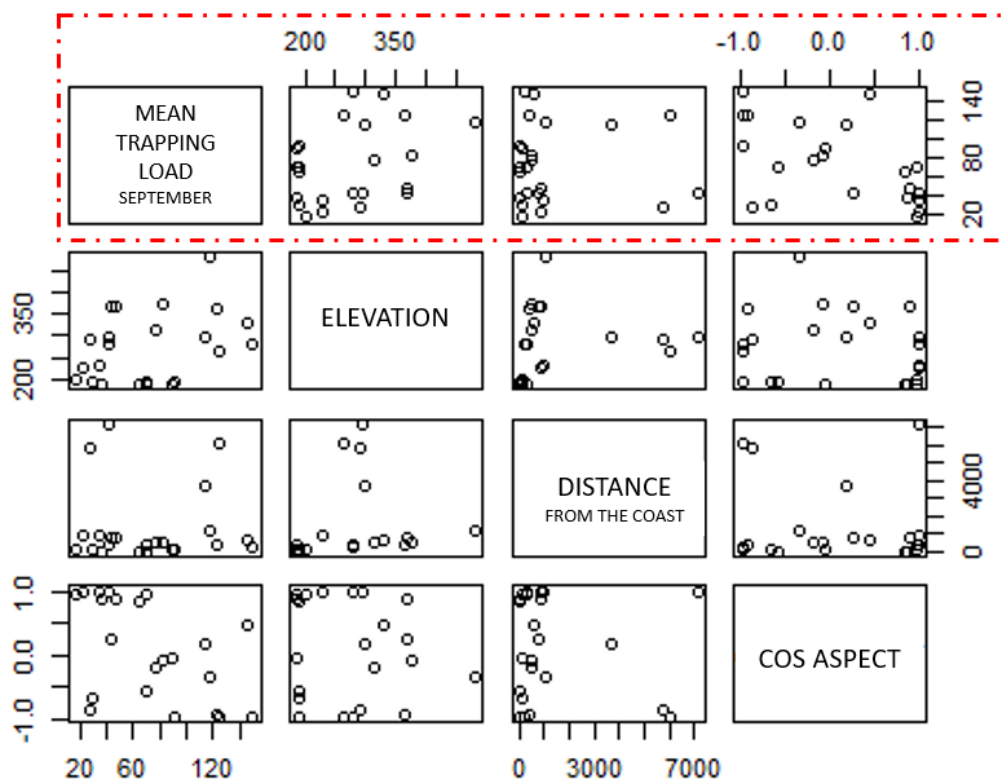


Figure 12 -Scatter plots showing the relation between the target variable and the predictors (red box) and the relation between each of the dependent variables.

From the visual analysis of the plots there was no evident linear relation between the OFF infestation and the three geographic variables in fact the points appeared to be randomly distributed.

To get a statistic measure of the strength of linearity the Pearson's correlation coefficient was also computed (Table 8).

Table 8 - Pearson's correlation coefficients between the Mean trapping load of September and the selected geographic variables.

PEARSON' S CORRELATION COEFFICIENT

Elevation	Distance	Cosaspect
0.351	-0.003	-0.531

The resulting Pearson's coefficients values shown a moderate positive linear correlation with the Elevation, a weak negative relation with the Distance from the Coastline and a moderate negative relation with the Cosine of the Aspect (Ratner, 2009).

Because from the first analysis of the data there was no evidence of an existing strong linearity, the Neural Network regression method was selected to account for any possible non-linear complex relation among the target and the geographic variables (Mas & Flores, 2008).

After data normalization a Leave One Out Cross Validation for the three-layers perceptron with an increasing number of hidden Neurons (NN1-NN10) was implemented in RStudio (Appendix B.2)

The average Sum of Squared Error (SSE) calculated using the error cost function (Equation 3), and the Mean Absolute Error (MAE) were computed for each iteration of the LOOCV process. The resulting errors were then averaged (Table 9) and used as the criterion for the selection of the model topology.

Table 9 - Average Sum of Squared Error and Mean Absolute Error resulted from the Leave One Out Cross-Validation of the different model topologies (1-10 hidden neurons)

LOOCV ERRORS MEAN		
N HIDDEN NEURONS	SSE	MAE
1	<u>0.0481</u>	<u>0.2340</u>
2	0.0876	0.3140
3	0.0514	0.2435
4	0.4100	0.4798
5	0.1281	0.4085
6	0.1774	0.3636
7	0.1136	0.3468
8	0.0910	0.3277
9	0.3757	0.5567
10	0.1759	0.4462

Neural Networks topologies with 1 and 3 Neurons respectively performed very similarly. However the best model accuracy was obtained by the simplest model (NN1), with only one hidden neuron (SSE=0.0481; MAE=0.2340).

Model topologies with 4 and 9 neurons shown much higher error compared to the rest of the models but generally, the variation of the errors in relation to the increasing complexity of the model architecture did not show any clear trend (Figure 13)

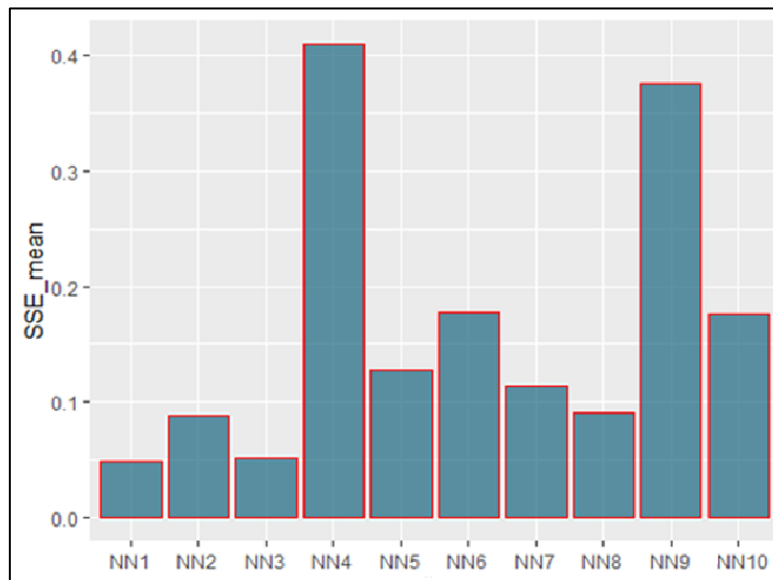


Figure 13 – Histogram showing the mean SSE resulting from the cross-validation. NN1 to NN10 represents the model's architectures with an increasing number of hidden neurons.

Discussing the NN architecture (for classification problems) Mas & Flores (2008) stressed that too simple networks topologies might not be able to interpret the internal structure of the data, resulting in lower accuracies, while too large Networks are likely to overfit the training data set

Nevertheless, from the results of the cross-validation the highest model accuracy, was obtained with the simplest possible model architecture. This might be related to the “simple” nature of our model (only 3 dependent variables) and the small size of the input data set.

The model NN1 (Figure 14), with one hidden neuron, was selected to produce the thematic map of the predicted trapping load for the month of September in the study area. To help the interpretation of the LOOCV resulting error for the selected model, the MAE was transformed back to the original units (47 Adults of Olive Fruit Flies).

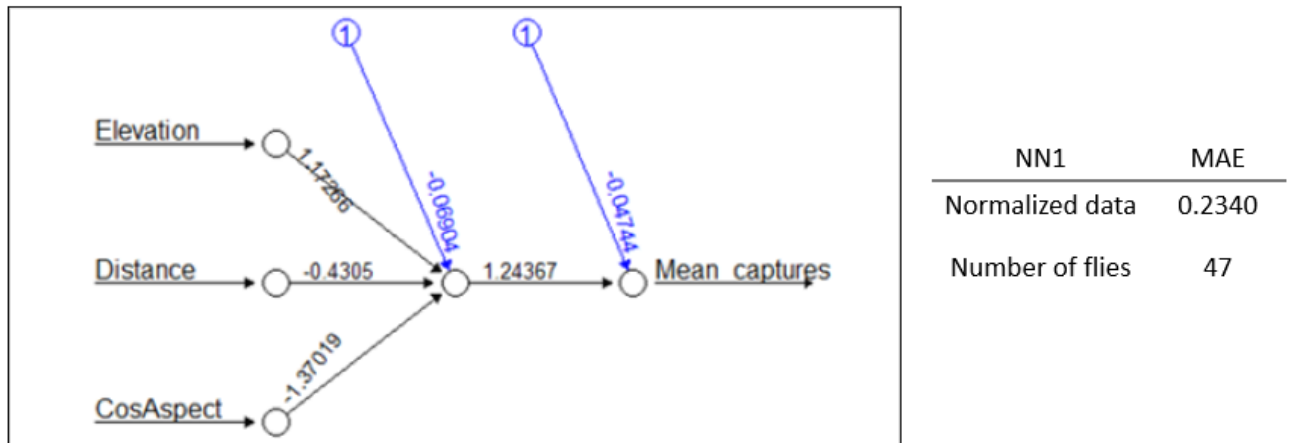


Figure 14 - Selected model topology (NN1) and the relative Mean Absolute Error (MAE).

When compared to the average trapping load of all sample stations, it resulted in a 65% error which might be problematic for practical applications.

However, the model needs to be tested on an external dataset before assessing its capability to predict the OFF-population density in the study area. An external data set was not available at the time of this research (trapping data from 2020 were not considered because they did not cover the entire month of September) and so it will be necessary to collect more data in the next years in order to test the model performance.

Even though the model needs to be further tested, a first version of the OFF-risk map was produced to get a first impression of the spatial distribution of the pest according to the model predictions.

The resulting raster map was drawn in RStudio (Appendix B.2). It shows the predicted trapping load for the month of September in the suitable elevation range for the olive trees in the study area, from the lake level (187m) to 600 m, with a spatial resolution of 100x100m (Figure 15).

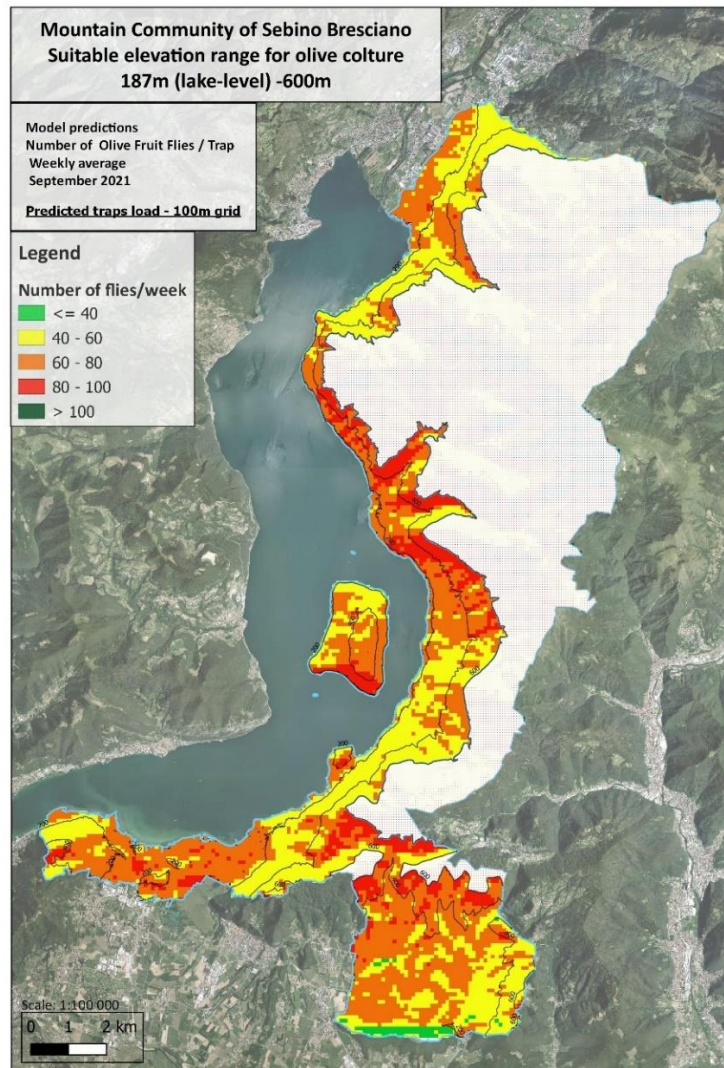


Figure 15 - Map of the predicted average trapping load for the month of September in the study area.

The resulting maps shown a predicted high OFF infestation in all the area with a minimum predicted value of 34 and a max of 122 flies.

From the resulting patchy pattern of the predicted trapping load, it is hard to detect any clear relation between the pest population and the topographic factors in the area. Nevertheless, it is possible to notice that a high OFF infestation (>80 individuals) at high altitude was predicted similarly to the results reported by Castrignano et al. (2012) and Kounatidis et al. (2008) for the same period in a different region. At the same time, very high numbers of flies were predicted also at the lake level in line with the data collected during the monitoring campaign.

It is worth to mention that these results were surprising when compared with the local knowledge of olive farmers, who claimed that in the past the OFF never exceeded the intervention threshold at higher altitudes.

In addition to that, it is important to stress that the olive production for the year 2021 in the area registered an estimated loss of 70-80% compared to the previous year according to local information sources (Romele, 2021).

According to local expert knowledge from the local technicians the yield reduction was due to a combination of factors:

- Alternate bearing typical of the olive trees;
- Local drought after flowering;
- Attacks of both *Halyomorpha halys* Stål
- Attacks of the Olive Fruit Fly;

Even though the estimated loss and the related causes might not be very accurate, due to the lack of scientific works addressing the topic, it is reasonable to hypothesize that the very low production influenced the OFF infestation.

In fact, in a study conducted by Delrio & Lentini (2016), the recorded OFF infestation rate was generally higher (up to 100%) in correspondence to years of low olive production.

Moreover, the lack of fruits availability might have influenced the dispersion rate of the adults' flies which could have migrated to higher altitude looking for a source of food and a place to lay their eggs. In support to this hypothesis, Voulgaris et al., (2013) found that Mature Olive fruit fly can spread over long distances, which, depending on the climate, the topography, and the availability of olives, can reach up to 10km.

Eventually the olives load can have an indirect effect on the OFF, influencing the management and number of treatments applied to the olive tree plantations.

This suggestion came from the information given by the farmers during the monitoring campaign. In fact, most of the producers did not take any action to counteract the damage caused by the Olive Fruit Flies because the harvest was already compromised and so they loss any economic interest in protecting the olives. The cumulated effect of insecticide treatments and other agricultural practices, was proposed by Castrignano et al. (2012) as a possible explanation for the randomness in the distribution of OFF trapping rates.

This work provided a baseline model to predict the spatial distribution of the OFF infestation in the study area focusing on the second generation of the pest which is the target of most of pest management actions.

It was a first attempt to create the most suitable model to support local pest managers improving the information on the pest distribution without making major changes, and investments, in their monitoring protocol.

The simplicity of the resulting model might have negatively influenced its capability to capture the complexity of factors influencing the OFF-population dynamics. For further research on the topic additional variables such as the estimated olive production, management and weather data should be included to increase the ability of the model to capture the complexity of the system.

Moreover, the model resulted from this work still needs to be independently validated because of the lack of more trapping data in the area.

The availability of a consistent and reliable dataset of trapping measures from several years would improve the model capability of generalization and would make it possible to include the effect of the olives load on the OFF infestation. In fact, due to the physiological alternate bearing of olive trees, the high load of olives normally happens once every two years.

The data collected for this research could be an initial database to be enriched during the next year's monitoring campaigns and used for further development of pest risk's models.

Eventually there is a need for longer time series of trapping data to assess whether the changes in the behaviour of the pest reported by the farmers was due to the exceptional conditions of this season or the pest will shift towards higher elevation in the future as a consequence of climate change.

4. Conclusions

Nowadays and more than ever, there is a call for innovative solution in agriculture to achieve the goals of sustainability in the context of the ecological transition, to assure the access to healthy food and reduce the impact of pesticides on the environment.

The reduction of yields losses caused by crop pests still threaten the food security in many developing Countries, while it causes unacceptable economic loss for farmers all over the world.

Area Wide Pest Management strategies can help to reduce losses in the production and to improve the efficiency of pest control practices, especially when they are coupled with consistent information on pest ecology. The influence of climatic variables, especially temperature, on pest population dynamics has been of great interest by entomologists and a wide range of mathematical models have been used to study the complexity of factors influencing the pest development in an agroecosystem. The Olive Fruit Fly (OFF) is a major pest of Olive trees in all olive cultivated areas however, the intensity of damages varies in function to the location of the olive groves, and the associated climatic factors. Modelling the OFF dynamics in relation to the climatic variables can provide useful information to technicians, and farmers to support their decisions in the management of the insect and so reduce economic losses and the abuse of pesticides.

This research integrated the use of trapping data, satellite images and mathematical models to predict the OFF infestation in the Mountain Community of Sebino (Italy).

The local technicians of the Office of Forest and agriculture raised the need for a better understanding of the pest distribution in the area to enable early-detection of pest's outbreaks and to drive their technical support to the most vulnerable areas. Moreover, the location of OFF low-risk areas through the creation of a risk assessment map would permit to highlight suitable zones for new olive groves to be planted. This would be of great interest to produce organic olive oil as they would not require any chemical treatment for the control of the OFF.

Many of the works in literature addressing the relation between the OFF infestation and climatic factors made use of temperature data from weather stations or sensors which were not available in the study area at a sufficient spatial density. Land Surface Temperature data from satellite shown the same problem as their spatial resolution did not fit the rough morphology of the region.

To overcome the lack of temperature data and to account for the mountainous topography of the study site, (experiment I) NDVI-derived metrics from Sentinel-2 image data and (experiment II) topographic variables (i.e. elevation, aspect and distance from the coast) were selected as a proxy of temperature, to predict the OFF infestation.

The computation of the median NDVI values for the 22 sample groves was performed using the Google Earth Engine web platform, while the analysis of the time series was done in the TIMESAT 3.3 software package. The results of the first experiment shown that NDVI data from Sentinel-2 images does not meet the spatial requirements for the extraction of seasonality parameters for sample olive groves in the study area. This is probably due to the small size, the irregular shape, and the presence of non-vegetation pixels in the neighbourhoods of the parcels.

Comparing the parcel's size of the sample groves with the average size of the total of olive groves in the Mountain Community, obtained by visual interpretation of orthophoto, one could say the sample is quite representative of the overall situation and so the proposed methodology cannot be applied to the study area. Nevertheless, the possibility to use the NDVI metrics, from Sentinel-2 image data, to predict the OFF distribution could provide useful outcomes to pest managers and should be tested in other areas characterized by large and regular olive trees plantations such as the South of Italy.

In the second experiment, a multi-layer perceptron Neural network was used to investigate the relation between the main topographic factors influencing temperature and the OFF infestation. The analysis was conducted using the statistics software RStudio and the model topology was selected by a Leave One Out Cross-Validation (LOOCV) process. According to the results of the LOOCV, the simplest model having just one hidden neuron in the hidden layer, shown the lowest error both in terms of SSE (0.0481) and MAE (0.2340). The selected model topology was then used to create a map of the OFF distribution in the CMS area (expressed as the predicted average trapping load for the month of September). From the visual assessment of the map, it was not possible to detect any clear pattern in relation to the selected geographic factors, but the predicted infestation was high all over the study area in accordance with the data obtained during the monitoring campaign 2021.

Although the resulting model will require an external dataset to be tested, this study provided a baseline model to predict the OFF infestation which could be used as a starting point for further research. The trapping data collected for this study provided a first layer of information to be enriched in the following years to create a consistent database for the use of local technicians and researchers. It is worth to stress that the year when this study took place was characterized by a dramatic loss in the olive production due to a combination of factors. That is why there is a need for trapping data from several years to assess whether, and how, the olive production has influenced the OFF population during this season. Accounting for these last remarks, information on the olive load and management practices should be considered in further studies to include more factors influencing the OFF-population dynamics.

5. References

- Acharya, M. C., & Thapa, R. B. (2015). Remote sensing and its application in agricultural pest management. *Journal of Agriculture and Environment*, 16, 43–61.
- Agea. (2015). *Ortofoto Agea 2015*. Ortofoto Agea. https://www.geoportale.regione.lombardia.it/metadati?p_p_id=detailSheetMetadata_WAR_gptmetadataportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&_detailSheetMetadata_WAR_gptmetadataportlet_uuid=%7B6187332D-EF71-469D-9D87-D8214BA7E85E%7D
- Badr, G., Hoogenboom, G., Davenport, J., & Smithyman, J. (2015). Estimating growing season length using vegetation indices based on remote sensing: A case study for vineyards in Washington State. *Transactions of the ASABE*, 58(3), 551–564. <https://doi.org/10.13031/trans.58.10845>
- Belcari, A. (2019). *La mosca delle olive* (Fondazione Clima e Sostenibilità (ed.)).
- Blum, M., Lensky, I. M., Rempoulakis, P., & Nestel, D. (2015). Modeling insect population fluctuations with satellite land surface temperature. *Ecological Modelling*, 311, 39–47.
- Bontempi, R. (2016). *Il Sebino bresciano: aspetti naturalistici* (R. e ambiente S.r.l. (ed.)).
- Castrignano, A., Boccaccio, L., Cohen, Y., Nestel, D., Kounatidis, I., Papadopoulos, N. T., De Benedetto, D., & Mavragani-Tsipidou, P. (2012). Spatio-temporal population dynamics and area-wide delineation of *Bactrocera oleae* monitoring zones using multi-variate geostatistics. *Precision Agriculture*, 13(4), 421–441.
- Castrignanò, A., Boccaccio, L., Cohen, Y., Nestel, D., Kounatidis, I., Papadopoulos, N. T., de Benedetto, D., & Mavragani-Tsipidou, P. (2012). Spatio-temporal population dynamics and area-wide delineation of *Bactrocera oleae* monitoring zones using multi-variate geostatistics. *Precision Agriculture*, 13(4), 421–441. <https://doi.org/10.1007/s11119-012-9259-4>
- Cawley, G. C., & Talbot, N. L. C. (2004). *Fast exact leave-one-out cross-validation of sparse least-squares support vector machines*. 17, 1467–1475. <https://doi.org/10.1016/j.neunet.2004.07.002>
- Comunità Montana del Sebino Bresciano. (2021). *Home page*. <https://www.cmsebino.bs.it/>
- Damos, P., & Savopoulou-Soultani, M. (2012). Temperature-driven models for insect development and vital thermal requirements. *Psyche*, 2012.
- Delrio, G., & Lentini, A. (2016). Dinamica e fattori di regolazione delle popolazioni della mosca delle olive. *Atti Accademia Nazionale Italiana Di Entomologia*, 64, 55–62.

- Devaux, N., Crestey, T., Leroux, C., & Tisseyre, B. (2019). Potential of Sentinel-2 satellite images to monitor vine fields grown at a territorial scale. *Oeno One*, 53(1), 51–58. <https://doi.org/10.20870/oeno-one.2019.53.1.2293>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., & Martimort, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.
- Eklundh, L., & Jönsson, P. (2017). TIMESAT 3.3 with seasonal trend decomposition and parallel processing Software Manual. *Lund and Malmo University, Sweden*, 1–92. <http://www.nateko.lu.se/TIMESAT/>
- COMMISSION IMPLEMENTING REGULATION (EU) 2019/1090 of 26 June 2019 concerning the non-renewal of approval of the active substance dimethoate, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the plac, (2019).
- Ferrari, M., Marcon, E., & Menta, A. (2006). *Fitopatologia, entomologia agraria e biologia applicata*. Edagricole scolastico.
- García-Chapeton, G. A., Toxopeus, A. G., Olivero, J., Ostermann, F. O., & de By, R. A. (2020). Combining favorability modeling with collaborative geo-visual analysis to improve agricultural pest management. *Transactions in GIS*.
- Geoportale Regione Lombardia. (2015). *DTM - 5x5 Digital Terrain Model (2015)*. https://www.geoportale.regione.lombardia.it/metadati?p_p_id=detailSheetMetadata_WAR_gptmetadataportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&_detailSheetMetadata_WAR_gptmetadataportlet_uuid=%7BFC06681A-2403-481F-B6FE-5F952DD48BAF%7D
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guidotti, D., Ragolini, G., & Petacchi, R. (2003). Analysis of spatio-temporal *Bractocera oleae* infestatio distributions obtained from large-scale monitoring network and its importance to IPM. *Proceedings of the Meeting Comptes Rendus de La Réunion at / à Chania (Crete, Greece) 29-31 May 2003*, 28(May 2003), 29–31.
- Günther, F., & Fritsch, S. (2010). Neuralnet: training of neural networks. *R J.*, 2(1), 30.
- Gutierrez, A. P., Ponti, L., & Cossu, Q. A. (2009). Effects of climate warming on olive and olive fly (*Bactrocera oleae* (Gmelin)) in California and Italy. *Climatic Change*, 95(1), 195–217.

- Haghighian, F., Yousefi, S., & Keesstra, S. (2020). Identifying tree health using sentinel-2 images: a case study on Tortrix viridana L. infected oak trees in Western Iran. *Geocarto International*, 1–11.
- Haniotakis, G. E. (2005). Olive pest control: present status and prospects. *IOBC Wprs Bulletin*, 28(9), 1.
- Hird, J. N., & Mcdermid, G. J. (2009). Remote Sensing of Environment Noise reduction of NDVI time series : An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113(1), 248–258. <https://doi.org/10.1016/j.rse.2008.09.003>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Isagro. (2020). *Catalogo - Trappola cromotropica per il monitoraggio e la cattura massale di Ditteri Tripetidi* (pp. 112–113).
- Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, 3(1), 89–93. <https://doi.org/10.7763/ijcte.2011.v3.288>
- Kalamatianos, R., Karydis, I., & Avlonitis, M. (2019). Methods for the identification of microclimates for olive fruit fly. *Agronomy*, 9(6), 1–17. <https://doi.org/10.3390/agronomy9060337>
- Kalamatianos, R., Kermanidis, K., Karydis, I., & Avlonitis, M. (2017). Treating stochasticity of olive-fruit fly ' s outbreaks via machine learning algorithms. *Neurocomputing*, 280, 135–146. <https://doi.org/10.1016/j.neucom.2017.07.071>
- Karungi, J., Adipala, E., Nampala, P., Ogenga-Latigo, M. W., & Kyamanywa, S. (2000). Pest management in cowpea. Part 3. Quantifying the effect of cowpea field pests on grain yields in eastern Uganda. *Crop Protection*, 19(5), 343–347.
- Kounatidis, I., Papadopoulos, N. T., Mavragani-Tsipidou, P., Cohen, Y., Tertivanidis, K., Nomikou, M., & Nestel, D. (2008). Effect of elevation on spatio-temporal patterns of olive fly (*Bactrocera oleae*) populations in northern Greece. *Journal of Applied Entomology*, 132(9-10), 722–733.
- Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173–189.
- Malbêteau, Y., Merlin, O., Gascoin, S., Gastellu, J. P., Mattar, C., Olivera-Guerra, L., Khabba, S., & Jarlan, L. (2017). Normalizing land surface temperature data for elevation and illumination effects in mountainous areas: A case study using ASTER data over a steep-sided valley in Morocco. *Remote Sensing of Environment*, 189, 25–39.
- Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely

- sensed data. *International Journal of Remote Sensing*, 29(3), 617–663.
- Olsson, P.-O., Lindström, J., & Eklundh, L. (2016). Near real-time monitoring of insect induced defoliation in subalpine birch forests with MODIS derived NDVI. *Remote Sensing of Environment*, 181, 42–53.
- Petacchi, R., Marchi, S., Federici, S., & Ragolini, G. (2015). Large-scale simulation of temperature-dependent phenology in wintering populations of *Bactrocera oleae* (Rossi). *Journal of Applied Entomology*, 139(7), 496–509. <https://doi.org/10.1111/jen.12189>
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution*, 20(9), 503–510. <https://doi.org/10.1016/j.tree.2005.05.011>
- Prabhakar, M., Prasad, Y. G., & Rao, M. N. (2012). Remote sensing of biotic stress in crop plants and its applications for pest management. *Crop Stress and Its Management: Perspectives and Strategies*, 517–545.
- Ratner, B. (2009). The correlation coefficient: Its values range between 1/1, or do they. *Journal of Targeting, Measurement and Analysis for Marketing*, 17(2), 139–142. <https://doi.org/10.1057/jt.2009.5>
- Rice, E. R. (2000). Bionomic of the Olive Fruit Fly *Bactrocera* (*Dacus*) *oleae*. University of California. *Plant Protection*, 10, 1–5.
- Rolfi, G. (2003). *Sviluppo e valorizzazione dell'olivicoltura nel territorio del Sebino Bresciano* (Comunità Montana Sebino Bresciano (ed.)).
- Romele, A. (2021). Stagione nera per l'olivicoltura. A rischio la certificazione Dop. *Bresciaoggi*.
- RStudio Team. (2020). *RStudio: Integrated Development for R*. <http://www.rstudio.com/>.
- Sciarretta, A., & Trematerra, P. (2014). Geostatistical tools for the study of insect spatial distribution: practical implications in the integrated management of orchard and vineyard pests. *Plant Protection Science*, 50(2), 97–110.
- Shao, Y., Lunetta, R. S., Wheeler, B., Iiams, J. S., & Campbell, J. B. (2016). Remote Sensing of Environment An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data ☆. *Remote Sensing of Environment*, 174, 258–265. <https://doi.org/10.1016/j.rse.2015.12.023>
- Vajsová, B., Fasbender, D., Wirnhardt, C., Lemajic, S., & Devos, W. (2020). Assessing spatial limits of Sentinel-2 data on arable crops in the context of checks by monitoring. *Remote Sensing*, 12(14).

<https://doi.org/10.3390/rs12142195>

- Viana, C. M., Oliveira, S., Oliveira, S. C., & Rocha, J. (2019). Land Use/Land Cover Change Detection and Urban Sprawl Analysis. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier Inc. <https://doi.org/10.1016/b978-0-12-815226-3.00029-6>
- Voulgaris, S., Stefanidakis, M., Floros, A., & Avlonitis, M. (2013). Stochastic modeling and simulation of olive fruit fly outbreaks. *Procedia Technology*, 8, 580–586.
- Weber, D., Schaepman-Strub, G., & Ecker, K. (2018). Predicting habitat quality of protected dry grasslands using Landsat NDVI phenology. *Ecological Indicators*, 91(March), 447–460. <https://doi.org/10.1016/j.ecolind.2018.03.081>
- Yan, Y., Feng, C.-C., Wan, M. P.-H., & Chang, K. T.-T. (2015). Multiple regression and artificial neural network for the prediction of crop pest risks. *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*, 73–84.
- Yang, L., Peng, L., Zhong, F., & Zhang, Y. (2008). A study of paddystem borer (*Scirpophaga incertulas*) population dynamics and its influence factors base on stepwise regress analysis. *International Conference on Computer and Computing Technologies in Agriculture*, 1519–1526.

6. List of abbreviations

ANN: Artificial Neural Network

AWPM: Area Wide Pest Management

CDD: Cumulated Degree Days

CMS: Comunità Montana del Sebino (Mountain Community of Sebino)

IPM: Integrated Pest Management

LOOCV: Leave One Out Cross Validation

LST: Land Surface Temperature

MLP: Multi-Layer Perceptron

NDVI: Normalized Difference Vegetation Index

NN: Neural Network

OFF: Olive Fruit Fly

RS: Remote Sensing

VI: Vegetation Index

7. List of figures

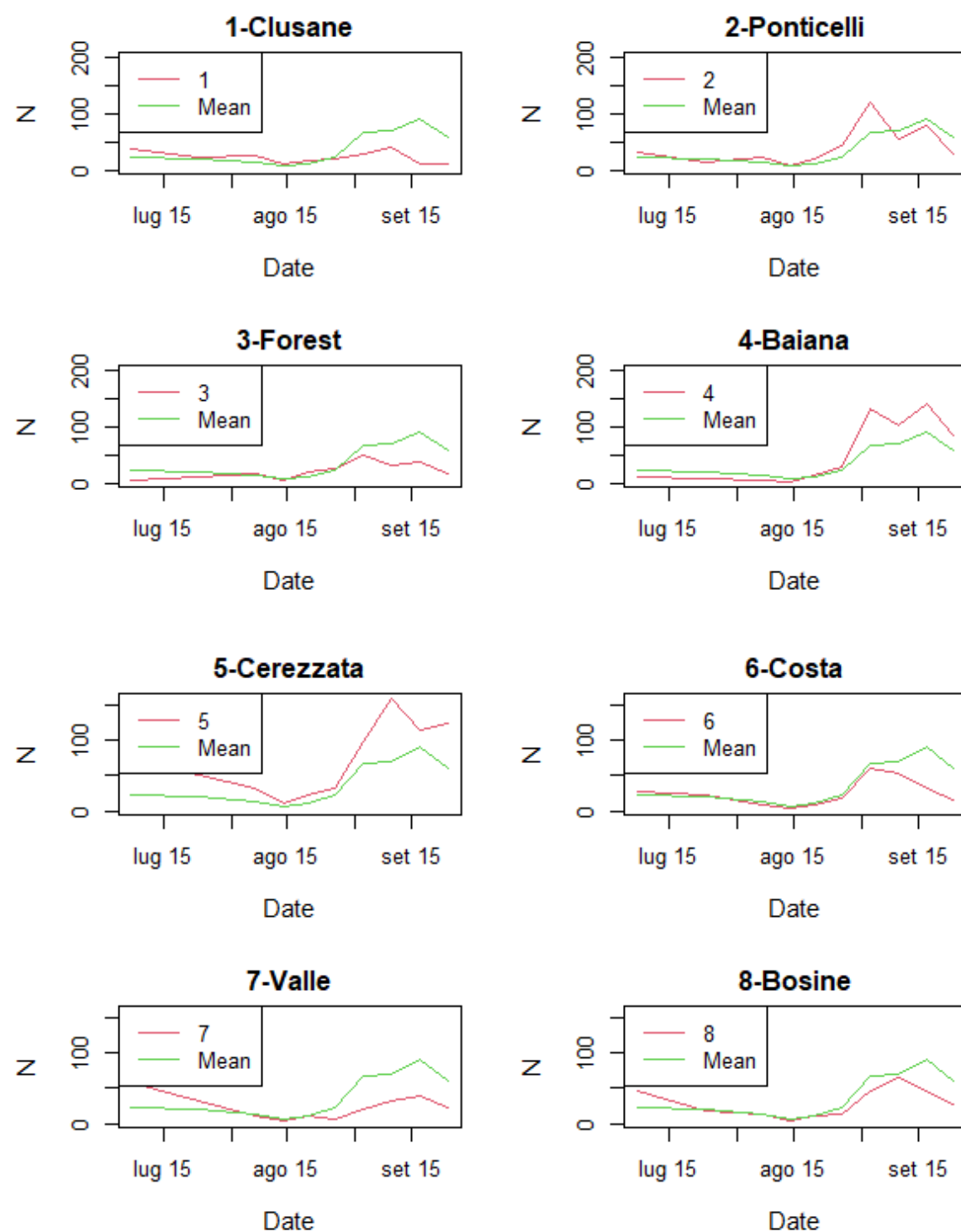
Figure 1 - Smoothed annual NDVI curve for vegetation. The NDVI values start rising since the beginning of the growing season, it reaches a maximum and then decreases in respondents to the start of the vegetation senescence. Adapted from Pettorelli et.al (2005).....	5
Figure 2 - Geolocation and the administrative borders of the study area. It is possible to distinguish the 9 municipalities composing the Comunità Montana del Sebino Bresciano.....	11
Figure 3 – Map showing the Classified olive groves in the CMS area and the location of the OFF-monitoring traps used in this research. A zoom-in to trap number 3 is shown to provide an example of a sample olive grove.	13
Figure 4 -a) One of the wing traps used for the monitoring of the OFF population. It is possible to see the ammonia and the pheromones bates in the upper part of the trap. b) Males OFF adults captured by a trap.....	14
Figure 5- Schematic representation of the four steps for the planned methodology. Extraction of the NDVI time series for the sample groves (1-2) and seasonality parameters (3). Neural Network regression including OFF trapping data (4).....	16
Figure 6 - Data structure of ASCII files. The first line includes the number of years, the number of data per year and the number of time series (nts). The time series are then given line by line (y)(Eklundh & Jönsson, 2017).	18
Figure 7 – A hypothetical schematic representation of the 1 year artificial time-series.	18
Figure 8 –Multi-layer perceptron Neural Network with one hidden layer. Black numbers represent the weights while blue numbers are the so-called Bias representing the intercept.	19
Figure 9 - Spatial distribution of the predictors in the CMS: Distance from the lake (1), Cos of the Aspect (2) and Elevation (3).....	21
Figure 10 – Line plots showing the number of adults OFF per each monitoring visit (N) from 07-2021 to 10-2021, for four sample locations (red lines). The Green line shows the average trapping load among the 23 sample stations.....	23
Figure 11 - Sample groves 1-6 and respective NDVI curves from TIMESAT. The red line showing the double logistic smoothed curve.....	25
Figure 12 -Scatter plots showing the relation between the target variable and the predictors (red box) and the relation between each of the dependent variables.	29
Figure 13 – Histogram showing the mean SSE resulting from the cross-validation. NN1 to NN10 represents the model's architectures with an increasing number of hidden neurons.	31

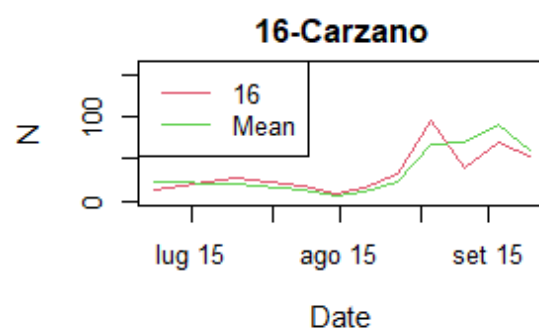
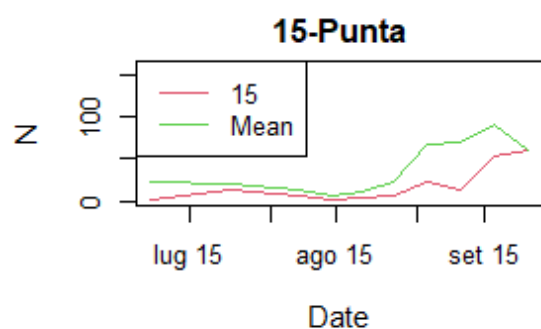
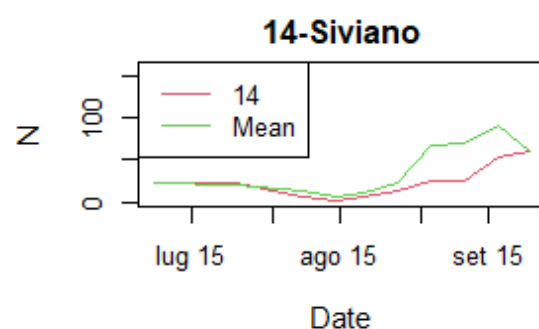
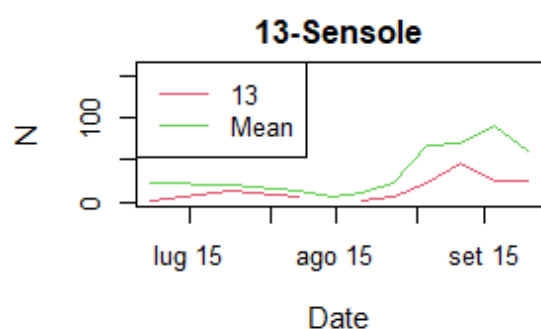
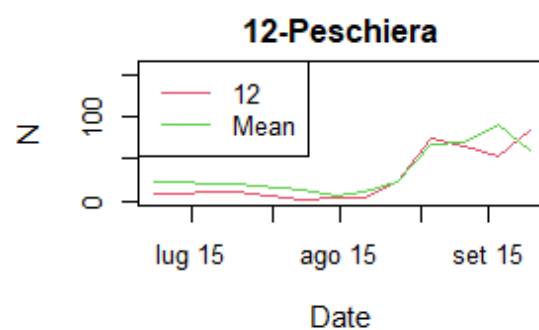
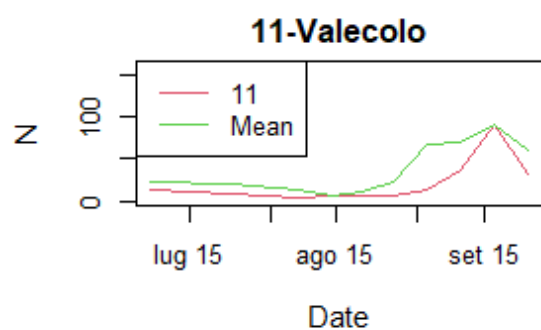
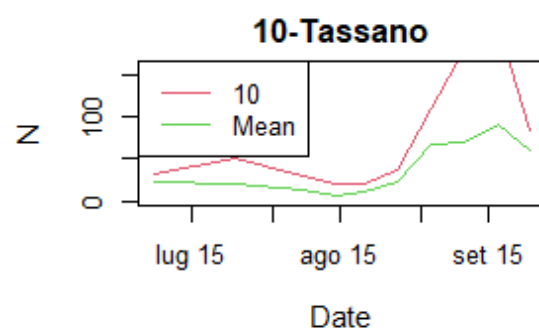
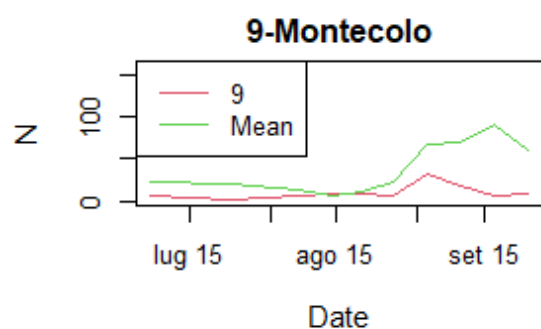
Figure 14 - Selected model topology (NN1) and the relative Mean Absolute Error (MAE).....	32
Figure 15 - Map of the predicted average trapping load for the month of September in the study area.	33

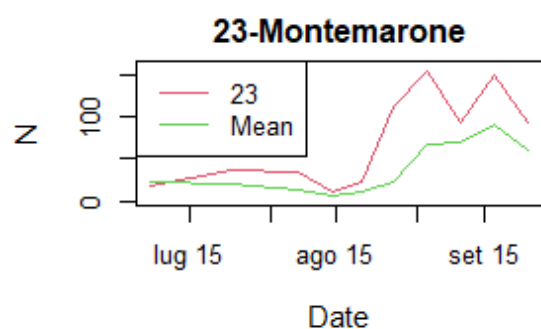
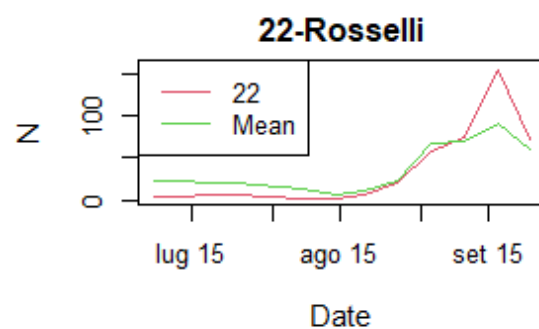
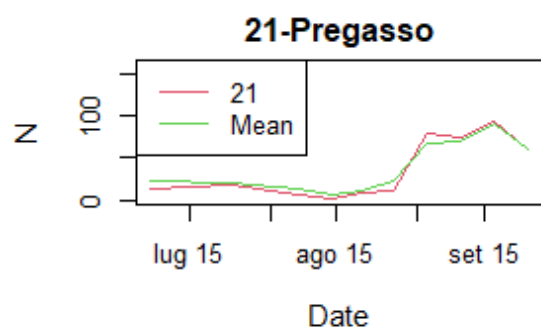
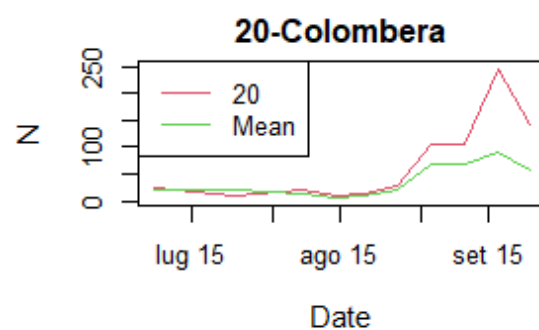
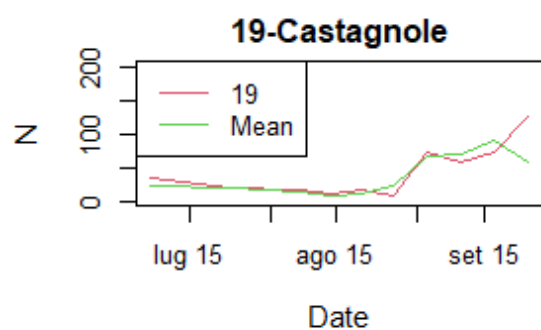
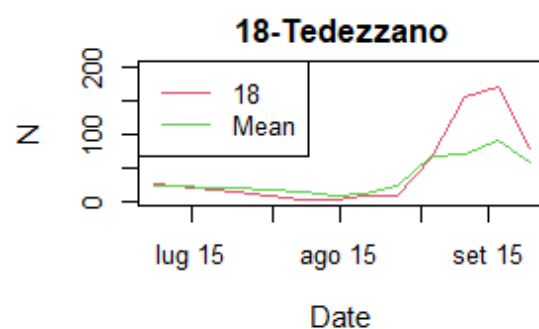
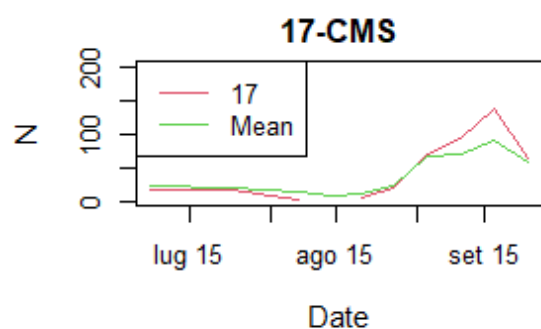
8. List of tables

Table 1 Temperature thresholds for the development of the Olive Fruit Fly (Rice, 2000)	2
Table 2 - Examples of NDVI derived indices and their ecological meaning. Adapted from Pettoirelli et.al (2005)	6
Table 3 - Sentinel 2 Bands and Spatial resolution (ESA, 2015)	15
Table 4 - NDVI median values for the 22 sample grove for the month of January 2021. The first row includes the number of years (3), number of observations per year (66) and the number of time-series (22). -9999 values are missing values or values outside the [0-1] range which were excluded from the analysis.....	24
Table 5 - View of a selection of sample groves from Orthophoto and S2 image, and relative NDVI curves.....	27
Table 6 -Comparison of the sample olive groves with the total.....	28
Table 7 - Descriptive statistics of the data set used for the spatial analysis	28
Table 8 - Pearson's correlation coefficients between the Mean trapping load of september and the selected geographic variables.....	30
Table 9 - Average Sum of Squared Error and Mean Absolute Error resulted from the Leave One Out Cross-Validation of the different model topologies (1-10 hidden neurons)	30

9. Appendix A: Time-series of trapping data during the monitoring campaign 2021







Appendix B: Scripts

1. B.1 GEE Script: Extraction of the median NDVI time-series for the sample groves

```
//NDVI band calculation function
function addnd(input) {
  var ndvi = input.normalizedDifference(['B8', 'B4']).rename('ndvi');
  return input.addBands(ndvi); }
var ndvi_palette = 'FFFFFF, CE7E45, DF923D, F1B555, FCD163, 99B718, 74A901, 66A000,
529400, ' + '3E8601, 207401, 056201, 004C00, 023B01, 012E01, 011D01, 011301';
//Cloud masking function
function maskCloudAndShadows(image) {
  var cloudProb = image.select('MSK_CLDPRB');
  var snowProb = image.select('MSK_SNOWPRB');
  var cloud = cloudProb.lt(5);
  var snow = snowProb.lt(5);
  var scl = image.select('SCL');
  var shadow = scl.eq(3);
  var cirrus = scl.eq(10);
  var mask = (cloud.and(snow)).and(cirrus.neq(1)).and(shadow.neq(1));
  return image.updateMask(mask); }
//Loading and filtering Sentinel-2 Images
var S2_collection = ee.ImageCollection("COPERNICUS/S2_SR")
  .filterBounds(Study_area)
  .filterDate('yyyy-mm-dd', 'yyyy-mm-dd')
  .filterMetadata('CLOUDY_PIXEL_PERCENTAGE', 'less_than', "desired %")
  .map(addnd)
  .map(maskCloudAndShadows);
//Select NDVI band and show layers on the map
var S2_ndvi = S2_collection.select('ndvi');
print(S2_ndvi);
var S2_ndvi_mosaic = S2_ndvi.median().clip("Study_Area");
Map.addLayer(S2_ndvi_mosaic, {min: -0.1, max: 1, palette: ndvi_palette}, 'NDVI S2');
Map.addLayer(groves, {color: 'yellow'}, 'sample olive groves');
Map.centerObject(CMS);
//NDVI time series_Percentiles [5,95,50]
```

```

var percentiles = S2_ndvi.map(function(image) {
  var withStats = image.reduceRegions({
    collection: groves,
    reducer: ee.Reducer.percentile([5,95,50]).setOutputs(['p5','p95','p50']),
    scale: 10  })
  .map(function(feature) {
    var ndvi = ee.List([feature.get(['p5','p95','p50']), -9999])
      .reduce(ee.Reducer.firstNonNull());
    return feature.set('imageId', image.id());  });
  return withStats;
}).flatten();
//Formatting the data frame of Median NDVI Values
var format = function(table, rowId, colId) {
  var rows = table.distinct(rowId);
  var joined = ee.Join.saveAll('matches').apply({
    primary: rows,
    secondary: table,
    condition: ee.Filter.equals({
      leftField: rowId,
      rightField: rowId  })  });
  return joined.map(function(row) {
    var values = ee.List(row.get('matches'))
      .map(function(feature) {
        feature = ee.Feature(feature);
        var p= ee.List([feature.get('p50'), -
          9999]).reduce(ee.Reducer.firstNonNull())
        return [feature.get(colId), feature.get('p50')]
      });
    return row.select([rowId]).set(ee.Dictionary(values.flatten()));
  });  });
var timeSeriesResults = format(percentiles, 'fid', 'imageId');
print(timeSeriesResults.first())
//Exporting .csv
Export.table.toDrive({
  collection: timeSeriesResults,
  description: 'P50_NDVI_Series',
  folder: 'earthengine',
  fileNamePrefix: 'Median_ndvi_series',
  fileFormat: 'CSV'})

```

2. B.2 R-Script: Multi-layer perceptron LOOCV and creation of the OFF-risk output raster

```
setwd("....")
data<-read.csv2("Trapping_data.csv")
colnames(data)<-c("id","mean","elevation","distance","cosaspect")
##Loading required packages
library(tidyverse)
library(neuralnet)
library(GGally)
library(Metrics)
library(plyr)
library(raster)
library(rgdal)
library(tidyverse)
library(ggplot2)
library(gridExtra)

#####
scale01 <- function(x){
  (x - min(x)) / (max(x) - min(x))
}
scaled<- data %>%
  mutate_at (c("mean","elevation","distance","cosaspect"), scale01)
#####
#Leave-One-Out Cross validation
#NN1
SSE1 <- NULL
MAE1<- NULL
RMSE1<- NULL
for(i in 1:nrow(scaled)){
  Train <- scaled[-i,]
  Val <- scaled[i,]
  set.seed(1211)
  NN1 <- neuralnet(mean ~ elevation + distance + cosaspect,
                    data = Train,
                    hidden=1,
                    act.fct = "logistic",
```

```

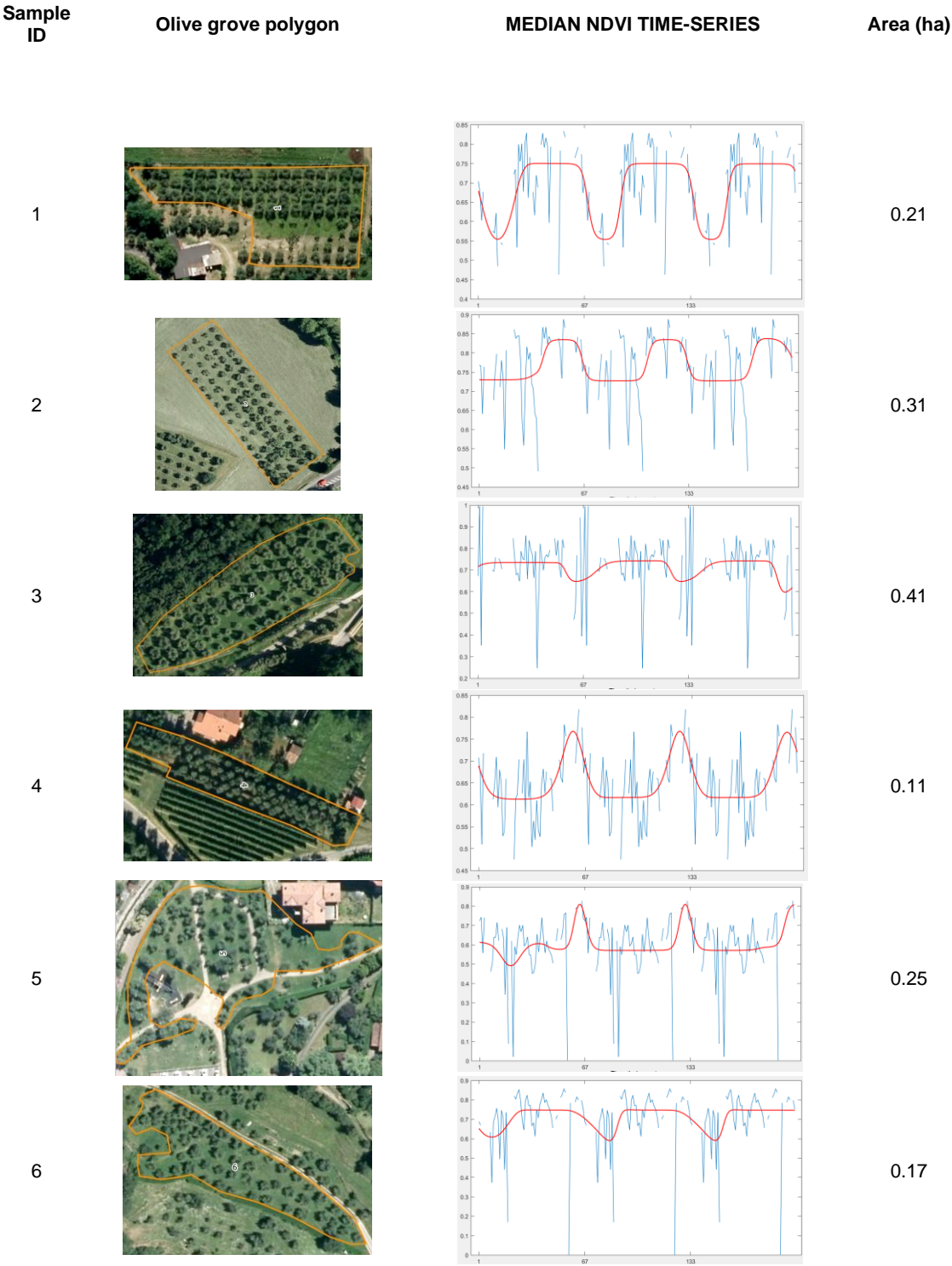
        linear.output = TRUE)

    NN1_pred <- predict(NN1, Val[,3:5])
    SSE1[i]<- sum((NN1_pred - Val$mean)^2)/2
    MAE1[i]<-mae(Val$mean, NN1_pred)
  }
#####
Repeat for 1-10 Hidden Neurons
#####
#LOOCV error calculation
SSE_CV<-cbind.data.frame(SSE1,SSE2,SSE3,SSE4,SSE5,SSE6,SSE7,SSE8,SSE9,SSE10)
ID<-data.frame("NN1","NN2","NN3","NN4","NN5","NN6","NN7","NN8","NN9","ZNN10")
ID = t(ID)
SSE_mean<-data.frame(mean(SSE1),mean(SSE2),mean(SSE3),mean(SSE4),mean(SSE5),
                      mean(SSE6), mean(SSE7),mean(SSE8),mean(SSE9),mean(SSE10))
SSE_mean=t(SSE_mean)
MAE_mean<-data.frame(mean(MAE1),mean(MAE2),mean(MAE3),mean(MAE4),mean(MAE5),
                      mean(MAE6), mean(MAE7),mean(MAE8),mean(MAE9),mean(MAE10))
MAE_mean = t(MAE_mean)
Errors_stat<-cbind.data.frame(ID,SSE_mean,MAE_mean)
#####
#Barplot error mean
SSE_mean_plot<- ggplot(Errors_stat, aes(x=ID, y=SSE_mean)) +
  geom_bar(stat="identity",color="red",
    fill=rgb(0.1,0.4,0.5,0.7))
MAE_mean_plot<-ggplot(Errors_stat, aes(x=ID, y=MAE_mean)) +
  geom_bar(stat="identity",color="red",
    fill=rgb(0.3,0.8,0.9,0.9))
grid.arrange(SSE_mean_plot,
  ncol = 1, nrow = 1)
#####
#Reverse normalization to the get the error in the original units
max<-max(data$mean)
min<-min(data$mean)
Errors_stat$MAE_mean_r <-((Errors_stat$MAE_mean)*
  (max-min)+min)

```

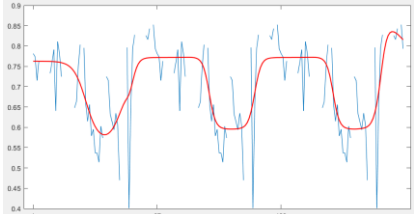
```
#####
#Creation of the output Raster
#####
cosaspect.r <- raster("Cosaspect.tif")
elevation.r<-raster("Elevation.tif")
distance.r<-raster("Distance.tif")
#Resampling to get the same resolution of 100x100m
elevation.r=resample(elevation.r,distance.r,"bilinear")
cosaspect.r=resample(cosaspect.r,distance.r,"bilinear")
#Creation of the spatial dataframes
elevation<-as.data.frame(elevation.r, xy=TRUE)
cosaspect<-as.data.frame(cosaspect.r,xy=TRUE)
distance<-as.data.frame(distance.r,xy=TRUE)
predictors<-data.frame(distance,cosaspect$Cosaspect_def,elevation$Elevation_def)
predictors<- na.omit(predictors)
#Data scaling
predictors.n<-predictors %>%
  mutate_all(scale01)
predictors.n<-predictors.n[,c(1,2,5,3,4)]
#Predict_raster_output using the NN1 model
NN1_raster_output <- predict(NN1,predictors.n[,3:5])
predictors.n$output<-NN1_raster_output
# Reverse normalization to the get the predictions in the original units
predictors.n$rescaled <-((predictors.n$output)*(max-min)+min)
raster.df<-as.data.frame(predictors.n[,c(1,2,7)])
#Rasterization of the resulting spatial dataframe
raster<-rasterFromXYZ(raster.df, res=c(NA,NA), crs="", digits=5)
writeRaster(raster,"NN1_ouput_map",format="GTiff", overwrite=TRUE)
```

10. Appendix C: NDVI curves for the sample olive groves

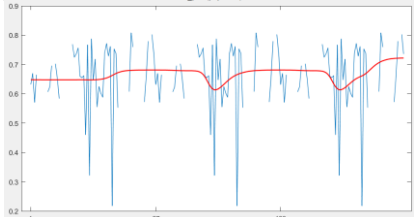


Sample ID	Olive grove polygon	MEDIAN NDVI TIME-SERIES	Area (ha)
-----------	---------------------	-------------------------	-----------

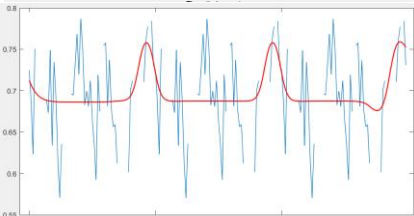
7			0.29
---	---	--	------



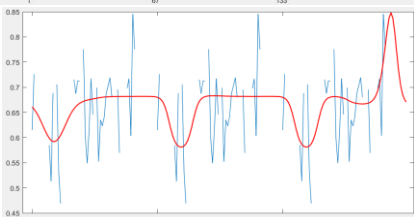
8			0.18
---	---	--	------



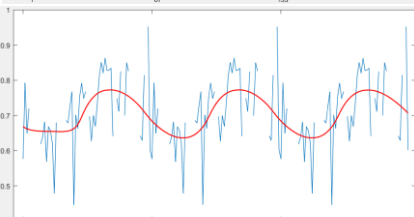
9			0.19
---	--	---	------



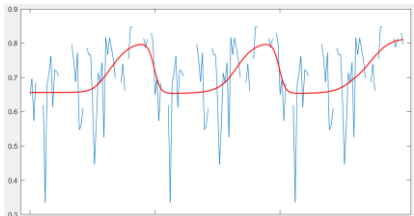
10			0.07
----	---	--	------


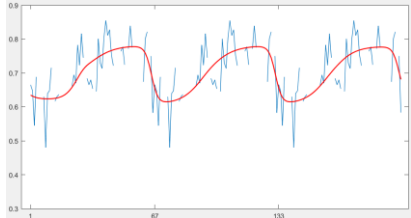

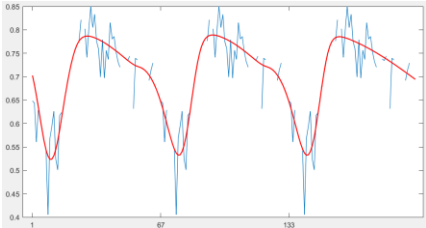

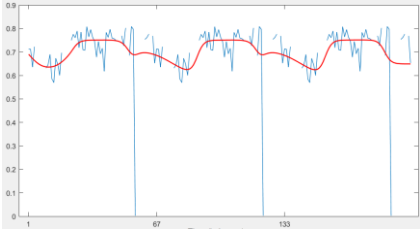

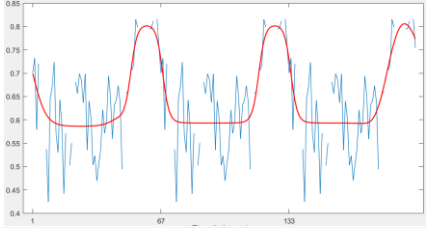

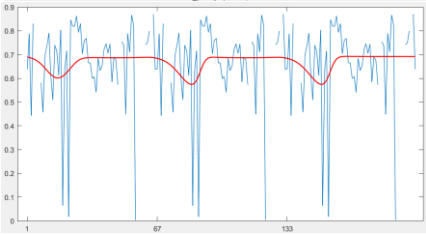



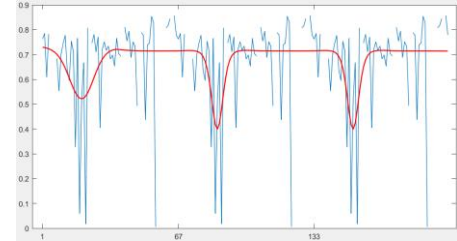

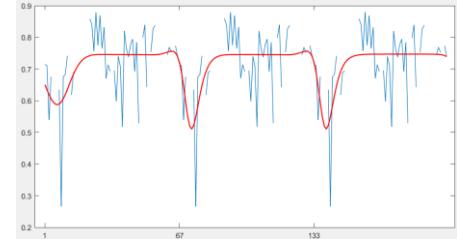

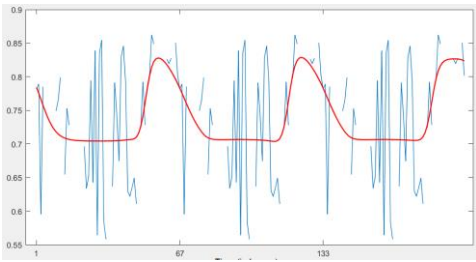

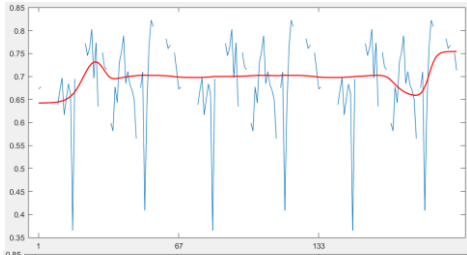

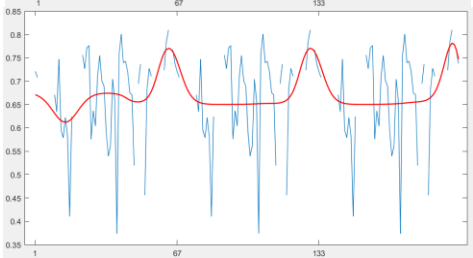
11			0.5
----	---	--	-----



12			0.16
----	---	--	------



Sample ID	Olive grove polygon	MEDIAN NDVI TIME-SERIES	Area (ha)
13			0.05
14			0.21
15			0.23
16			0.52
17			0.46

Sample ID	Olive grove polygon	MEDIAN NDVI TIME-SERIES	Area (ha)
18			0.72
19			0.09
20			0.24
21			0.35
22			0.12

