Heliyon 10 (2024) e31648

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Precision agriculture for wine production: A machine learning approach to link weather conditions and wine quality

Giovanna Maria Dimitri *,[1], Alberto Trambusti [1]

Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Universitá di Siena, Via Roma 56, 53100 Siena, Italy

ARTICLE INFO

Dataset link: https://github.com/GiovannaMariaDimitri/WineQualityMachineLearning

Keywords:
Precision agriculture
Machine learning
Clustering
Wine production
Climate change

ABSTRACT

The agricultural sector, in particular viticulture, is highly susceptible to variations in the environment, crop conditions, and operational factors. Effectively managing these variables in the field necessitates observation, measurement, and responsive actions. Leveraging new technologies within the realm of precision agriculture, vineyards can enhance their long-term efficiency, productivity, and profitability. In our work we propose a novel analysis of the impact of pedoclimatic factors on wine, with a case study focusing on the Denomination of Controlled and Guaranteed Origin Chianti Classico (DOCG), a prime wine-producing region located in Tuscany, between the provinces of Siena and Florence. We first collected a novel dataset, where geographic information as well as wine quality information were collected, using publicly available sources. Using such geographic information retrieved and an unsupervised machine learning approach, we conducted an in-depth examination of pedoclimatic and production data. To collect the whole set of possibly relevant features, we first assessed the region's morphological attributes, including altitude, exposure, and slopes, while pinpointing individual wineries. Subsequently we then calculated crucial viticultural indices such as the Winkler, Huglin, Fregoni, and Freshness Index by utilizing daily temperature records from Chianti Classico, and we further related them to an assessment of wine quality. In addition to this, we designed and distributed a survey conducted among a sample of wineries situated in the Chianti Classico area, obtaining valuable insights into local data. The primary goal of this study is to elucidate the interrelationships between various parameters associated with the region, considering influential factors such as the environment, viticulture, and field operations that significantly impact wine production. By doing so, wineries could potentially unlock the full potential of their resources. In fact, through the unsupervised and correlation analysis we could elucidate the relationships existing between the pedoclimatic parameters of the region, considering the most important factors such as viticulture and field operations, and relate them to wine quality as for instance using the survey data collected. This study represents an unprecedent in the literature, and it could pave the path for future studies focusing on the importance of climatic factors into production and quality of wines.

* Corresponding author.
E-mail address: giovanna.dimitri@unisi.it (G.M. Dimitri).
[1] Equal contribution.

https://doi.org/10.1016/j.heliyon.2024.e31648
Received 5 November 2023; Received in revised form 17 May 2024; Accepted 20 May 2024

Available online 21 May 2024

1. Introduction

The territory serves as the fundamental foundation for any agricultural activity, representing a multifaceted system where various elements coexist within the same space, intricately interconnected with each other [27]. Agricultural endeavors constitute an exceptionally intricate domain, marked by the convergence of environmental fluctuations that impact both the soil and crops. Within this context, operational variations, the timing of interventions, and the periodic intensity of agricultural tasks are all factors that coexist and contribute to the complexity [3]. The outcomes of these agricultural operations can take on multiple forms, primarily influenced by factors such as the specific location and terrain type. However, they are also significantly shaped by climatic conditions and the choice of tools and techniques employed in the process [48].

Agriculture is viewed from a rational perspective that strives to optimize the utilization of all productive resources while taking into account both internal and external constraints, whether they are direct or indirect [3]. Effective resource management is guided by decisions that aim to achieve environmental sustainability, technological advancement, and operational efficiency. In this context lives also viticulture. Viticulture is a complex activity integrated into the production of a commodity highly esteemed on the international market. To maximize and manage resources, a decision-making model is established, drawing on knowledge from a wide array of available techniques and technologies [30]. Nowadays farmers operate within a technological landscape that ranges from traditional tools like hoes to advanced satellite-based systems, each serving a specific purpose in response to the prevailing conditions. Given the distinctive biological characteristics of this sector, viticulturists strive to understand the agronomic impact of operational activities on the cropping system over the short, medium, and long term. Through control and management systems rooted in knowledge and continuous observation, wineries can methodically execute every step of the production process, ensuring it is conducted in an optimal and effective manner [48,41]. In this context a paramount importance is played by climate in the overall quality of wine and viticulture. Viticulture and wine making activity represent a great part of the production sector and the impact of climate change can dramatically affect the production and quality of the final product. For instance some clear effects of climate change on wine production can already be seen. The harvest dates have been dramatically shifted in advance, with the consequence of [8,22,25]. Furthermore several new diseases have been spreading due to the presence of regions which are more and more wet, as for instance fungines infestations and similar virulent diseases [36] with the need of investigating and implementing newer strategies for the fighting of such issues, as for instance microbial activities [4]. On top of this the increase of extreme events has dramatically affected the production quality, due to the increasing number of floods, and out of trend seasonal events as for instance spring frost damages in several wine-growing regions [54,39,46]. In this paper, we conducted a comprehensive study that involved the research and subsequent analysis of data pertaining to pedoclimate and wine production, with a focus on the various processes within environmental, viticultural, and operational domains. Furthermore using Geographic Information Systems (GIS), we examined in depth the region of Chianti, taking into account its morphological characteristics, which encompassed aspects like altitude, exposure, and slopes. This analysis also involved the identification and mapping of individual wineries within the area [43]. In order to study the pedoclimatic characteristic of the area, significant thermal bioclimatic indices for viticulture, as for instance the Winkler, Huglin, Fregoni and Freshness indexes, were calculated in order to evaluate the thermal seasonality in Chianti Classico [31,45]. On top of this a survey was carried out on a sample of wineries, to analyze the production of DOCG Chianti Classico wine. Through a questionnaire, data relating to business aspects, viticulture and wine production in the period 2011-2021 were collected. The paper presents several and important novelties. First of all a comprehensive data gathering was performed, from public sources and with the reconstruction of pedoclimatic indices. Moreover a novel survey was designed and distributed with a comprehensive collection of unprecedentedly available data for the Chianti Region. Using all of such newly collected information and a unsupervised machine learning approach, we were able to disentangle the relationships between the collected features and wine quality and production. To the best of our knowledge no previous studies of such type have been proposed for the Chianti Area under examination. The paper is structured as follows. In Section 2 we report the principal background studies concerning viticultural applications and relationships with climatic factors, including an extensive literature review. Furthermore, in Section 3 we reported the methodologies applied in our study, summarizing the pedoclimatic indicators, also describing an overview of the machine learning approach used in the analysis. In Section 4 we report the datasets overview, describing the sources from which we gathered the relevant data, together with the survey structure and definition. In Section 5 we further describe the experimental settings and the results obtained, discussing findings and implications. Eventually in Section 6 conclusions are drawn, with a thorough description also of the limitations and future developments of our study.

2. Background

2.1. Literature review

The impact of climate change in wine production and wineyards management is seeing an ever more growing interest in the literature. Extreme events, are, in fact, more and more frequent, increasing the number of issues in vineyards management and production. Such effect is in fact highly visible in several countries located worldwide. For instance in [37] the authors analyze the impact of climate change and wine production in the region of Czechia, evaluating how the impact of climate change might affect the area dedicated to the production and wineyards cultivation. In the paper the authors made use of historical data concerning the cultivar in the area, and evaluated possible socio-economical implications of the growing economy. In this context several studies have been conducted in the area of Portugal where wide regions are dedicated to the wine cultivations. For such studies, most of them have focused on high extreme events could affect the growing of the plants, the quality of the wines as well as the definition

Table 1
Summary of Literature Review on Pedoclimatic Factors and Wine.

Reference	Description and Summary	Year
[1]	Examines the impact of climate change on wine production and quality, providing a comprehensive overview of current research in the field.	2023
[5]	Investigates the potential of microbial activities as mitigating strategies for climate-related changes in wine quality, highlighting the importance of food quality in the wine sector.	2019
[8]	Compares the effect of temperature on grapevine phenology between different vineyards, providing insights into temperature variations' impact on wine production.	2021
[16]	Assesses the exposure of Portuguese viticulture to weather extremes under climate change, highlighting the vulnerability of wine regions to changing climatic conditions.	2023
[18]	Utilizes bioclimatic indicators to evaluate climate change impacts on the Spanish wine sector, emphasizing the importance of climate adaptation strategies for sustainable wine production.	2023
[20]	Conducts a bibliometric analysis of climate change influence on grapes and wines, providing insights into research trends and areas for future investigation.	2023
[22]	Explores the manipulation of postharvest periods to enhance vine productivity, addressing practical approaches to mitigate climate-related challenges in wine production.	2006
[25]	Examines the relationship between viticulture and climate in Greece, highlighting recent climate trends' impacts on harvest date variation and wine production.	2014
[29]	Investigates wine quality control parameters and the effects of regional climate variation on sustainable production, emphasizing the importance of climate-resilient practices in wine-making.	2023
[36]	Studies predominant mycotoxins and mycotoxigenic fungi related to climate change and their implications for wine safety, addressing emerging food safety concerns in wine production.	2018
[37]	Assesses the impact of climate change on wine production sustainability and consumption patterns in Czechia, providing insights into adaptive strategies for maintaining wine industry resilience.	2023
[38]	Proposes a mix-method model for climate change adaptation in Italian wine farms, offering a case study-based approach to address climate-related challenges in agricultural practices.	2017
[39]	Analyzes the probability of unprecedented high rainfall in wine regions of northern Portugal, highlighting the need for climate risk assessment and management in wine production.	2023
[46]	Investigates the impact of climate change on viticulture and wine quality, providing insights into climate adaptation strategies and their implications for wine-making practices.	2016
[54]	Examines how extreme heat affects premium wine production in the United States, highlighting the vulnerability of wine regions to climate change-induced heat stress.	2006
[56]	Assesses adaptation options for Chianti wine production in Tuscany (Italy) under climate change using a model-based approach, providing insights into adaptive strategies for wine-growing regions.	2016

of new areas where wine production can be conducted [1,39,16]. Similarly in other European countries, where wine is produced, as for instance Spain and Greece, a plethora of studies have been focusing on the urgence of defining new cultivations parameters and possible countermeasures to take into account the changing scenarios [18,20,29]. Soil pH changing, together with the rise in temperature can in fact lead to new cultivar cycles and the quality of the final product can dramatically be affected by such newly born scenarios. In the context of the Italian wine production an analogous analysis was conducted, and an increasing interest in analyzing the potential effects of climate change into wine production and quality. Among others several studies focus on the development of a model for adaptation to the new conditions. For instance [56] presented an interested model of re-organization of the wineyards in the Chianti Area, which could dynamically adapt to the case of climate changes. Similarly [38] presented an interested study where a decision support system was developed in order to suggest possible changes in the farming organization for Chianti area producers, in light of the future climate change. In this context, further study also focuses, not only on the redesigning of the relevant areas of the farming structures, but also to the definition of the economic impact which such a change could bring in the contexts [6]. So far a comprehensive analysis on the effect of climate change in the Chianti area is not present in the literature, to the best of our knowledge, and in this context our study could contribute significantly to the spread of knowledge and awareness in the population. A summary of the main studies is reported in Table 1.

2.2. Viticulture: soil and climatic factors

The term pedology refers to the complex of physical conditions of the soil in the layer in direct contact with the aerial environment, influenced by climatic factors, such as temperature, rainfall, humidity and solar radiation [12,52]. The nature of the soil defines the variability of the expression of the character of the wine since the fruit of the vine expresses the particular characteristics of the territory. Soil, water, air and nutrients guarantee proper root development; atmosphere, light, heat and wind instead a correct development of the leaf part giving life to a lush plant in all its parts [35]. The soil is a fundamental factor as a reserve of water and mineral salts necessary for the life of the plant. The same vine produces grapes with different characteristics according to the different compositions of the soil that reflects the geological conformation of the area. In poor soils, the plant stimulates root development and reduces aerial development by producing a fruit with higher quality. In fertile soils, on the other hand, plants have a greater vegetative development and less root growth, so there is a greater quantity. The development of the roots is influenced by pedological parameters: porosity, retention capacity and water transfer due to positioning, texture, drainage and organic content. The vegetative energy depends on the root, the more the roots explore the soil and the stronger and more vigorous the plant is [35,53]. To define the soil, reference is made to two main characteristics: texture and chemical composition. Weaving consists of the percentage composition of sand, silt and clay. These components have particles of different diameters and their functionalities at the agronomic level are different:

- **Sand:** (0.02 to 2 mm) performs a mechanical action that makes the soil more porous, permeable and workable. It can cause fertility problems as it does not retain water and mineral salts.
- **Clay:** (<0.002 mm) has a finer texture capable of absorbing and gradually releasing water to the roots. It also retains fertilizers. It brings compactness and plasticity to the ground but in abundance risks making the soil compact and impermeable.
- **SIL:** (0.002 to 0.02 mm) has intermediate characteristics.

The ideal soil from an agronomic point of view is therefore the one that balances the three components, with a clay content not exceeding 20-25%. In addition, the presence of stoniness in the soil leads to a vine with branched roots with less vigor and better ripening; On the other hand, fertile soil involves taproot root with more vigor and delayed ripening. The chemical composition of the soil, on the other hand, concerns the presence of particular substances such as limestone, marl, clay, which together with the type of texture transmit to the grapes and consequently to the wine characteristics of color, aroma, acidity, structure. In general, they change the peculiarities and quality of the wine. The composition of the soil influences the type of wine produced in the following ways:

- **Calcareous-marly soils:** wines tend to have good structure, intense color, low acidity and richness in alcohol percentage;
- **Calcareous-arenaceous soils:** wines tend to be balanced and low longevity;
- **Calcareous-clayey** soils: wines tend to be of good quality;
- **Clayey soils:** wines tend to richness in alcohol percentage, softness and longevity;
- **Sandy soils:** Wines tend to have little color, light structure, good acidity and low longevity [26].

The pH of a soil can be considered a synthetic indicator of the chemical fertility of a soil. The most common agricultural plants prefer a pH close to neutrality, precisely because in that range there is the best availability of nutrients present in the aqueous solution. In the presence of such pH values, plants live better, can be considered less stressed and such pH promotes their energy metabolism. The pH in the viticultural field is important to obtain quality wines $PH = 6 - 8$ [7]. Therefore, healthier and more structured soils favor the development of more balanced plants and less vulnerable to the unexpected. The vine is a plant particularly sensitive to weather conditions and climate. In particular, it fears excessive summer heat and late frosts, needs good lighting and low levels of rainfall. The resistance to cold is limited, in fact the screw does not withstand temperatures below -15 °C. In particular, temperatures below -3 °C are harmful during budding and before flowering below 2.5 °C compromising the harvest and causing damage to the plant. In addition, in the first years of the plant's life, too low temperatures can compromise the growth process not yet completed. The optimal temperature for the vine is about 25 °C but changes during the vegetative period depending on the biological cycle: in May, June and July it varies from 25° and 32 °C and from September it varies between 20° and 25 °C. The lower vegetative limit is considered to be at 10 °C, below which the vine stops its development [26]. The specific thermal requirements of grape varieties are the most important element for their distribution in wine regions. A thermal indicator is described by the Winkler Index, calculated as the sum of the average temperatures, decreased by 10, in the vegetative period of the vine until ripening and harvesting. Through the value of this index it is possible to classify the world vineyards into 5 zones [23,55]:

- **Cool temperate zone:** $IW < 1390\,^{\circ}\mathrm{C}$;
- **Temperate zone**: $1391\,^{\circ}\mathrm{C} < IW < 1670\,^{\circ}\mathrm{C}$
- **Temperate-warm zone:** $1671\,^{\circ}\mathrm{C} < IW < 1950\,^{\circ}\mathrm{C}$;
- **Hot zone:** $1951\,^{\circ}\mathrm{C} < IW < 2220\,^{\circ}\mathrm{C}$;
- **Very hot area:** $IW > 2220\,^{\circ}\mathrm{C}$.

In addition, the thermal trend of a significant season for the vineyard can also be expressed by the Huglin Index. This index takes into account both average temperatures and maximum temperatures measured from April to October [24]. In addition, a high temperature range between day and night hours in the ripening phase of the fruit, which takes place in September for the Chianti Classico, is one of the determining factors for having a good ripening of the grapes and a good concentration of aromatic substances in the skin of the berries. For this evaluation, the Fregoni Index is used, which takes into account the differences between minimum and maximum temperatures recorded in September. Together with the Fregoni Index, also the Freshness Index is taken into account. Such indicator represents the average of the minimum temperatures recorded in September, the month of maturation [17,44]. The temperature ranges, measured by the Night Freshness Index (IFN), mostly appropriate for viticulture are:

- **Hot nights:** $IFN \geq 18\,^{\circ}\mathrm{C}$

  temperatures are high for all types of grape varieties;

- **Temperate nights:** $14\,^{\circ}\mathrm{C} < IFN \leq 18\,^{\circ}\mathrm{C}$

  Late vines are advantaged;

- **Cool nights:** $12\,^{\circ}\mathrm{C} < IFN \leq 14\,^{\circ}\mathrm{C}$

The temperatures are favorable to the ripening of the grapes;

- **Very cool nights:** $IFN \leq 12\,^{\circ}\mathrm{C}$

Good daytime sunshine is necessary to ensure a good level of ripeness. Exposure to light due to the incidence of sunlight on the leaf surface of the vine is also essential to allow photosynthesis. In addition, another important factor regarding the planting of the vineyard is the number of plants present in a hectare of surface. A high planting density leads to the formation of small plants, but more robust and richer in reserve substances, less productive but which guarantee complete maturation. A low density, on the other hand, involves large, very productive and very demanding plants with the consequence of higher costs and less quality. In addition, a less dense vineyard yields less and ages earlier. For example, a vineyard can be defined low density when less than 3000 plants per hectare are present, or high density when around 6000 plants per hectare are available. The vine, in terms of water needs, is much less sensitive than it is to heat and light. However, it is not a plant suitable to be grown in areas that are too humid or too arid. The first water availability for the vine is determined by the rain that in abundance could lead to the development of pests and the reduction of the quality of the crop. Very dry areas, on the other hand, can be considered a limit for the survival of the vine unless you intervene with emergency irrigation [34].

2.3. Precision agriculture

The first definition of precision agriculture was provided in 1997 by the US House of Representations:

"*Precision Farming is an integrated information- and production-based farming system that is designed to increase longterm, site-specific and whole farm production efficiency, productivity and profitability while minimizing unintended impacts on wildlife and the environment"* [33]. Following the definition, we can assess that precision agriculture as an integrated system where information based on the specific farm and production site are tightly integrated, for increasing efficiency, productivity and profitability in the long term, minimizing unwanted impacts on wildlife and the environment. Following this it can be therefore defined a management system that uses multiple methods to achieve these goals through integrated knowledge, based on observation, measurement and response to variables in the field. Site-specific production keeps in mind that the characteristics of the plots are different and even every single square meter can be optimized to avoid waste [19].

Information technology applied to agriculture creates potential benefits that extend beyond the farm, such as environmental performance measurements, quality monitoring and product traceability [2]. With the help of digital technologies, farmers could achieve higher yields, with less use of resources, addressing the challenges of climate change and the scarcity of natural resources. For what concerns Italy the spread of precision agriculture is taking place in this decade and is consolidating in the most specialized contexts such as viticulture.

2.4. Precision viticulture

The vineyard is considered a living element in which environmental, cultural and anthropic factors are recognized and valued to best express the character of the wine [48]. The goal of precision viticulture is to recover this decision-making model by relying on modern detection and analysis technologies. Today's winery, through these tools, will be able to find numerous data to have an in-depth knowledge on a large scale. The most widely used technologies and tools in precision viticulture are:

- **GPS, Global Position System** can determine the exact position has allowed to locate exactly every feature of the fields;
- **GIS, Global Information System,** are computer systems that allow you to geo reference data. Through geostatistical algorithms it is possible to create maps of issues related to the characteristics of the field;
- **GNSS Global Navigation Satellite System** allows support for driving and precision machining of operating vehicles in the field and on the road;
- **VRT, Variable Rate Technology,** are machines that allow you to vary the amount of processing. This technology is used in machinery for soil management, weed control and fertilization;
- **Multispectral and hyperspectral cameras** are intended to provide information on vegetation indices, including chlorophyll content and stress level, which vary in space and time;
- **Sensors** for measuring environmental, crop, operational and machine parameters;
- **UAV, Unmanned Aircraft System**, commonly called drones, are small remotely piloted aircraft: these can be associated with sensors or cameras for aerial detection of information [2,48].

The monitoring and subsequent analysis of the information obtained, allow to manage the production processes as a whole, including environmental, viticultural, mechanical and operational aspects [48].

2.5. Chianti classico

The production area of the Chianti Classico Denomination of Controlled and Guaranteed Origin is located in Italy, in the center of the Tuscany Region with an extension of 718 square kilometers 1. The District includes 30400 hectares of territory under the province

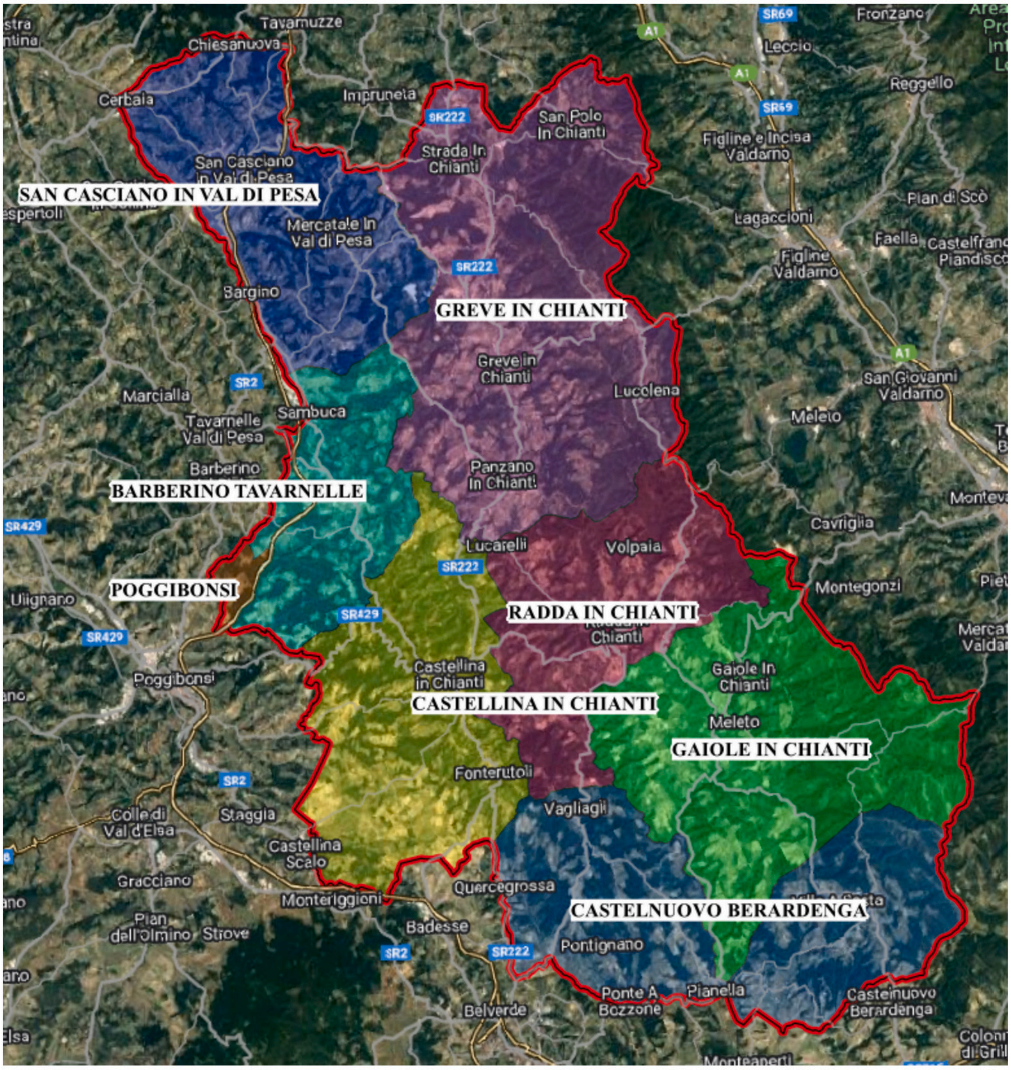


Fig. 1. Figure representing the Chianti Area in Tuscany taken from [10].

of Florence and 41400 hectares in the province of Siena. It is entirely part of the municipalities of: Greve in Chianti, Castellina in Chianti, Radda in Chianti and Gaiole in Chianti; while it partially includes the municipalities of: San Casciano Val di Pesa, Tavarnelle Val di Pesa, Barberino Val d'Elsa and Castelnuovo Berardenga (see Fig. 1) [9].

The territory is limited to the north by the river Greve, to the west by the river Pesa and Elsa, to the south by the sources of the river Ombrone and Arbia and to the east is bordered by the Chianti Mountains. At the morphological level, the environment is mostly hilly, with minimum altitudes not less than 200 m s.l.m., and maximum altitudes on average not exceeding 600 m [9]. From a geological point of view, the soil is shallow, recent and brown. The composition consists of clayey shale, with insertions of scaly clays that are alternated with fine calcareous sandstones, and alberese. At the chemical level, there is a modest amount of organic substance, well endowed with exchangeable cations and reduced presence of assimilable phosphorus ([9]). The territory of Chianti is a land of ancient wine traditions, with Etruscan and Roman testimonies linked to wine. In medieval times it was a land of conflict between the city of Florence and Siena, in which they settled with the construction of villages and castles. At the end of the Middle Ages the cultivation of the vine acquired importance and international fame. In 1716, by the Grand Duke of Tuscany Cosimo III, the wine production area was defined by law [9]. In 1924, a group of producers created the Consortium for the Defense of Chianti Wine, to protect it from plagiarism that tried to imitate it in other parts of Tuscany. The Chianti Classico Wine Consortium is an organism that encompasses all categories of wine production and represents 90% of the production itself. In 1996 the autonomy of Chianti was recognized with a specific production specification, shown below [9].

2.5.1. Production regulations for chianti classico DOCG wines

The Chianti Classico DOCG wine production regulations establish the rules and parameters for viticulture and winemaking. A well-defined geographical limit has been established within which those who adhere to this specification produce a wine with the

Chianti Classico denomination. The Specification works to safeguard the quality of the wine, to guarantee the origin and ensure the effectiveness of the controls, with the aim of protecting the reputation of the Chianti Classico DOCG wine. The denomination of controlled and guaranteed origin is distinguished by the "Black Rooster" brand, which is shown on the bottles following the procedures established by the Consortium. The Chianti Classico DOCG is attributed to the red wine that has the conditions and requirements established by this specification. The Chianti Classico denomination may be followed by the mention "Riserva" or "Gran Selezione" based on specific criteria of grape processing to ensure better quality (Consorzio Vino Chianti Classico, 2014).

2.5.2. Rules for chianti classico viticulture

Chianti Classico wine is made from grapes originated from vineyards located in the demarcated area of Chianti Classico. The ampelographic composition is formed by the native Sangiovese vine, with a percentage between 80% and 100%. A maximum of 20% of individual holdings may use grape varieties present in the vineyard register as "complementary grape varieties", for example: Cabernet, Canaiolo, Ciliegiolo, Colorino, Merlot, Malvasia, etc. [9]. The vineyards must have a hilly position, with a suitable sun exposure and be located at an altitude of less than 700 m. These are planted in soils consisting mainly of arenaceous substrates, clay shale, sand, calcareous marly and pebbles. Vineyards located in wetlands or at the bottom of the valley and planted in predominantly clayey soils are not suitable for this production [9].

The cultivation techniques must be such as not to modify the peculiar characteristics of the grapes. The vineyards can be used for the production of grapes for Chianti Classico starting from the third year after planting. The planting layouts, the training forms and the pruning systems must be traditional, such as not to change the characteristics of the grapes and wine. The traditional forms of farming are goyot, Tuscan bow and spurred tail. The minimum planting density is set at 4400 vines per hectare [9].

2.5.3. Rules for chianti classico winemaking

The operations of vinification, conservation, aging, bottling and aging in bottle, must be carried out in cellars located within the delimited production area of Chianti. Only local, loyal and constant practices are allowed in winemaking ([9]). Chianti Classico can be placed on the market from October 1st of the year following the harvest. Chianti Classico Riserva wine must undergo at least 24 months of aging, of which at least 3 months of aging in bottle. Chianti Classico Gran Selezione wine must undergo at least 30 months of aging, of which at least 3 months of aging in bottle. The ageing period is calculated from 1 January of the year following the harvest.

3. Methods

In this section we will go through an in depth description of the methodologies used in our paper. First we will briefly give an overview of the Geographic Information System approach to retrieve and analyze spatially referenced data. Secondly we will summarize pedoclimatic quality indicators, briefly mentioned in the background section above. Then we will give an overview of the unsupervised machine learning clustering approach used in the paper.

3.1. Territorial analysis with GIS systems

The Geographic Information System, known as GIS-Geographical Information System, can effectively manage and analyze spatially referenced data. Geographic information allows a spatial analysis activity through a cartographic, digital and quantitative approach, characterized by an organization of data in logical layers connected to a descriptive component. The Information System is a multidimensional archive of geographical and alphanumeric data for the management and analysis of the territory [32]. In GIS, geographic data is stored in the form of two macro-models, vector and raster model. In the vector model, information consists of geometric shapes, such as points, lines, and polygons, defined by the coordinates (x, y) of the points at each vertex. Each shape represents an object to which a database is associated, structured in tables, within which the significant attributes are stored [32].

The raster model consists of pixels that represent the unit, useful for discretizing the territory into square cells, more or less wide side according to precision. Each of them is associated with a numerical value that can represent a physical characteristic, or the measure of a phenomenon, for example the elevation above sea level or the slope of a slope [32]. The Open Source Quantum GIS (QGIS) software, through a graphical interface, manages the numerous data in various formats on geographical maps. QGIS has significant integrations with the plug-in *Processing toolbox* which allows you to use different geo-algorithms interfacing with the functions of R, SAGA, GRASS, and others [32]. In the paper we used The version of the QGIS software used is the 3.28 Firenze version, on a Windows operating system.

3.2. Climate analysis

Climate is the set of atmospheric conditions over time, in a certain place on the earth's surface. The physical parameters that allow to describe it are temperature, intensity and duration of solar radiation, atmospheric pressure and type/frequency of precipitation. Focusing on the vegetative growth and ripening of the fruits of the vine, the main factors to be measured for the quality of wine products are the temperature, light radiation and rainfall of the territory under analysis. Among the climatic factors, the air temperature plays a decisive role on the overall ripening of the grapes and in particular on the aromas and flavors, with important repercussions on the characteristics of the wines [47,50,14,49]. In our analysis, we focused solely on the temperature variable, as it represents the culmination of numerous climatic influences. The consideration of the water factor was omitted, as we regard the vine as inherently drought-resistant. Additionally, in instances of severe water scarcity, irrigation systems can be implemented by humans to address any deficiencies.

Light energy is a factor difficult to model because solar radiation reaching the ground is only a fraction of the energy; This varies greatly depending on the cloud cover, the purity of the atmosphere, the altitude and the slope of the sun's rays, due in turn to the latitude and time of year [24].

Three reference climatic levels have been determined:

- **Macroclimate**, referred to latitude and distance from seas and mountains for large-scale areas
- **Mesoclima,** deals with local characteristics such as altitude, exposure and distance from the valley floor;
- **Microclimate**, given by atmosphere in direct contact with the vine and the bunches of grapes [47].

To represent the agrometeorological situation of Chianti Classico, it is advisable to focus on the mesoclimate, as it represents the climatic evolution referred to a specific vineyard area at a given time.

3.2.1. Thermal bioclimatic indices

The bioclimatic indices represent a fundamental factor of the seasonal climate trend and are extremely useful for classifying the areas of wine production. Below, the most significant indices for viticulture are reported, based on thermal trends.

Winkler index The Winkler index consists of the sum of the average temperatures, subtracted by 10, measured in the vegetative period of the vine until maturation and harvest, considered from 1 April to 30 October. Days with temperatures below 10 °C are placed with no contribution because the vine stops its development. The Winkler index is expressed in degree days .[°C • $perday$]

$$IW = \sum_{1/4}^{31/10} (T_{Med} - 10)$$

$$se\ \left(T_{Med} - 10\right) < 0 \rightarrow T = 0$$

Where:
IW: Winkler index
T_{Med}: medium (average) daily temperature
T: calculated temperature

This parameter of the viticultural potential is linked to the ripening of the berries and the potential for accumulation of sugar, favored by the suitable temperatures in a given environment [55].

Huglin index

The Heliothermal Index of Huglin is composed of the sum of the average and maximum daily temperatures above the minimum threshold of 10 °C, in the period between pre-ripening and harvest, from 1 April to 30 September for the northern hemisphere. The maximum temperature is linked to the activity of the vine during the daily hours. In addition, there is a coefficient of "length of day" to take into account the hours of light in the region on the globe, depending on latitude. It takes values between and between 40° and 50° latitude. Chianti Classico is located at a latitude of about 43°, so the coefficient is set to $k = 1.03$

$$IH = \sum_{1/04}^{30/09} \left[\frac{(T_{Med} - 10) + \left(T_{Max} - 10\right)}{2} \right] \bullet k$$

$$se\ \left(T_{Med} - 10\right) < 0 \rightarrow T = 0$$

$$se\ \left(T_{Max} - 10\right) < 0 \rightarrow T = 0$$

Where:
IH: Huglin index
T_{Med}: average daily temperature
T_{Max}: maximum daily temperature
T: calculated temperature

This index is more suitable to describe the hilly areas, characterized by strong temperature ranges, taking into analysis mainly the daytime temperatures related to the photosynthetic activity of the vine. The Huglin index has been shown to have a greater correlation with the sugar level present in harvested grapes than the Winkler index [24].

Fregoni index

The simplified Fregoni index is composed of the sum of the daily thermal oscillations in September, derived from the difference between maximum and minimum temperature, multiplied by the number of days with temperature below 10 °C.

$$IF = \sum_{1/9}^{30/9} \left(T_{Max} - T_{Min}\right) \bullet\ Nd_{(T<10\,°\mathrm{C})}$$

Where:

IF: Fregoni index
T_{Med}: average daily temperature
T_{Min}: daily minimum temperature
T: calculated temperature
$Nd_{(T<10\,^{\circ}\mathrm{C})}$: number of days with average temperature below 10 °C

In the ripening phase of the grapes, which takes place in September, the thermal oscillations and the cool days are favorable to the synpaper of aromatic substances, polyphenols, such as tannins and anthocyanins, and finally to the maintenance of an acidity such as to guarantee a longevity of the wine. The latter are all elements that guarantee a higher quality of the final product [17].

Freshness Index of Nights

The Cool Night Index, or Cool Night Index, is the average of the minimum temperatures of the terminal period of ripening of the grapes that takes place in September. This index is associated with the quality of the harvest since it quantifies the development of polyphenols and aromas in grape ripening [44].

$$IFN = \underline{T}_m^9$$

Where:
$\underline{T}_m^9$ average daily minimum temperatures for September.

3.3. Unsupervised machine learning and clustering

Clustering is a technique that is part of the family of unsupervised machine learning algorithms. Clustering methods can be applied to group observations with similar characteristics and separate those with different characteristics widely used in several application fields [15,13,40,42]. The goal of clustering is to produce groups of observations with a high degree of similarity within each group, and a low degree of similarity between different groups. A secondary goal of clustering is to reduce data complexity so that it can be easily explored across subgroups [28,40]. In our study we used the K-Means clustering which is a well known partition methods for dividing elements into subgroups according to their distance [11,21].

3.3.1. Partition of Voronoi

The partition of a set consists in forming subsets of mutually disjointed, such that their union is the same. Let $Q \subset \mathbb{R}^d$ be a convex polytope, $P = \{p_1, p_2, \ldots, p_N\}$ a finite set of nodes, where $p_i \in Q$, and $I_N = \{1, \ldots, N\}$ a set of indices. The standard Voronoi partition, generated by P with respect to the norm, is the Euclidean set $\{V_i(P)\}_{i \in I_N}$ defined as:

$$V_i(P) = \{p_i \in Q \mid \|q - p_i\| \le \|q - p_j\|, \forall p_j \in P\}$$

where $\| \cdot \|$ represents the Euclidean norm.

In the Voronoi partition the space to be partitioned is a subset connected to Euclidean space, the nodes are points of space and the measure of distance is Euclidean [51]. Voronoi diagrams are related to the k-nearest neighbors (k-NN) algorithm, an algorithm used in classification, regression, and clustering problems. The algorithm is based on the characteristics of objects close to the one considered. Since Voronoi diagrams divide the space into polygons containing the points closest to each point, the edges of Voronoi cells correspond exactly to the decision limits of a simple k-nearest neighbors problem.

4. Dataset

We will hereby describe the process we carried out to obtain the relevant dataset we used in our analysis. Data and code are anonymised and available in the folder: https://github.com/GiovannaMariaDimitri/WineQualityMachineLearning

4.1. Map data

To carry out the territorial analysis, the QGIS software was used, working with geo–referenced information starting from subsequent data. We used QGIS version 3.28, Firenze, on a Windows Operating System.

4.1.1. Basic cartography

For basic maps, WMS Services have been used, which allow you to connect to cartographic services on the Web, so you can investigate them without downloading them. The services offered by Google have been used, with particular attention to Google Satellite, Google Hybrid and Google Label.

4.1.2. Cartographic data Tuscany region

In the cartographic portal of the Tuscany Region, called GEOscopio ([2]), countless cartographic data are stored for free consultation, extraction, reproduction and modification in compliance with the terms of the Creative Commons license with which they were

[2] https://www502.regione.toscana.it/geoscopio/cartoteca.html.

released. These cartographic data, produced by the Tuscany Region and disseminated through web services, are transferred in copy pursuant to Executive Decree n.663/2014 (GEOscopio).

The following Datasets have been downloaded from the GEOscope:

- **Administrative areas of the Tuscany Region**

(Data source: Tuscany Region – Administrative areas) (Regione Toscana)

- **DOC and DOCG wine production areas in Tuscany**

(Data source: Tuscany Region – Wine production areas) (Regione Toscana)

4.1.3. Chianti classico wineries

In the following study, 280 wineries registered with the Chianti Classico Wine Consortium were analyzed. Through an online search, on the relative sites of each company, data relating to the location of the registered office, telephone and e-mail contacts were obtained, useful for the provision of the questionnaire (the full questionnaire is reported in the Supplementary Material). Each location address has been converted into latitude and longitude values to be entered as geographical coordinates in the geographic information system.

4.1.4. Vineyards Tuscany region

Through the Open Data of the Tuscany Region it was possible to download a file containing information on the vineyards registered in the Register of Vineyards 2022, claimed according to ARTEA, the Tuscany Regional Agency for Agricultural Disbursements from the Tuscany Region local government.

4.1.5. Digital land elevation model DEM

To carry out the morphological analysis it was essential to search for the digital soil elevation file. The National Institute of Geophysics and Volcanology, INGV, offers DEM images, Digital Elevation Model, with 100 m resolution, for the whole of Italy. Identified the area of interest in the system, the files were downloaded: W48065 and W48070.

4.2. Climate data

In the SIR, Tuscany Region Hydrological System, it is possible to consult the meteorological data updated in real time for each meteorological station present in the regional territory. Within the SIR filing system, it was possible to obtain historical data of climatic factors.[3] Initially, through the cartography of the system, the stations located within the Chianti Classico and in the surrounding area were identified. Subsequently, we discarded weather stations with insufficient data for analysis. We later constructed a dataset containing the stations of interest, in particular containing the distinctive information, such as ID, station name, type of measurement, and geographical coordinates, in latitude and longitude, in order to use this as input in the GIS program. For the following analysis, the data relating to the daily temperatures recorded in the decade 2011-2021 have been downloaded. In this way we were able to create a dataset with 4019 values of minimum and maximum daily temperatures, measured by 16 thermometric stations distributed in the Chianti Classico territory.

4.3. Data on Chianti Classico wineries

4.3.1. Survey collection

As part of the data collection process, we designed and administered a novel survey to Chianti Classico wineyards in order to further investigate the impact of pedoclimatic factors into wine quality, directly from producers. The data collection took place through a questionnaire addressed to the wineries present in the Chianti Classico area. The population for the statistical study is composed of 280 companies present in the territory. The wineries of interest for this study have been identified in the register of the Chianti Classico Wine Consortium, an entity that guarantees and controls the quality of the wine produced [9].

general, technological, viticultural and wine aspects. The whole questionnaire structure is released in the supplementary material of the paper. The survey was structured avoiding choice biases and designing the questionnaire structure in such a way that most of the possible informations could be retrieved. The large dimensionality of the sample considered allowed us to introduce possible sampling biases. The questionnaire consists of 5 sections:

- **Generic aspects about the company**
- **Viticulture**,
- **Chianti Classico DOCG wine produced from 2011 to 2021**;
- **Technologies**.

[3] Regione Toscana, 2023.

Table 1
Questionnaire sections with their question number. The full questionnaire is reported in the Supplementary Material.

SECTION	QUESTION NUMBER
Generic aspects	1 - 10
Viticulture	11 - 25
Chianti Classico DOCG wine:	
Year 2011	26 - 38
Year 2012	39 - 51
Year 2013	52 - 64
Year 2014	65 - 77
Year 2015	78 - 90
Year 2016	91 - 103
Year 2017	104 - 116
Year 2018	117 - 129
Year 2019	130 - 142
Year 2020	143 -155
Year 2021	156 - 168
Technologies	169- 181

In each section, 10/15 questions were presented, with the aim of obtaining the attributes necessary for the analysis. All of the questions submitted in the questionnaire are listed, the structure of which is illustrated in Table 1.

4.3.2. Delivery method of the questionnaire

The questionnaire was developed with the help of the "Google Forms" platform, through which it was possible to obtain answers from wineries in an effective and immediate way. Data collection began on 3 April and ended on 1 June 2023. Initially, a telephone call was made directly to each company, in order to present the research to involve and stimulate more participation. Subsequently, an e-mail was sent with what was said in writing in the telephone call and with the questions contained in the questionnaire attached in a pdf file. The latter allowed companies a first view of the proposed questions in order to leave the time necessary for data recovery. Finally, the questionnaire form which we named: "UniSi - DOCG Chianti Classico", was administered by e-mail. This form made it possible to receive direct responses from the participating wineries. The data obtained from the questionnaire allowed the production analysis of Chianti Classico. The full questionnaire is available in the supplementary material.

4.3.3. Data preparation

To facilitate analysis, the data were divided into areas, creating multiple datasets:

- **Winery Dataset**;
- **Viticulture Dataset**
- **Wine Production Dataset**
- **Wine Quality Dataset**

The attributes of these datasets are composed of numeric variables, with continuous values, and nominal character values. For a matter of privacy, the data will be shown in aggregate, without showing the name of the winery in relation to the data provided. Only companies that have consented to the dissemination of information referred to them have been mentioned.

5. Experiments and results

5.1. Pedoclimatic analysis

In this study, the area is limited to the Chianti Classico area. To carry out the analysis of the territory, the Open Source Quantum GIS Georeferenced Information System was used. The physical structure of the database, within the QGIS, consists of geographical elements in vector and raster format linked to alphanumeric attributes. The information layers used were effective in describing and analyzing the pedoclimatic aspects of Chianti Classico. The database of the information system consists of:

- **WMS basic cartography;**
- **Provincial and municipal administrative limits;**
- **Border area of the DOCG Chianti Classico;**
- **Chianti Classico wineries;**
- **Vineyards Tuscany Region;**
- **Weather stations;**
- **Digital Elevation Model, DEM**.

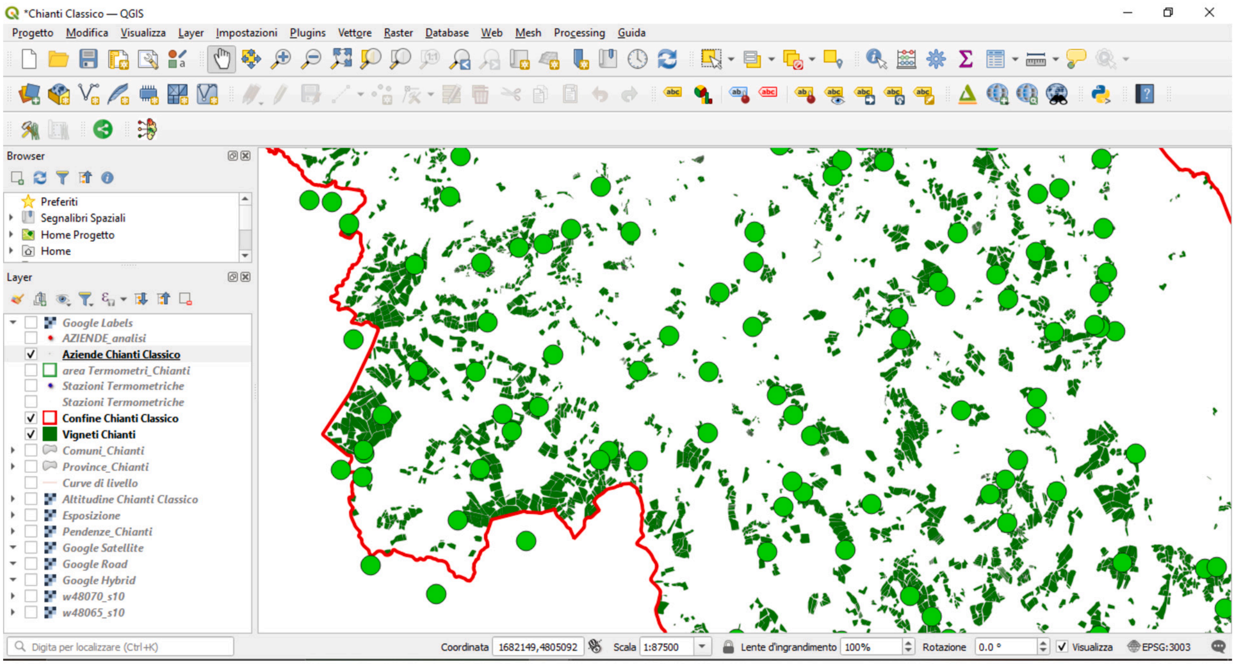


Fig. 2. QGIS; Wineries and vineyards in the Chianti Classico area.

Through the analysis and processes developed with the QGIG Software, the following maps relating to the Chianti Classico territory have been created:

- **Map of vineyards** (shown in Appendix A);
- **Map of the area of influence of thermometric stations** (shown in Appendix B);
- **Map of altitudes** (shown in Appendix C);
- **Exhibition map** (shown in Appendix D);
- **Map of the slopes** (shown in Appendix E)

Initially, in the QGIS it was set as the Reference System (SR) of the coordinates of the Monte Mario project / Italy zone 1, EPSG:3003. To visualize the study area, the following scales were used: 1:200 °000 for the entire Chianti territory, between 1:200 °000 and 1:100 °000 for the municipal territories and 1:50 °000 to frame the company extension. The basic maps were uploaded to the QGIS through the web-connected plugin, QuickMapService. The services offered by Google have been used, with particular attention to Google Satellite, Google Hybrid and Google Label that report geographical information combined with satellite images. Through these maps it was possible to have a qualitative and direct comparison of the various characteristics of land use and the positioning of structures and plants. Once the various thematic maps were produced, reported below, Google Label was used in overlay, in order to identify the denominations present in the area.

Vineyard map Through the cartographic data of the Tuscany Region it was possible to identify the boundaries of the study area related to the Chianti Classico DOCG. This limit was used to crop all subsequent layers loaded into the QGIS, using Geoprocessing tools. This process was also carried out for the administrative areas, identifying the provinces of Florence and Siena and the municipalities falling within the Chianti area. The dataset of the Chianti Classico wineries has been uploaded to QGIS to obtain a series of geolocated points that identify the location of the Chianti Classico wineries. The file of the Tuscany Region Vineyards uploaded to QGIS represents, in the form of vector polygons, the geolocated vineyards in the regional territory, showing geometric and informative parameters. The vineyards present in the Chianti Classico territory were then selected, using cutting-out Geoprocessing tools as shown in Fig. 2. In the Vineyard List all vineyards are reported without making distinction to the Chianti Classico denomination and other DOC and IGT denominations. By geolocating wineries and extending vineyard boundaries, we were able to qualitatively and holistically analyze the viticultural characteristics of the region.

Appendix F shows the vineyard map relating to the Chianti Classico area. Through this map, it is possible to observe that, in the Chianti Classico territory, the area covered by vineyards is very extensive and has a high density along the slopes of the valleys. The wineries are located within the area cultivated with vines. The areas not intended for vineyards have more concerted wooded areas, both in the central part (between Radda and Castellina), and along the eastern border of the Chianti Classico territory.

Map area of influence thermometric stations The meteorological station dataset was then uploaded to the QGIS system, geolocating the weather stations on the reference cartography as shown in Fig. 3.

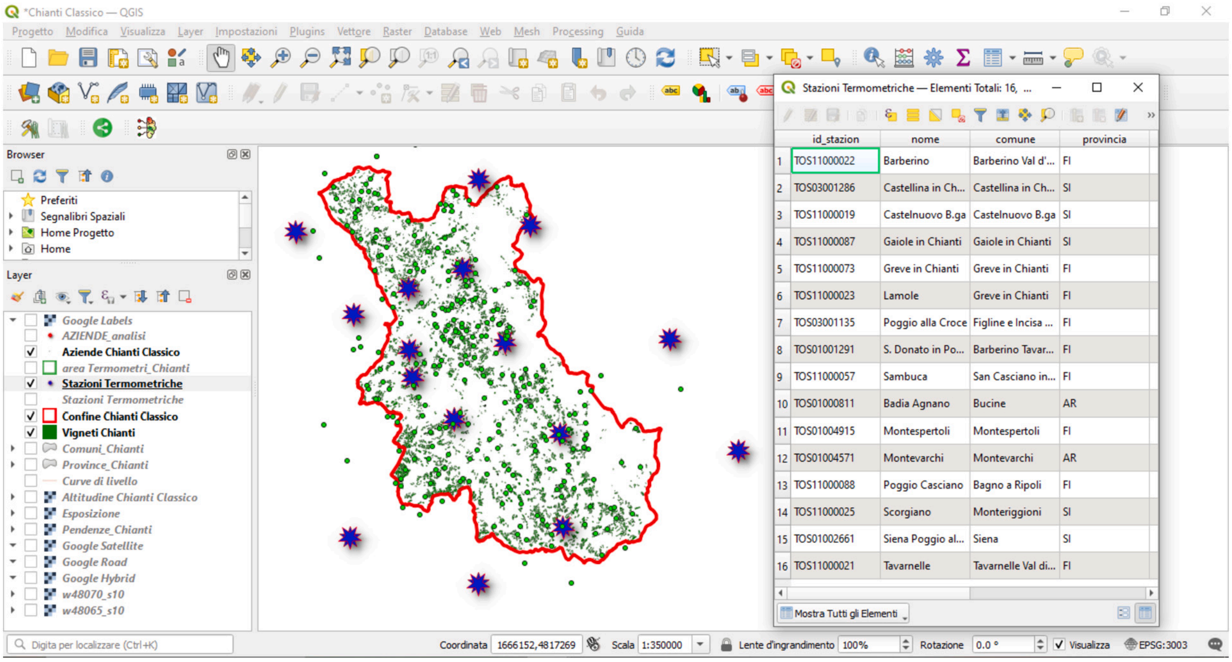


Fig. 3. QGIS: geolocation of thermometric stations in Chianti Classico.

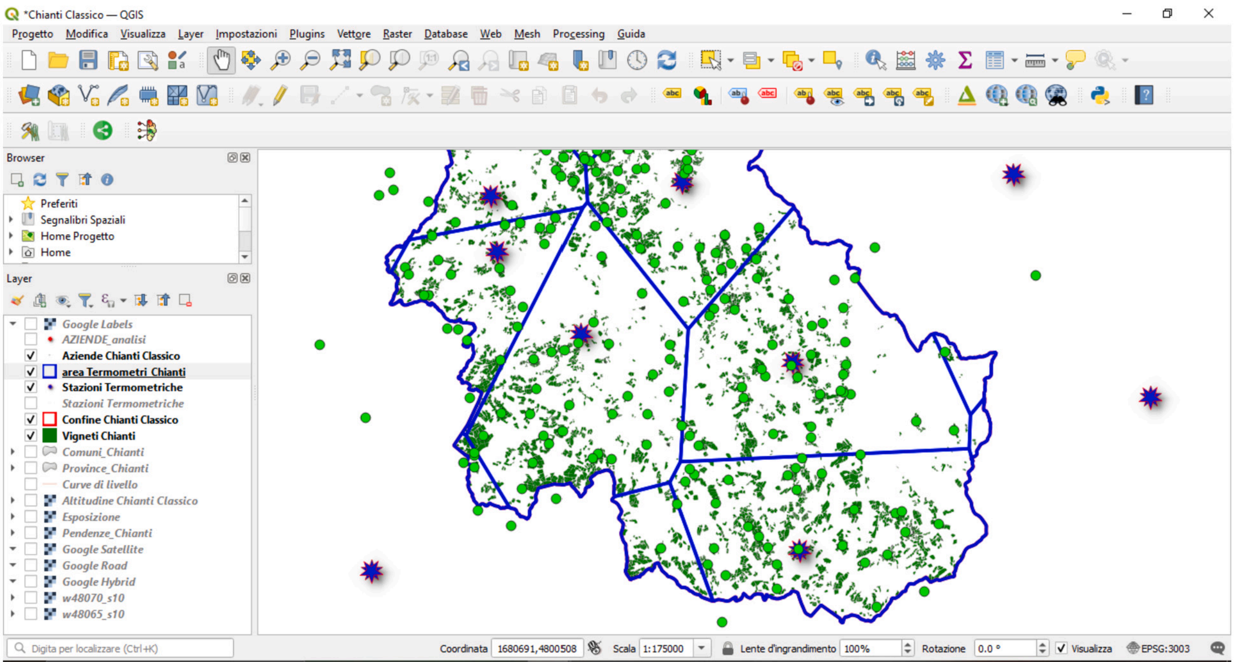


Fig. 4. QGIS; thermometer stations and polygons of Voronoi.

In QGIS, the Geoneous Polygons function has been used by the Geometry Tools, *which consists of an algorithm that generates a polygonal* vector containing the corresponding Voronoi polygons, starting from a vector of points. These polygons represent the area of influence of each detection station, defined as the part of the plane closest to that point compared to any other as shown in Fig. 4.

Appendix B shows the map of the area of influence thermometric station relating to the Chianti Classico area. Through this map, it was possible to identify, in a graphic way, the reference thermometric stations for each winery. The monitoring stations were assigned to the companies located within the relevant polygon of Voronoi. This relationship was considered to attribute the thermal indices of Winkler, Huglin, Fregoni and Freshness calculated to each company examined.

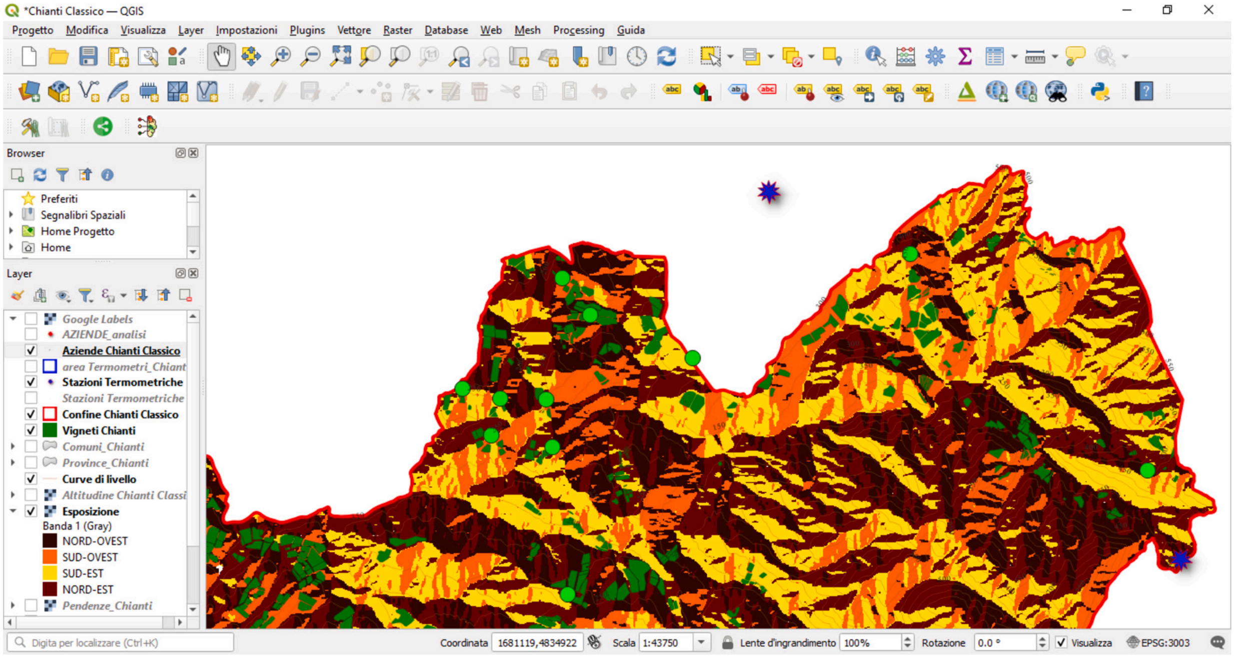


Fig. 5. QGIS; exposure of morphological slopes in the San Polo in Chianti area.

Map of altitudes The digital terrain elevation model file, imported to the QGIS, is represented by the raster model, in which each pixel is associated with a terrain elevation value they have been. Through raster mixing tools, it was possible to merge the two layers into a single file. Subsequently, using raster extraction tools, it was possible to cut out the DEM for the area of interest, using the boundaries of Chianti Classico as a mask. The morphology of the territory has been improved with the contour lines, lines that connect points at the same altitude. Starting from the Digital Terrain Elevation Model, with the raster tool of Extraction of Contour Lines, it was possible to create a vector containing the isoipse with an altitude range of 50 m. The elevation map (Appendix C) has been refined by changing the properties of the raster layer: Symbology - Display type - Single band false color - Interpolation: Linear - Color scale: Grey - Fashion: Equal range - 11 Classes (>150 m a.s.l.; 150-200 m a.s.l.; 200-250 m a.s.l.; 250-300 m a.s.l.; 350-350 m a.s.l.; 350-400 m a.s.l.; 400-450 m a.s.l.; 450-500 m a.s.l.; 500-600 m a.s.l.; 600-700 m a.s.l.; > 700 m a.s.l.).

Appendix C shows the map of altitudes relating to the Chianti Classico area. Through this map, it is possible to observe that the territory varies between an altitude of 150 m to 900 m. The territory has hills of low relief along the north, west and south border. To the east there is a chain of mountains that divides Chianti from Valdarno; moreover, in the central west there is a ridge that divides the center of Chianti from the outermost part. The presence of these mountains coincides with wooded areas. The vineyards in Chianti Classico are mostly located on the slopes of the hills at an altitude between 200 m and 450 m. Various characteristic areas can be identified. The area of San Casciano, Vagliagli and Castelnuovo are mainly in a hilly area at low altitude. The Greve area is situated in a valley enclosed in the hills. On the slopes of this is located Lamole to the east and Montefioralle to the west. In the central part there are Radda and Gaiole surrounded by hills of medium altitude. Finally, it is possible to note that the areas of Panzano, Castellina and San Donato in Poggio are located at the top of the hilly ridge.

Exhibition map Through the raster analysis tool with the support of the GDAL Plugin, Geospatial Data Abstraction Library, the exposure of the territory was calculated with the Digital Terrain Model as input. In the raster of the exposure it is composed of square cells of size 10 m in reality, to which is associated a value from 0 to 360 that expresses the direction of the slope, starting from the North (0°) and continuing clockwise. The Exposure Map (Appendix D) has been refined by changing the properties of the layer: Symbology - Display Type - Single band false color - Interpolation: Discreet - Color Scale: Grey - Fashion: Range equal to 4 Classes. In total 4 significant exposures were identified as shown in Fig. 5:

- **0-90° North-West;**
- **90°-180° Southwest;**
- **180°-270° South-East;**
- **270°-360° North-East.**

Appendix D shows the Exhibition Map relating to the Chianti Classico area. Through this map, it is possible to see how the Chianti Classico territory turns out to be a very varied hilly trend, presenting exposure of the slopes towards the four cardinal directions. Thanks to this, it is possible to understand the trend of each hill with an accuracy of 10 m for a change in exposure. On

Stazioni Termometriche	IW_2011	IW_2012	IW_2013	IW_2014	IW_2015	IW_2016	IW_2017	IW_2018	IW_2019	IW_2020	IW_2021
BARBERINO	2062.2	2062.2	1889.5	1807.4	2024.7	1911.9	2178.9	2088.9	2121.7	2015.9	1965.5
CASTELLINA						1647.7	1830.7	1855.3	1738.6	1641.0	1645.2
CASTELNUOVO	2024.4	2024.4	1871.7	1773.6	1985.5	1895.1	2097.3	2096.7	1997.8	1932.5	1912.5
GAIOLE	1655.6	1655.6	1602.2	1459.3	1646.8	1564.6	1739.9	1841.7	1706.0	1606.9	1546.2
GREVE	1942.7	1942.7	1819.2					2332.9	1957.8	1868.9	1874.2
LAMOLE	1720.4	1720.4	1623.6	1552.7	1738.1	1683.1	1891.9	1938.6	1810.7	1742.0	1727.0
P.CROCE						1773.4	1972.8	2023.9	1860.6	1793.7	1783.8
SANBUCA	2180.3	2180.3	2048.2	1938.9	2129.3	2038.8	2194.4	2248.4	2140.8	2017.2	2047.4
MONTESPERTOLI	2059.2	2059.2	1994.5	1843.8	2050.2	1973.0	2027.4	2164.3	2020.9	1922.5	1870.8
P.CASCIANO	2101.9	2101.9	2008.4	1908.4	2108.5	2018.2					
SCORGIANO	1893.8	1893.8	1825.4	1735.4	1913.7	1747.4	2002.5	2080.1	1977.8	1874.6	1879.2
SIENA P.VENTO								2354.8	2224.7	2173.5	2116.7
TAVARNELLE	1986.4	1986.4	1864.0	1761.7	1976.0	1887.5	2386.8	2175.5	2166.5	2015.9	2015.7
B.AGNANO								1879.4	1745.8	1635.6	1630.1
MONTEVARCHI								2210.7	2071.8	2000.5	1960.6
S.DONATO	2022.6	2022.6	1902.0								

Fig. 6. Winkler index computation.

the basis of these observations, it is possible to graphically identify the areas with a more favorable or less favorable exposure for the development of the vines. The Chianti Classico vineyards are mainly located on the south-facing slopes.

Slope map Through the raster analysis tool with the support of the GDAL Plugin, Geospatial Data Abstraction Library, the slopes of the territory were calculated with the Digital Terrain Model at the entrance. The algorithm calculates the angle of inclination of the terrain from a DEM raster, expressing the slope in degrees. The Slope Map (Appendix E) has been finalized by changing the properties of the layer: Symbology - Display Type - Single band false color - Interpolation: Discreet - Color scale: Grey - Fashion: Range equal to 6 Classes.

The slopes have been classified into:

$>2\%$ Plain; 2-10%; 10-20%; 20-30%; 30-45%; $>45\%$.

Appendix E shows the map of the slopes of the Chianti Classico area in scale. Through this map, it is possible to recognize the intensity of slope present in the various hills and identify the flatter areas. You can see that most Chianti Classico vineyards are located on slopes between 10% and 20%. To conclude, all the maps produced with the help of QGIS are a support to the understanding of the territory in its morphological characteristics. Thanks to the precision of the layers loaded into the system, the analysis could be carried out with a vision scale focused on each individual plot of vineyard, to understand its physical characteristics and manage operations in the best possible way. The georeferenced data also allow to associate alphanumeric information to the layers, so each winery could create a cartographic database of each situation monitored in the field, in order to increase the management efficiency of the cultivated areas.

5.1.1. *Thermal analysis*

To carry out the climatic analysis in the Chianti Classico territory, the most important temperature factor was considered. Through the Winkler, Huglin and Fregoni indices it was possible to synthesize the thermal trend during the season, to compare several years and different areas. In the archive of the Tuscany Region Hydrological System, the data relating to the daily temperature measurements recorded in the meteorological stations of interest for Chianti Classico have been downloaded. Then a dataset was created consisting of the time series of maximum and minimum temperatures recorded from 01/01/2011 until 31/12/2021 for 16 thermometric stations. Some of them had missing data for a long time. Therefore, the years when there was incompleteness of measurements were discarded. There are an average of 10 measuring stations for each year. If a few days presented missing data, these points were replaced with the average values of the previous and subsequent days.

Winkler index calculation The Winkler Index represents the sum of the average temperatures, scaled by 10, measured from April until October, discarding values below 10 °C. The calculation of this index was carried out through a code on the Rstudio calculation software, starting from a dataset composed of 2,354 maximum and minimum temperature values, measured in 16 thermometric stations distributed in Chianti Classico. In Fig. 6 of calculated values of the Winkler Index for the 2011-2021 years relating to the thermometric stations of interest for Chianti Classico was computed.

On the basis of the classification of vineyards, using the Winkler Index, the Chianti Classico territory can be classified from temperate to very hot zone. By analyzing the average Winkler Index, it is possible to see the different climatic zones within the territory. It is possible to note that the areas adjacent to the detection station of Badia Agnano, Scorgiano, Poggio alla Croce, Lamole, Gaiole and Castellina fall within the temperate-warm zone. The areas adjacent to the remaining stations fall into the hot zone. The area with the lowest thermal sum is in Gaiole, while the one with the largest is in Siena. Taking into consideration the Winkler Index relating to the measuring station located in Gaiole, it is possible to define that the year with lower temperatures was 2014, with such as to consider the temperate zone; while the year with the highest temperatures was 2018. With regard to the measuring station located in Tavarnelle, the highest value of the Winkler Index was recorded in 2017, equal to $IW = 2387\,°C$, such as to bring this area into a very hot area. The average value of the Winkler Index, calculated for the period 2011-2021 in the Chianti Classico area,

Stazioni Termometriche	IH_2011	IH_2012	IH_2013	IH_2014	IH_2015	IH_2016	IH_2017	IH_2018	IH_2019	IH_2020	IH_2021
BARBERINO	2425.2	2363.5	2170.9	2047.2	2384.2	2242.5	2530.6	2392.0	2394.0	2404.1	2308.5
CASTELLINA						2064.8	2259.3	2222.1	2091.9	2109.0	2077.9
CASTELNUOVO	2597.6	2564.7	2328.8	2189.8	2515.7	2405.4	2698.8	2605.0	2502.6	2533.0	2504.5
GAIOLE	2388.8	2326.7	2189.4	2007.9	2330.2	2216.5	2509.6	2478.1	2347.6	2358.3	2269.8
GREVE	2568.2	2474.2	2294.3					2892.8	2502.5	2513.8	2505.1
LAMOLE	2095.9	2031.1	1906.1	1831.4	2097.1	2001.4	2335.9	2313.8	2164.1	2209.3	2154.2
P.CROCE						2141.5	2355.4	2339.1	2171.5	2221.1	2165.0
SANBUCA	2638.1	2619.4	2431.3	2279.5	2583.9	2465.2	2646.2	2641.5	2526.1	2495.6	2506.3
MONTESPERTOLI	2764.4	2683.9	2546.8	2355.8	2679.0	2592.3	2694.4	2729.6	2602.5	2585.6	2515.2
P.CASCIANO	2621.9	2586.8	2411.8	2274.7	2589.2	2481.9					
SCORGIANO	2505.7	2460.7	2292.9	2183.1	2459.5	2226.4	2674.6	2633.5	2528.3	2527.4	2514.7
SIENA P.VENTO								2719.2	2579.8	2637.5	2541.3
TAVARNELLE	2351.3	2307.3	2139.5	1999.3	2319.1	2190.6	2726.2	2488.3	2456.4	2423.6	2397.6
B.AGNANO								2533.6	2394.1	2376.5	2363.4
MONTEVARCHI								2710.7	2571.1	2597.7	2545.4
S.DONATO	2481.9	2454.1	2254.8								

Fig. 7. Computed values for the Huglin index for the various stations considered in our study.

Stazioni Termometriche	IF_2011	IF_2012	IF_2013	IF_2014	IF_2015	IF_2016	IF_2017	IF_2018	IF_2019	IF_2020	IF_2021
BARBERINO	0.0	0.0	0.0	0.0	0.0	0.0	512.0	565.2	0.0	855.9	0.0
CASTELLINA						0.0	2543.4	915.9	305.7	1509.0	0.0
CASTELNUOVO	861.8	335.9	399.8	664.8	984.3	762.2	3366.9	1726.4	408.7	1987.5	0.0
GAIOLE	3118.2	5008.8	5179.9	5611.2	7453.8	4507.0	6784.0	2422.5	5752.8	3775.2	7855.5
GREVE	2272.5	1430.4	773.4					2800.2	855.8	3005.8	1889.6
LAMOLE	0.0	0.0	0.0	0.0	984.0	0.0	2565.0	1005.6	314.7	1266.0	0.0
P.CROCE						0.0	500.2	580.6	261.8	1651.2	0.0
SANBUCA	0.0	297.5	0.0	0.0	0.0	0.0	296.4	722.8	329.7	1383.6	0.0
MONTESPERTOLI	1503.0	1957.0	1302.3	3254.4	4228.0	2727.6	3305.7	2403.5	1326.6	2673.0	1878.8
P.CASCIANO	0.0	322.7	0.0	324.0	0.0	0.0					
SCORGIANO	2298.5	1370.0	1683.6	2165.4	3388.5	1981.5	4624.8	2673.6	1342.8	2144.0	1364.7
SIENA P.VENTO								344.0	0.0	1307.6	0.0
TAVARNELLE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	299.2	0.0	900.9	0.0
B.AGNANO								2999.4	4803.0	2772.6	7329.0
MONTEVARCHI								2158.0	407.7	1610.4	447.1
S.DONATO	0.0	591.6	0.0								

Fig. 8. Fregoni index calculation.

is equal to $IW = 1934\,°C$, thus falling into a warm temperate zone. This turns out to be favorable for the ripening of Sangiovese grapes.

Huglin index calculation The Huglin Index takes into account, in the thermal sum, both the average and maximum temperature, measured from April to September. The calculation of this index was carried out through a code on the Rstudio calculation software, starting from a dataset composed of 2013 maximum and minimum temperature values, measured in 16 thermometric stations distributed in Chianti Classico, excluding measurements not detected.

In Fig. 7, the table of calculated values of the Huglin Index for the 2011-2021 Years relating to the thermometric stations of interest for Chianti Classico is reported.

It is possible to note that the value of the Huglin Index, for the Chianti Classico territory, falls within a range between a minimum of $IH = 1831\,°C$, recorded in Lamole in 2014, and a maximum of $IH = 2892\,°C$, measured in Greve in 2018. In the period 2011-2021 this index reported the following average values calculated for the relevant stations. The value indicated for a good ripening of the grapes is equal to $IH = 2100\,°C$. Only the area adjacent to the thermal detection station located in Lamole has a Huglin index value close to the indicated threshold, while all other areas have relatively higher values. The average value calculated for the Huglin Index in the Chianti Classico territory is $IH = 2404\,°C$.

Fregoni index calculation The Fregoni index evaluates the thermal behavior in September multiplied by a coefficient that takes into account days with temperatures below 10 °C. The calculation of this index was carried out through a code on the Rstudio calculation software, starting from a dataset consisting of 330 maximum and minimum temperature values, measured in 16 thermometric stations distributed in Chianti Classico, excluding undetected measurements. In Fig. 8 we present the calculated values of the Fregoni Index for the 2011-2021 Years relating to the thermometric stations of interest for Chianti Classico was computed.

No day with temperature below 10 °C was recorded in September, while higher values were influenced by this multiplication factor. The highest value of this index was recorded at the station located in Gaiole in 2021; this area has significantly higher values over the years than the others. It can be noted that areas enclosed between the hills, such as Gaiole and Greve have higher values

Stazioni Termometriche	IFN_2011	IFN_2012	IFN_2013	IFN_2014	IFN_2015	IFN_2016	IFN_2017	IFN_2018	IFN_2019	IFN_2020	IFN_2021
BARBERINO	17.20	15.77	16.86	15.09	14.80	16.02	13.55	16.38	16.36	16.29	16.83
CASTELLINA						14.27	11.66	14.57	13.71	13.94	14.50
CASTELNUOVO	14.95	14.24	13.15	13.48	13.37	14.12	11.88	13.49	13.47	13.75	13.86
GAIOLE	11.49	10.91	10.18	10.85	10.27	11.15	9.77	11.27	10.66	11.03	10.34
GREVE	13.86	13.34	12.91					13.92	12.94	12.92	12.63
LAMOLE	15.62	14.38	14.14	13.89	13.64	14.82	11.97	14.53	13.92	14.23	14.55
P.CROCE						15.77	13.00	15.76	15.26	15.18	15.47
SANBUCA	16.63	15.40	15.08	14.73	14.48	15.58	13.45	15.46	15.37	14.83	15.41
MONTESPERTOLI	13.61	13.40	12.75	12.39	11.96	13.22	12.18	13.09	12.95	12.93	12.56
P.CASCIANO	15.71	14.86	14.46	14.11	14.29	14.98					
SCORGIANO	13.63	13.65	12.38	12.83	11.80	13.19	11.53	13.07	12.76	13.05	13.23
SIENA P.VENTO								16.41	15.86	16.09	16.64
TAVARNELLE	16.81	15.47	15.63	15.07	15.02	16.26	14.69	16.36	16.39	15.79	16.31
B.AGNANO								11.30	10.78	11.44	10.56
MONTEVARCHI								14.07	13.52	14.09	13.39
S.DONATO	15.95	14.62	14.85								

Fig. 9. Freshness index for the Chianti area.

than places located in an external area of low hills, such as Castelnuovo and Tavarnelle. This shows that in the most hilly areas there is a greater temperature range. On average, in the Chianti Classico area, the Fregoni index calculated in the period 2011-2021, is $IF = 1402\,^{\circ}\mathrm{C}$

Calculation of freshness index of nights The Freshness Index of the Nights evaluates the thermal behavior in September considering the average of the minimum temperatures, i.e. the temperatures measured during the night. The calculation of this index was carried out through a code on the Rstudio calculation software, starting from a dataset consisting of 330 maximum and minimum temperature values, measured in 16 thermometric stations distributed in Chianti Classico, excluding undetected measurements. In Fig. 9 we report the computed Freshness Index

On the basis of the classification of the vineyards by means of the Freshness Index it is possible to divide the temperature bands. In Chianti Classico, on average, there are temperate, cool and very cool nights, with an average minimum temperature of 14 degrees. The minimum temperature was recorded in Gaiole in 2017, while the highest minimum temperature was recorded in Barberino in 2011. It is possible to note, with the Freshness Index, that the areas enclosed by the hills have more pronounced minimum temperatures than areas present in the high ridges of the hill or in the lower part of altitude. For example, in 2018, in Gaiole, a valley between pronounced hills, the temperature was $T = 11.3\,^{\circ}\mathrm{C}$, while in Castellina, a hilltop, it was $T = 14.6\,^{\circ}\mathrm{C}$, and in Tavarnelle, a low hill, it was $T = 16.4\,^{\circ}\mathrm{C}$. Through the average Fregoni Index, the areas with very cool nights are Gaiole and Agnano, while those with cool nights are recorded in Castellina, Castelnuovo, Greve, Montespertoli, Scorgiano and Montevarchi; the remaining stations are classified with temperate nights.

Production analysis

The Chianti Classico wine consortium provides a production specification, which establishes certain fees, with the aim of guaranteeing the quality of a wine with a unique character. We therefore proceeded with the distribution of a questionnaire to a sample of the companies, and information were obtained relating to a sample of 55 wineries located in the Chianti Classico area. For a matter of privacy, only some companies have been mentioned that have consented to the dissemination of data referring to their business reality. We further applied a first univariate descriptive statistical analysis, the clustering of wineries and the multivariate analysis will be reported to recognize the possible relationships of the various factors present in this environment. This analysis was carried out using the following datasets: Wineries, Viticulture, Wine production, Wine quality.

5.1.2. Univariate analysis

Univariate analysis was conducted with the intent to understand the behavior of each individual variable within the dataset through frequency distribution, trend measures and graphs.

Wineries The Chianti Classico territory has very diverse from a morphological characteristic so that each winery has its own peculiarity. Tradition and knowledge in producing wine is still a unique value of this territory that is handed down from generation to generation. The production of wine has always been practiced in Chianti since the Middle Ages and the castles and fortresses are testimony to this. The other wineries are born on average in the 80 s, from the most recent founded in 2018 to the oldest founded in 1870. In the territory there are companies of various sizes: small realities are often family-run and manage only a few hectares, while large realities are made up of companies that manage up to 350 hectares of vineyards, without distinction between DOCG and other denominations. From the data collected it was possible to note that the median is equal to 16 ha, while the average of hectares managed is equal to 37.2; The inter–quantile range varies from a minimum of 1.2 ha per hectare up to 40 ha. Beyond this limit, there are outliers, external values, which represent the companies that manage over one hundred hectares of vineyards. The wineries located in the Chianti area, in addition to producing DOCG Chianti Classico wine, also produce DOCG Chianti Classico Riserva and Gran Selezione. In addition, according to different characteristics of the vineyard and operations in the winemaking phase, white and rosé wines are produced. On average, a winery places 7 wine labels on the market, with a maximum of 18 labels.

Some companies apply the strategy of selling grapes directly from certified vineyards, without carrying out winemaking operations to produce a finished wine. In the larger companies, up to 70 people work full-time, while for the smaller ones, the wine business remains a job for only a few units. In the harvest season, the required personnel increases, so the labor required can even double. On average, 9 people work in the wineries. To understand the variety of sizes among the companies present in Chianti, the distribution of turnover recorded in 2022 consisted in more than 1.000.000 euros for 32.7% companies and only 9.6% between 500.000 and 1.000.000 euros.

Activities in the vineyard and in the cellar As far as the processing in the vineyard is concerned, traditional mechanical means are mostly used. It is possible to notice that many operations have remained manual, especially in finishing. Only 9 companies use mechanical means with precision systems for tillage and screw treatments. More than 70% of the respondents reports manual tillage method, with mechanical methods still used in more than 80% of the respondents. Pruning, given the complexity and importance of the operation, is done 100% manually without the aid of mechanical means. The grape harvest is carried out manually to preserve the quality of the product, but for the harvest of large quantities of grapes 13% of companies use the grape harvester. Once harvested, the grapes are transported to the cellar where the selection, selection and discarding of the non-quality product is carried out. This operation is carried out manually, only 15% use automatic screening systems or optical selection. With vinification and aging, the transformation from grapes into wine takes place. The type of container in which fermentation takes place affects the structure and organoleptic characteristics of the wine.The wineries in Chianti Classico use more for this operation, oak barrels French in 30% and stainless steel tanks in 30%. To conclude, the wine is bottled through automatic systems: 76.7% of the companies carry out this activity within their cellar, while the remaining 23.3% outsource it.

Precision instruments 46% of companies can boast of the presence of private meteorological stations located within their vineyards. These tools can guarantee precise and specific data for the plots, important for the management of field operations and for the control of wine quality. A company uses the instrument for the measurements of leaf wetness, useful for managing cultivation operations. In addition, some companies have installed GPS systems on mechanical vehicles, such as tractors and excavators, for remote detection of the position during field work. The meteorological stations of the Tuscany Region cover the whole territory, offering sufficient data, they can be used by the remaining wineries for a good management of the vineyard and operations.

Viticulture The Chianti Classico DOCG wine is produced exclusively with grapes from vineyards registered in the register of the Denomination. In this section the characteristics of these vineyards managed by the wineries present in the territory have been analyzed. On average, a company cultivates 26 hectares of vineyards used for Chianti Classico, for larger companies up to 110 ha while smaller realities up to one hectare for selected grapes. The vineyards for the production of Chianti have been planted on average around 2000, the most recent in 2023 and the most given in 1960. The ampelographic composition of the vineyards is established by the Production Regulations for 80% of Sangiovese, while the remaining 20% remains at the choice of the companies. The grape variety mix is the main ingredients to give a unique character to the wine. The most widely used complementary grape varieties are Cannaiolo nero, Colorino and Merlot. The planting density is on average 5000 plants per hectare, with a maximum of 8000 vines / ha and a minimum of 2200 vines / ha, for less dense plants. The winemakers of Chianti Classico use as a type of breeding the growth of the vine with Tuscan arch, guyot and spurred cordon; Only some companies use the sapling and candlestick type. The management of the vineyards destined for the Chianti Classico DOCG by the wineries of the territory is conducted for almost 59.3% with organic methods, 18.5% is converging to this mode of operation, while 22.2% is not organic. The morphological characteristics of the territory, on which the vineyard is planted, influence the development of the vine and the ripening of the grapes. The Chianti region is mostly hilly and varies between an altitude of 150 m a.s.l. and 900 m a.s.l. According to the Production Regulations, vineyards located in the valley floor and those located at an altitude of over 700 m are excluded from the production of Chianti Classico wine. According to the data provided by the wineries analyzed, no vineyard is located at altitudes below 250 m, and at altitudes above 600 m. It has been deduced, through the analysis, that more than 60% of the vineyards are located at an average altitude between 250 m and 400 m. Taking into account that the territory of Chianti is hilly, another important factor is the exposure to which the slope of the hill is addressed. This factor affects the intensity and amount of reception of solar radiation. The Chianti Classico vineyards are exposed more to the South, 35% to the South-East and almost 30% to the South-West. A vineyard is less likely to be exposed to the North, with more unfavorable exposure to the North-West. Moreover it was possible to gather information concerning the texture of the soil on which the grapes ripen for the production of Chianti wine is 70% composed of medium texture, 22.9% composed of clays, 4.2% of sands. As far as stoniness is concerned, in the vineyards destined for Chianti Classico there is an average stoniness for 50%, for 4% a low one, while for the remaining 46% by a coarse stoniness. The vineyards destined for Chianti Classico are more characterized, at a chemical level, by the presence of limestone with a pH value greater than 7.5.

Wine production

As far as the production of the wineries is concerned, the wine with the Chianti Classico denomination produced in the 2011-2021 Years was analyzed. For each Year, various factors were analyzed, including the structure of the wine, its acidity, color and alcohol content by volume. The information in this regard was obtained through a discrete scale from 1 to 5 for each attribute, corresponding to:

- **Structure**: 1 “light body”, ..., 5 “large body”;
- **Acidity:** 1 “$>$4.5 g/l”, ..., 5 “$<$ 6.5 g/l”;
- **Color:** 1 “dark”, ..., 5 “alive”;

- **Degrees:** 1 "12%", 2 "12.5%", 3 "13%", 4"13.5%", 5 "14%";

Afterwards, the characteristics of the Chianti Classico DOCG wine produced in the 2011-2021 Years will be reported.

Year 2011 In 2011, an average of 90,000 Chianti Classico DOCG bottles were produced, with a maximum of 470,000 and a minimum of 4,000. The grape yield per hectare of vineyard was 53.3 q/ha on average, with a maximum of 80 and a minimum of 30. Below are the frequency histograms of the relative characteristics of the wine produced in the 2011 year, such as structure, acidity, color and alcohol content.

Year 2012 In 2012, an average of 79,000 Chianti Classico DOCG bottles were produced, with a maximum of 400,000 and a minimum of 3,000. The yield of grapes per hectare of vineyard was 52.1 q/ha on average, with a maximum of 75 and a minimum of 30. Below are the frequency histograms of the relative characteristics of the wine produced in the 2012 Year, such as structure, acidity, color and alcohol content

Year 2013 In 2013, an average of 72,000 Chianti Classico DOCG bottles were produced, with a maximum of 500,000 and a minimum of 4,000. The grape yield per hectare of vineyard was 53.9 q/ha on average, with a maximum of 80 and a minimum of 30. Below are the frequency histograms of the relative characteristics of the wine produced in the 2013 Year, such as structure, acidity, color and alcohol content

Year 2014 In 2014, an average of 45,000 Chianti Classico DOCG bottles were produced, with a maximum of 500,000 and a minimum of 4,000. The yield of grapes per hectare of vineyard was 51.7 q/ha on average, with a maximum of 80 and a minimum of 30. Below are the frequency histograms of the relative characteristics of the wine produced in the 2014 Year, such as structure, acidity, color and alcohol content.

Year 2015 In 2015, an average of 56,000 Chianti Classico DOCG bottles were produced, with a maximum of 500,000 and a minimum of 4,000. The grape yield per hectare of vineyard was 55.9 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2016 In 2016, an average of 69,000 Chianti Classico DOCG bottles were produced, with a maximum of 500,000 and a minimum of 4,000. The grape yield per hectare of vineyard was 54.9 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2017 In 2017, an average of 57,000 Chianti Classico DOCG bottles were produced, with a maximum of 500,000 and a minimum of 4,000. The yield of grapes per hectare of vineyard was 47 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2018 In 2018, an average of 69 thousand Chianti Classico DOCG bottles were produced, with a maximum of 500 thousand and a minimum of 4 thousand. The grape yield per hectare of vineyard was 54.9 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2019 In 2019, an average of 55 thousand Chianti Classico DOCG bottles were produced, with a maximum of 500 thousand and a minimum of 4 thousand. The yield of grapes per hectare of vineyard was 56.4 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2020 In 2020, an average of 71 thousand Chianti Classico DOCG bottles were produced, with a maximum of 500 thousand and a minimum of 4 thousand. The grape yield per hectare of vineyard was 53.9 q/ha on average, with a maximum of 80 and a minimum of 30.

Year 2021 In 2021, an average of 81 thousand Chianti Classico DOCG bottles were produced, with a maximum of 500 thousand and a minimum of 4 thousand. The yield of grapes per hectare of vineyard was 54.2 q/ha on average, with a maximum of 80 and a minimum of 30.

5.1.3. Clustering wineries

The wineries operating in the Chianti Classico territory have uneven characteristics, so it was appropriate to divide them into 3 groups identifying small, medium and large realities. To perform this we applied the well known method of K-Means clustering [21], using as input features the following: turnover range, total cultivated vineyards, wine labels and full-time workers. We set the number of clusters equal to 3.

In Fig. 10 we show the result of the application of K-Means to our dataset. The horizontal axis represents the companies (represented numerically for privacy purposes), while the vertical axis shows the relative value of the factor under analysis. We applied the following coloring legends to the dots represented in the Figure: black marks represent large companies, the green ones represent medium-sized companies, while the red marks represent small companies.

Comparing the plots, presented in Fig. 10, it is possible to see how some companies, medium and large, with the same turnover, manage a smaller number of hectares sometimes employing less human resources. The results might suggest that realities defined as large, through the use of precision means and weather stations, can have greater operational efficiency with less use of resources. Focusing on the graph concerning the labels of Fig. 10, it is also possible to notice that the number of types of wines placed on the market is not strictly related to the level of turnover. The turnover attribute is therefore not the only parameter to be taken into consideration for the grouping of wineries.

Multivariate analysis We further performed a multivariate analysis in order to extrapolate information from the data and identify the correlation between the variables studied previously. Based on the data collected from the sample of wineries participating in the questionnaire, it was interesting to study which factors are linked to each other in order to take them into greater consideration in the management process of the activity. In particular in this way we could understand which attributes would have a direct correlation with the others present and which are independent of each other. To do this, we analyzed the Correlation Matrix, which represents

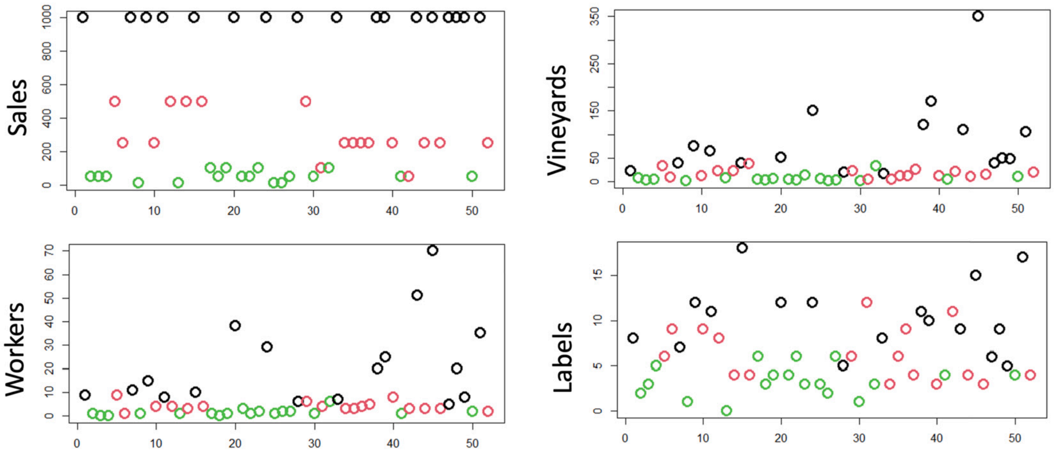


Fig. 10. Scatterplot charts for the 3 clusters obtained with K-Means.

the measure of the linear correlation between vectors of data, by reporting the Pearson linear correlation coefficient between each pair of variables.

Chianti classico DOCG wine production In order to carry out this analysis, it was appropriate to separate the data for the relevant reference years. The data relating to the annual production of wine with the Chianti Classico denomination, the values relating to the thermal indices calculated for the relative year and the data representing the characteristics of the vineyard and the main processes were included in a single dataset.

The dataset has fewer observations than in previous analyses, since rows with missing values were removed due to lost production in that year. Categorical variables were also converted to numerical values.

The attributes analyzed in the correlation analysis are as follows:

- **Quality:** 1 "low", ..., 5 "high";
- **Structure:** 1 "light body", ..., 5 "large body";
- **Acidity:** 1 ">4.5 g/l", ..., 5 "< 6.5 g/l";
- **Color:** 1 "dark", ..., 5 "alive";
- **Degrees:** 1 "12%", 2 "12.5%", 3 "13%", 4"13.5%", 5 "14%";
- **IW** (Winkler Index): numerical value;
- **IH** (Huglin Index): numerical value;
- **IF** (Fregoni Index): numerical value;
- **T_vendemmia** (harvest period): 1 "early", 2 "medium", 3 "late";
- **Yield** (grapes per hectare of vineyard q/ha): numerical value;
- **Altitude:** 150 - "below 250 m a.s.l.", 250 - "250-300 m a.s.l.", 350 - "350-400 m a.s.l.", 400 - "400-450 m a.s.l.", 500 - "500-600 m a.s.l.", 600 - "600-700 m a.s.l.";
- **Exposure:** 1 "North West", 2 "North", 3 "North East", 4 "West", 5 "East", 6 "South East", 7 "South", 8 "South West";
- **Density** (plants per hectare of vineyard): numerical value;
- **Weaving:** 1 "clayey", 2 "medium dough", 3 "sandy", 4 "alberese and marl";
- **Toniness:** 1 "low", 2 "medium", 3 "coarse";
- **Limestone:** 0 "absent", 1 "medium", 2 "high"
- **pH**: 1" < 7.5", 2" >7.5";
- **Treatments:** 1 "manual", 2 "traditional mechanical means", 3 "mechanical means with precision farming systems";
- **Harvest Type** (harvest operation): 1 "manual harvest", 2 "mechanical harvest";
- **Weather Monitoring Type:** 0 "No", 1 "Yes";
- **Method:** 0 "non-organic", 1 "in conversion", 2 "Organic";
- **Watering:** 0 "absent", 1 "watering".

In Fig. 11 the correlation matrix for the data relating to the 2021 Year is presented.

It can be observed that the quality of the end product, expressed with a value from 1 to 5, is in a positive relationship with the other characteristics of the wine. Moreover the cultivation method, between organic, in-conversion and non-organic, is directly linked to the quality of the wine produced. The processes in the vineyard, such as the use of mechanical machines for the harvest, show an inverse relationship with respect to quality. It should be noted that quality has a positive correlation with the presence of precision farming tools, such as the presence of meteorological monitoring stations and treatments with precision machinery. Moreover the Winkler thermal index is positively correlated with the yield of grapes per hectare and with the alcohol content recorded in the wine,

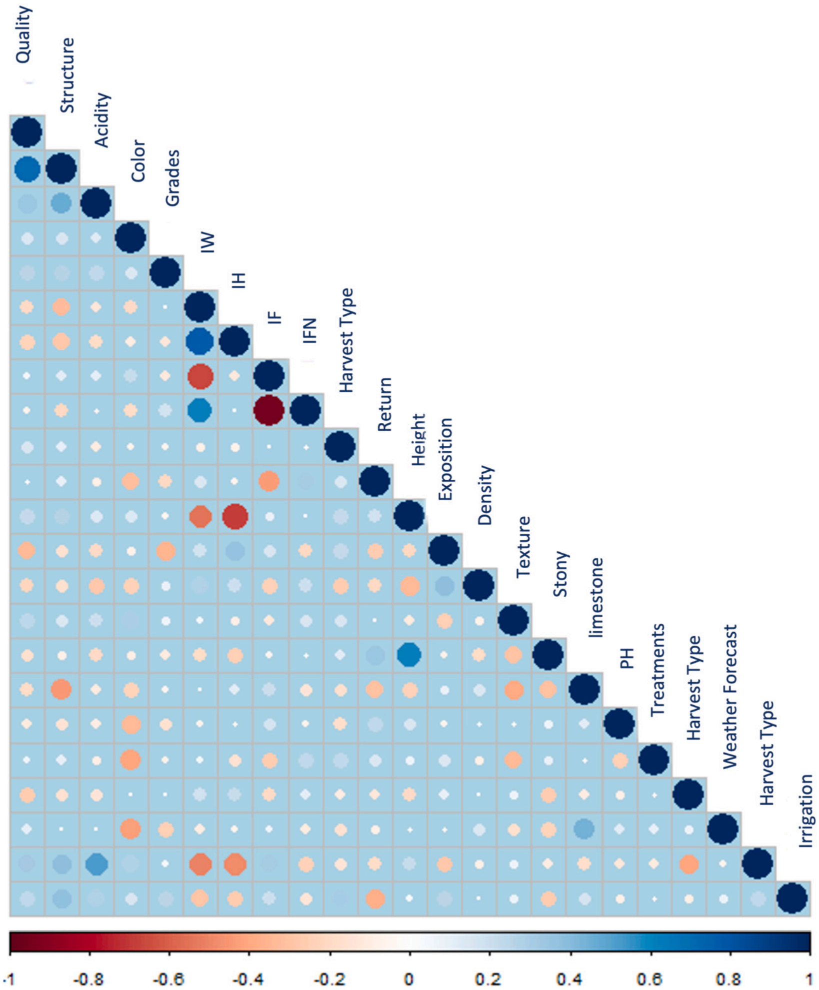


Fig. 11. Correlation matrix for the year 2021.

but negatively with other factors of the finished product such as quality, structure and color. Same can be said for the Huglin index having the same correlation but with a higher intensity. This result indicates that a season with high and prolonged temperatures could lead to lower results. Furthermore the Fregoni index has an opposite behavior compared to the other two indices, given that the temperatures of September must be mild with strong temperature ranges for a correct ripening of the grapes. This is positively correlated with acidity and color factors, but negatively with alcohol levels. While the Night Freshness Index is positively correlated with the structure and color of the wine. Vineyard characteristics such as altitude and soil texture have a positive correlation with wine attributes. Furthermore we can see that exposure is positively correlated with color, acidity and degrees. Furthermore the plant density is inversely proportional to the other factors, this means that denser plants use less available resources producing a lower result. The chemical characteristics of the soil such as the presence of limestone and pH are inversely proportional to the factors of the wine, excluding the presence of limestone which positively affects the alcohol content. To analyze the correlations between the attributes in the decade 2011-2021, the average values of the correlations obtained in each year were calculated. In Fig. 12 the average correlation matrix is reported to evaluate the dependence or not of the factors.

Wineries The attributes analyzed are as follows:

- **Turnover:** 1000 “over 1,000,000 euros”, 500 “between 500,001 and 1,000,000 euros”, 250 “between 250,001 and 500,000 euros”, 100 “between 100,001 and 250,000 euros”, 50 “between 50,001 and 100,000 euros”, 10 “between 10,000 and 50,000 euros”;
- **Total vineyards**: numerical value;
- **Labels**: numeric value;
- **Workers:** numerical value;
- **Harvest:** 1 “manual harvest”, 2 “mechanical harvest”;
- **Treatments:** 1 “manual”, 2 “traditional mechanical means”;

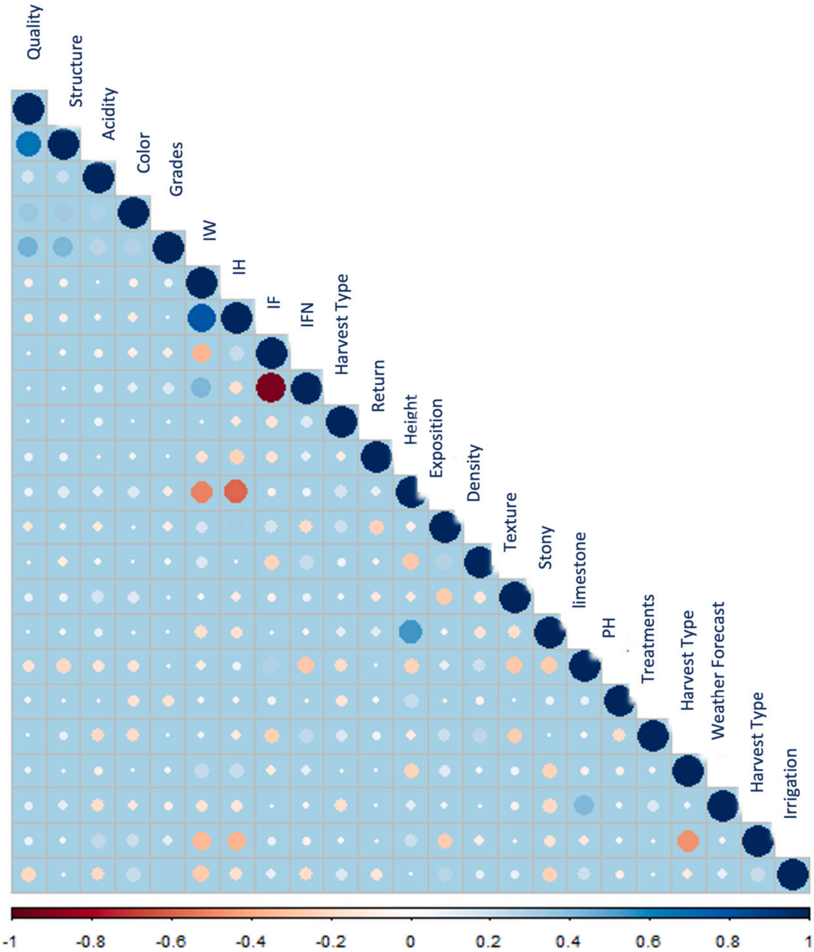


Fig. 12. Average correlation matrix between significant factors, relative to 2011-2021.

- **Method:** 0 “non-organic”, 1 “in conversion”, 2 “Organic”;
- **Rilevamento_meteo:** 0 “No”, 1 “Yes”;
- **Watering:** 0 “absent”, 1 “watering”;
- **Quality:** 1 “low”, ..., 5 “high”.

The correlation matrix was calculated to evaluate the relationships between pairs of attributes related to the generic aspects of the wineries present in Chianti Classico. To assess which factors were related to quality, 11 datasets were formed with the values related to the quality recorded in the years 2011-2021. As done in the previous analysis, the correlation matrix for each year and the data was computed and the average value of the correlation values were reported. Next, the mean matrix is reported to evaluate the dependence between the factors as shown in Fig. 13.

Through the correlation matrix (Fig. 13) it is possible to see how the profit of companies, measured with turnover, is positively correlated with the amount of hectares managed, number of labels produced and full-time workers. The cultivation method, organic or not, is negatively correlated with the use of precision machinery and large extensions of vineyards to be cultivated. This is due to the fact that organic cultivation requires more attention on the method of operation and on the products used. It is interesting to note how the presence of detection stations is positively correlated with factors of the company, demonstrating the fact that they make a positive contribution to the management of operational choices.

Theoretical and practical implications

The study we performed shows that quality is weakly correlated to the characteristics of the company, demonstrating the fact that a quality wine is produced both by small and large realities. A greater correlation with the quality of the fine product is due to the organic cultivation method. In addition, it can be seen that the presence of climate monitoring stations is more correlated with the quality factor. Similarly we can evidence how our study could have practical implications in the understanding of potential future threats and changes to be introduced in the companies structure, in order to tackle and address changes due to the newly climatic conditions.

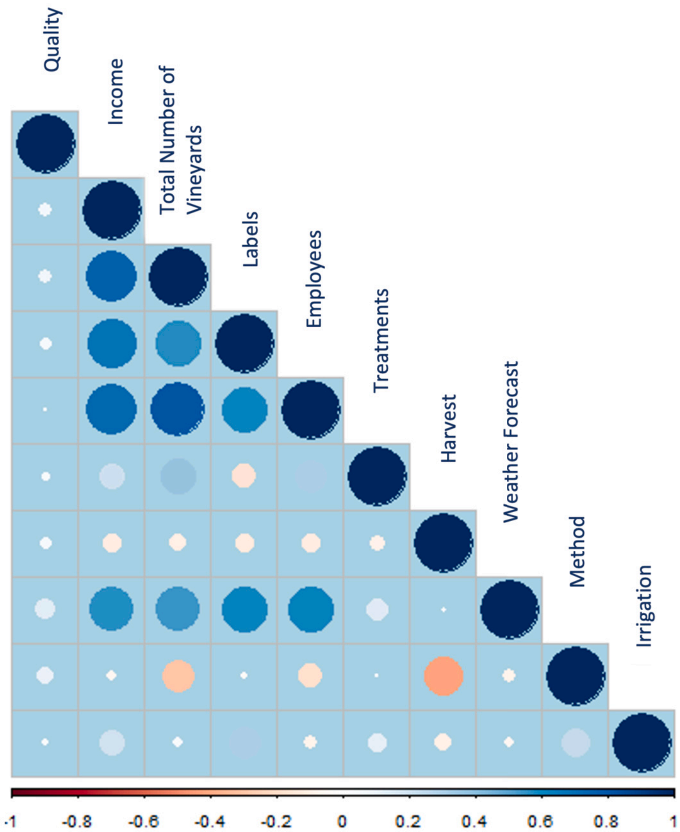


Fig. 13. Average correlation matrix between the significant factors of wineries, relating to 2011-2021.

6. Conclusion

In the paper the topic of precision agriculture referred to the specialized context of viticulture was treated. The vineyard can be considered a living element in which environmental, cultural and anthropic factors are recognized and valued to best express the character of the wine. Precision viticulture is a system based on observation, measurement and response to variables in the field. The case study analyzes the area with Denomination of Controlled and Guaranteed Origin Chianti Classico, a territory suited for the production of wine, located in Tuscany, between the provinces of Siena and Florence in Italy. With the help of the QGIS Open Source Geographic Information System, a spatial analysis was carried out using a cartographic, digital and quantitative approach. This allowed us to analyze the territory in its morphological aspects, including altitude, exposure and slopes, locating the individual wineries. A detailed picture of the characteristics of the different wine-growing areas in the Chianti Classico area was thus created. Thanks to the precision of the layers, loaded into the system, the analysis could be carried out with a vision scale focused on each individual plot of vineyard, to understand its physical characteristics and manage its operations. The georeferenced data also allow to associate alphanumeric information to the layers, so each winery could create a cartographic database of each situation monitored in the field, in order to increase the management efficiency of the cultivated areas. With the aim of evaluating the thermal seasonality in Chianti Classico, the significant thermal bioclimatic indices for viticulture were calculated, such as the Winkler, Huglin, Fregoni and Freshness indexes, derived from the values of daily temperatures measured in the meteorological stations of the Tuscany Region. The average value of the Winkler Index, calculated for the period 2011-2021, allows to classify the Chianti Classico territory on average as a temperate-warm zone, favorable for the ripening of Sangiovese grapes. According to the Huglin Index, it can be said that most of the areas in Chianti Classico have relatively higher thermal values than the theoretical figure for a good ripening of the fruit. Finally, through the Fregoni and Freshness Index, it is possible to demonstrate that the areas most enclosed by hills recorded in September, a greater temperature range than areas on the top of the ridges or tending to the plain, favoring better maturation. Secondly, a survey was carried out on a sample of wineries registered with the Chianti Classico Wine Consortium, to analyze the production of Chianti Classico DOCG wine. Through the questionnaire, data relating to business aspects, viticulture and wine production in the period 2011-2021 were collected. Through statistical methods and tools, the application of technologies and agronomic principles to manage the variability associated with wine production in the Chianti Classico territory was studied. Companies were classified into three groups, based on the attributes of turnover range, number of hectares conducted, number of wine labels traded and number of full-time workers. This has made it possible to identify small, medium and large companies through the clustering method, weighted with respect to these attributes. Finally, through a multivariate statistical analysis, it was possible to identify the linear correlations between the pedoclimatic and production parameters, to recognize the most influential characteristics in wine production. The characteristics of the vineyards, such as altitude and soil texture, have a positive correlation with the attributes related to the wine. Exposure is positively correlated with color, acidity and alcohol content. Moreover it was possible to note how the thermal index of Winkler and Huglin is positively correlated with the yield of grapes per hectare and with the alcohol content recorded in the wine, but negatively with the quality, structure and color of the product. Furthermore the Fregoni index shows a positive relationship with acidity and color factors, but negative with blood alcohol levels and the Night Freshness Index is positively connected with the structure and color of the wine. From the analysis performed we could also see ho quality has a positive correlation with the availability of precision farming tools. It could be therefore suggested said that these technologies are a positive support to the conduct of agricultural activity: this is partly shown by the positive correlation between quality and the presence of precision farming tools. The use of weather stations has proven to be an advantageous contribution to the management of operations and the quality of the finished product. Therefore, it can be interesting for companies, without this technology, to take advantage of the public weather stations, which cover the entire territory offering profitable data to be used in their internal decision making processes and analysis. Limitations of the work mainly include the absence of some information not released by the companies in the survey due to privacy concerns, which we then decided to not include in the set of questions. Moreover, as a future work possibility, we could probably compare the Chianti Area with similar but different Italian areas, implementing strategies of reorganization which have been actually suggested in order to assess the impact of climate change into the final wine production and therefore understand future endeavors of the product. To conclude, this paper aims to show the correlations between each individual aspect, so that each company can act and manage the production processes as a whole, including environmental, viticultural, instrumental and operational aspects, in order to increase efficiency, productivity and long-term profitability. These results, which emerged from the study, will be able to help implement decision-making models in the wine business thanks to the integration of the knowledge provided by professionals working in the sector, such as agricultural operators, agronomists and agricultural entrepreneurs.

CRediT authorship contribution statement

Giovanna Maria Dimitri: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alberto Trambusti:** Investigation, Formal analysis, Data curation.

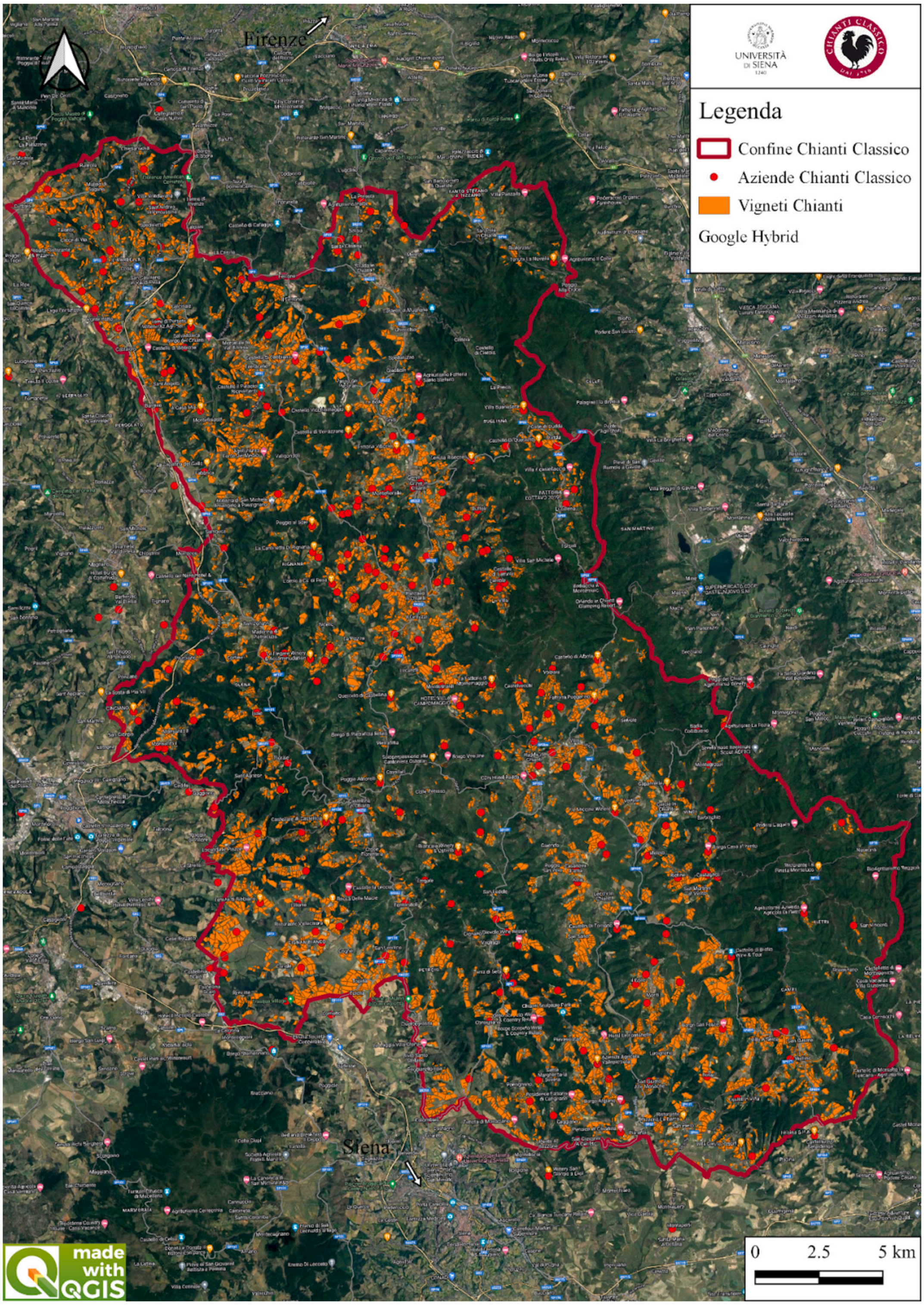
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

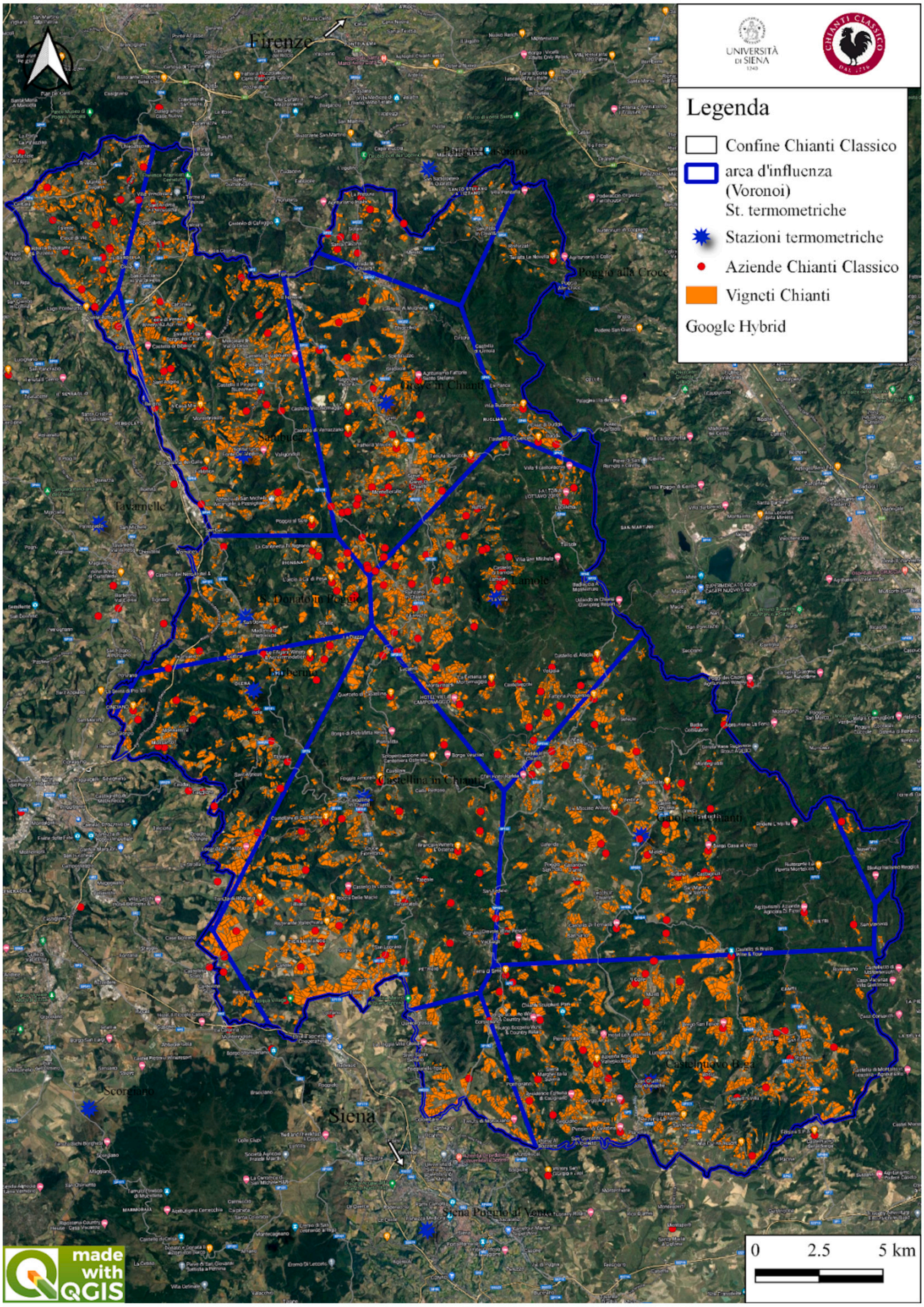
Data availability statement

The data and scripts presented in this work is available at:
https://github.com/GiovannaMariaDimitri/WineQualityMachineLearning.

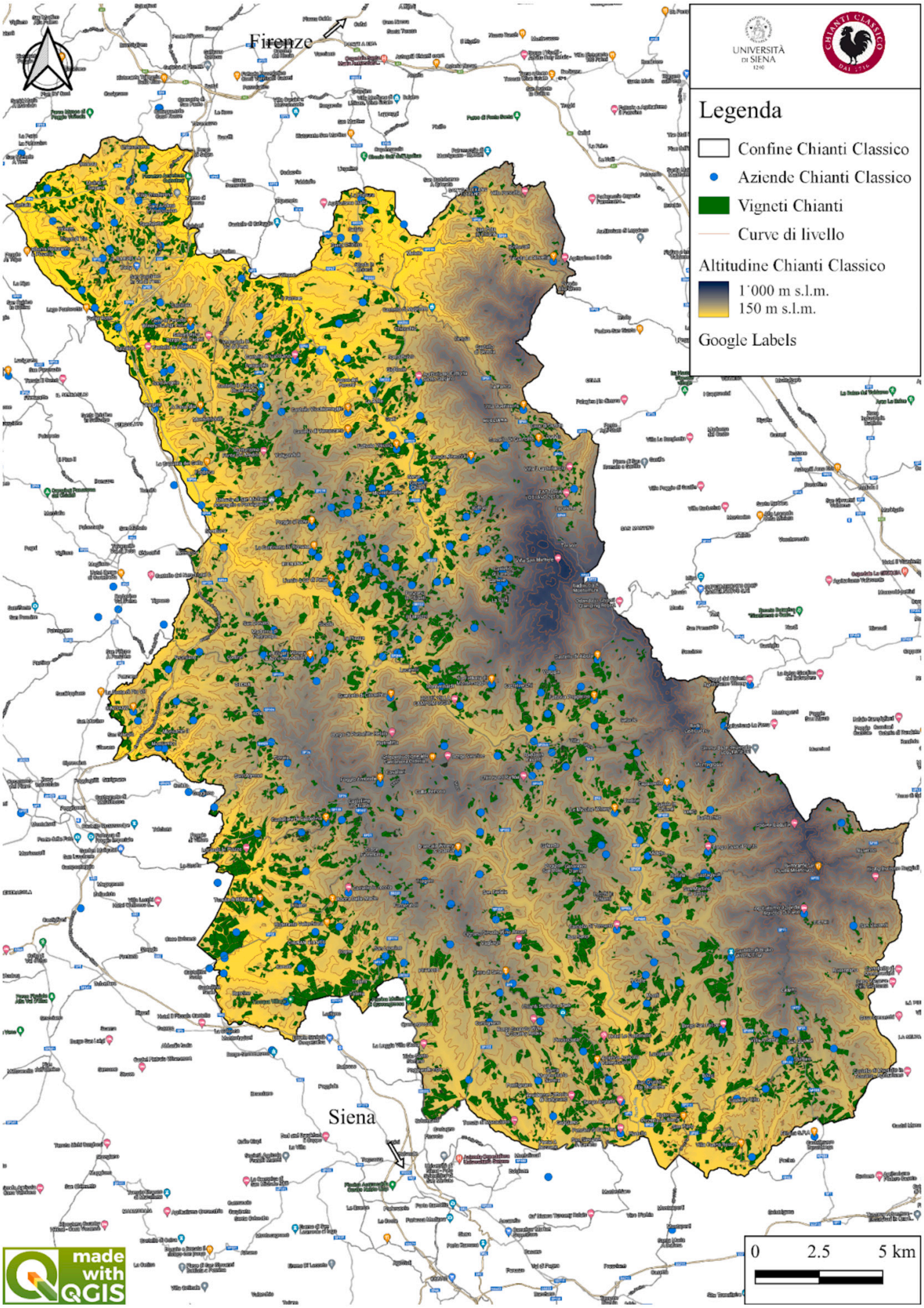
Appendix A. Wineyards map



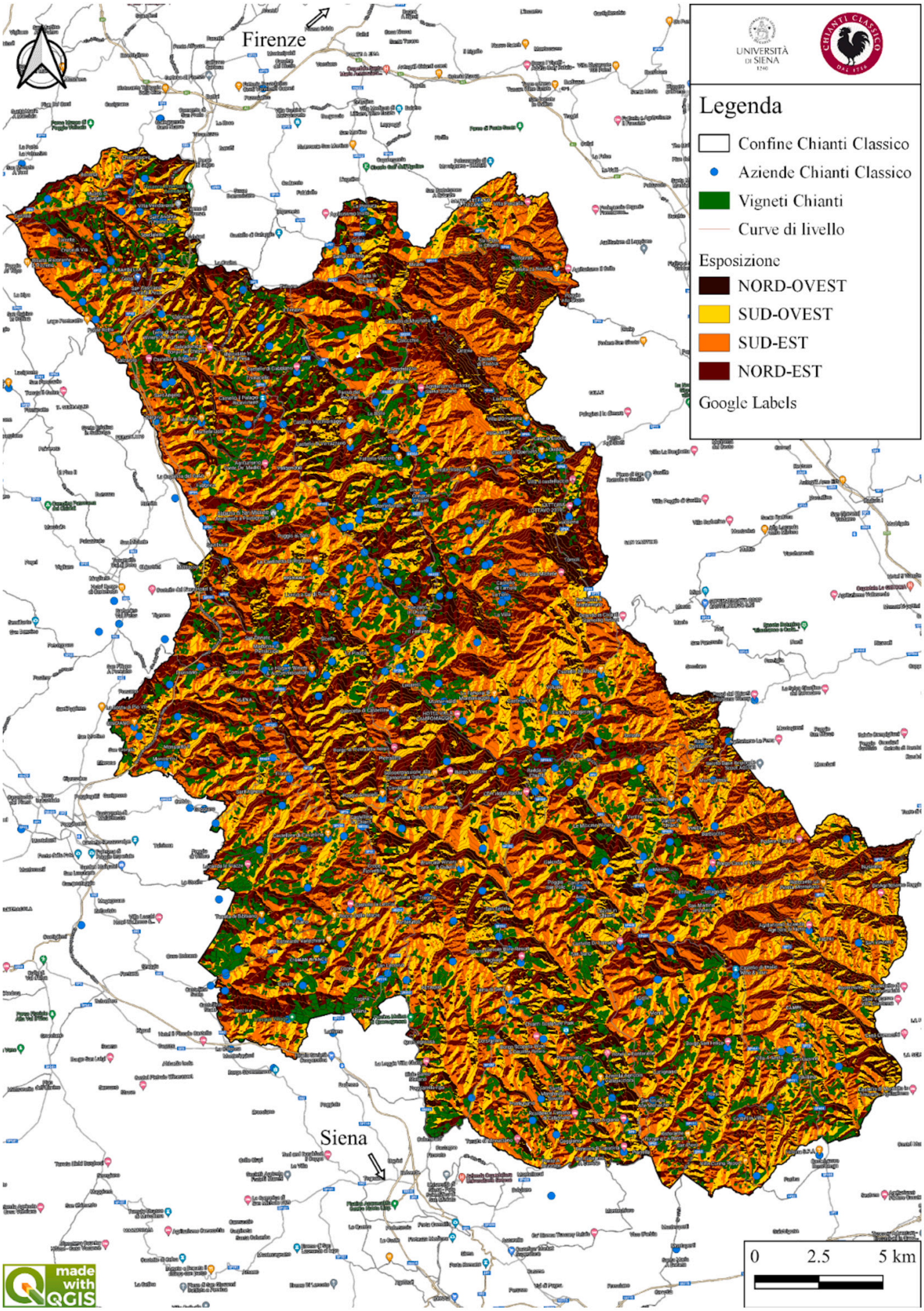
Appendix B. Area of influence map thermometric stations



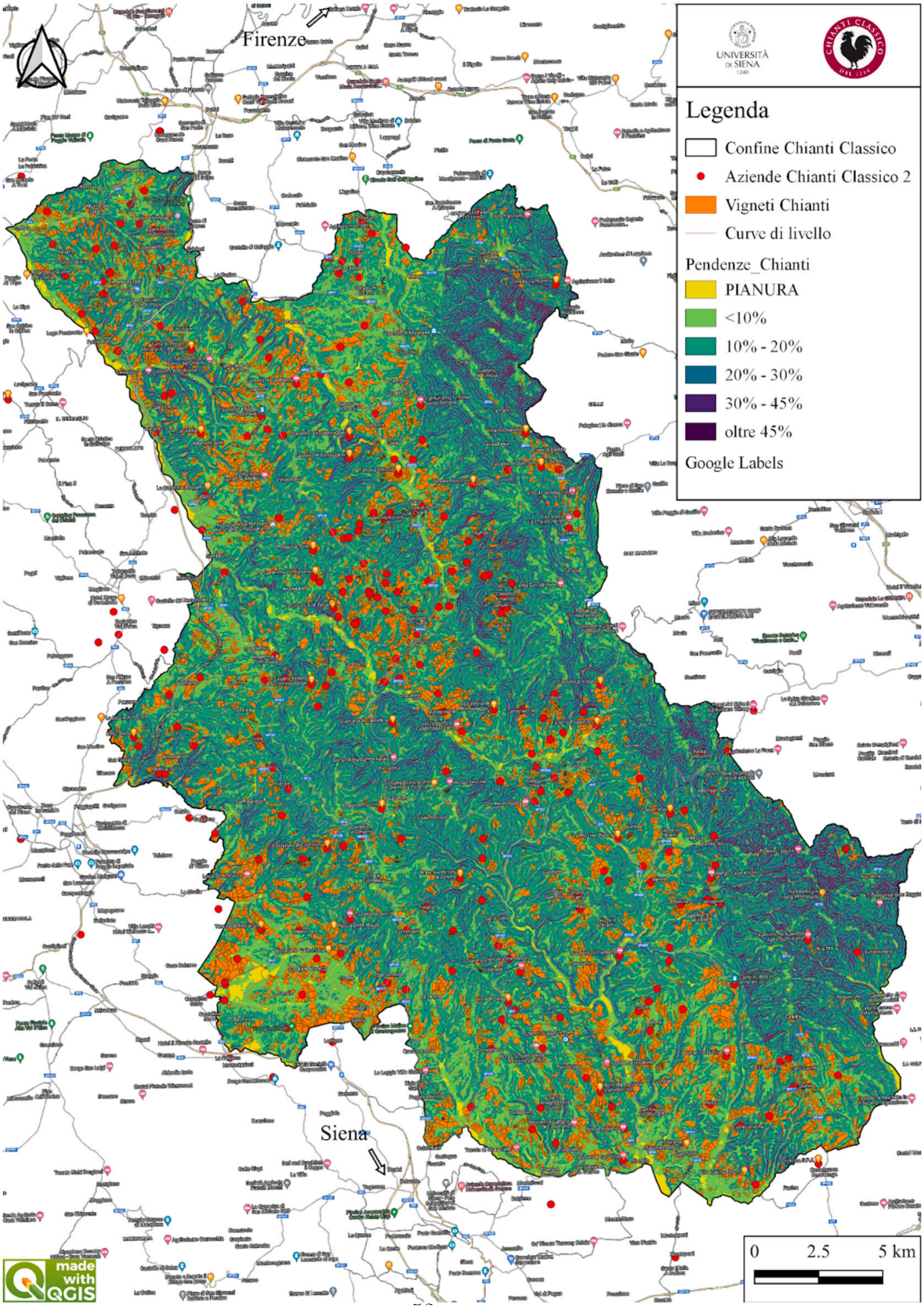
Appendix C. Altitude map



Appendix D. Exhibition charter



Appendix E. Map of slopes



Appendix F. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31648.

References

[1] Helena Albuquerque, et al., Wine landscape and climate change: the state of art, J. Wine Res. 34 (4) (2023) 272–290.
[2] R. Beluhova-Uzunova, D. Dunchev, Precision farming - concepts and perspectives, in: Zagadnienia Ekonomiki Rolnej / Problems of Agricultural Economics, 2019, pp. 142–155.
[3] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, Dionysis Bochtis, Machine learning in agriculture: a comprehensive updated review, Sensors 21 (11) (2021) 3758.
[4] Carmen Berbegal, et al., Climate changes and food quality: the potential of microbial activities as mitigating strategies in the wine sector, Fermentation 5 (4) (2019) 85.
[5] Carmen Berbegal, et al., Climate changes and food quality: the potential of microbial activities as mitigating strategies in the wine sector, Fermentation 5 (4) (2019) 85.
[6] Andrea Borghini, Nicola Piras, Beatrice Serini, Hot grapes: how to locally redesign geographical indications to address the impact of climate change, World Dev. Sustain. 2 (2023) 100043.
[7] F. Burroni, M. Pierucci, Non solo calcare: più attenzione al PH del terreno. Mille Vigne, 2022.
[8] Wendy Cameron, et al., A comparison of the effect of temperature on grapevine phenology between vineyards, Oeno One 55 (2) (2021) 301–320.
[9] Consorzio Vino Chianti Classico, Aziende, Tratto da Chianti Classico, 2014.
[10] Consorzio Vino Chianti Classico, Aziende, Tratto da Chianti Classico, 2023.
[11] Edoardo A.C. Costantini, Roberto Barbetti, Environmental and visual impact analysis of viticulture and olive tree cultivation in the province of Siena (Italy), Eur. J. Agron. 28 (3) (2008) 412–426.
[12] J.C. Dijkerman, Pedology as a science: the role of data, models and theories in the study of natural soil systems, Geoderma 11 (2) (1974) 73–93.
[13] Giovanna Maria Dimitri, Pietro Lió, Drugclust: a machine learning approach for drugs side effects prediction, Comput. Biol. Chem. 68 (2017) 204–210.
[14] Giovanna Maria Dimitri, Lorenzo Parri, Alessandro Pozzebon, Eleonora Vitanza, Ada Fort, Chiara Mocenni, Weair: wearable swarm sensors for air quality monitoring to Foster citizens' awareness of climate change, in: 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), IEEE, 2023, pp. 98–103.
[15] Giovanna Maria Dimitri, Simeon Spasov, Andrea Duggento, Luca Passamonti, Pietro Lió, Nicola Toschi, Multimodal and multicontrast image fusion via deep generative models, Inf. Fusion 88 (2022) 146–160.
[16] André Fonseca, Hélder Fraga, João A. Santos, Exposure of Portuguese viticulture to weather extremes under climate change, Clim. Serv. 30 (2023) 100357.
[17] M. Fregoni, L'indice bioclimatico di qualità Fregoni. Terrior Zonazione Viticoltura. Phytoline, 2003.
[18] Emma Gaitán, Mª Rosa Pino-Otín, Using bioclimatic indicators to assess climate change impacts on the Spanish wine sector, Atmos. Res. 286 (2023) 106660.
[19] R. Gebbers, V. Admchuck, Precision agriculture and food security, Science 829 (2010).
[20] Grazia Diego, et al., Grapes, wines, and changing times: a bibliometric analysis of climate change influence, Aust. J. Grape Wine Res. (2023) 2023.
[21] John A. Hartigan, Manchek A. Wong, Algorithm as 136: a k-means clustering algorithm, J. R. Stat. Soc., Ser. C, Appl. Stat. 28 (1) (1979) 100–108.
[22] Bruno P. Holzapfel, et al., Manipulating the postharvest period and its impact on vine productivity of semillon grapevines, Am. J. Enol. Vitic. 57 (2) (2006) 148–157.
[23] Jennifer M. Huggett, Geology and wine: a review, Proc. Geol. Assoc. 117 (2) (2006) 239–247.
[24] P. Huglin, Nouveau mode d'évaluation des possibilités héliothermiques d'un milieu viticole, in: Ministère de l'Agricolture de l'industrie Alimentaire, Symposium Int, sur l'ecologie de la vigne, Constanca, 1978.
[25] George Koufos, et al., Viticulture–climate relationships in Greece: the impacts of recent climate trends on harvest date variation, Int. J. Climatol. 34 (5) (2014) 1445–1459.
[26] M. Leder, La Vite e il Clima. Tratto da Quattrocalici, 2019.
[27] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, Dionysis Bochtis, Machine learning in agriculture: a review, Sensors 18 (8) (2018) 2674.
[28] S. Madhulatha, An Overview on Clustering Methods, 2012.
[29] Markopoulos Theodoros, et al., Wine quality control parameters and effects of regional climate variation on sustainable production, Sustainability 15 (4) (2023) 3512.
[30] Vishal Meshram, Kailas Patil, Vidula Meshram, Dinesh Hanchate, S.D. Ramkteke, Machine learning in agriculture domain: a state-of-art survey, Artif. Intell. Life Sci. 1 (2021) 100010.
[31] Marco Moriondo, G.V. Jones, Benjamin Bois, C. Dibari, R. Ferrise, G. Trombi, M. Bindi, Projected shifts of wine regions in response to climate change, Clim. Change 119 (2013) 825–839.
[32] Noti Valerio, GIS Open Source per geologia e ambiente: Analisi e gestione di dati territoriali e ambientali con QGIS, Dario Flaccovio Editore, 2014.
[33] U. S. House of Representatives, An act to reform, extend and repeal certain agricultural research. Title IV – New research, extension and education initiatives, 1997, p. 2534.
[34] A. Paliotti, S. Poni, O. Silvestroni, La nuova viticoltura. Innovazione tecniche per modelli produttivi efficienti e sostenibili, 2015.
[35] A. Palliotti, S. Poni, O. Silvestroni, Manuale di viticoltura, 2018.
[36] R. Russell, M. Paterson, et al., Predominant mycotoxins, mycotoxigenic fungi and climate change related to wine, Food Res. Int. 103 (2018) 478–491.
[37] Aleksandre Petriashvili, et al., The impact of climate change on the sustainability of wine production and the structure of its consumption in Czechia, Heliyon 9 (7) (2023).
[38] Sandro Sacchelli, et al., A mix-method model for adaptation to climate change in the agricultural sector: a case study for Italian wine farms, J. Clean. Prod. 166 (2017) 891–900.
[39] Michael G. Sanderson, et al., The probability of unprecedented high rainfall in wine regions of northern Portugal, Clim. Serv. 30 (2023) 100363.
[40] Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, Chin-Teng Lin, A review of clustering techniques and developments, Neurocomputing 267 (2017) 664–681.
[41] Abhinav Sharma, Arpit Jain, Prateek Gupta, Vinay Chowdary, Machine learning applications for precision agriculture: a comprehensive review, IEEE Access 9 (2020) 4843–4873.
[42] Ottavia Spiga, Vittoria Cicaloni, Giovanna Maria Dimitri, Francesco Pettini, Daniela Braconi, Andrea Bernini, Annalisa Santucci, Machine learning application for patient stratification and phenotype/genotype investigation in a rare disease, Brief. Bioinform. 22 (5) (2021) bbaa434.
[43] Elham Tayari, Amin Reza Jamshid, Hamid Reza Goodarzi, Role of gps and gis in precision agriculture, J. Sci. Dev. Res. 2 (3) (2015) 157–162.
[44] J. Tonietto, A. Carbonneau, A multicriteria climatic classification system for grapegrowing regions worldwide, Agric. For. Meteorol. (2004).
[45] János P. Tóth, Zsolt Végvári, Future of winegrape growing regions in Europe, Aust. J. Grape Wine Res. 22 (1) (2016) 64–72.
[46] Cornelis Van Leeuwen, Philippe Darriet, The impact of climate change on viticulture and wine quality, J. Wine Econ. 11 (1) (2016) 150–167.
[47] A. Vercesi, A. Castognoli, P. Dosso, Metodologia di caratterizzazione agrometereologica dei territori, Inf. Agrar. (2003) 13–16.
[48] M. Vieri, L'evoluzione tecnica e tecnologica nella moderna viticoltura imprenditoriale, Accademia dei Georgofili, 2003.
[49] Eleonora Vitanza, Giovanna Maria Dimitri, Federico Bizzarri, Chiara Mocenni, Investigating the impact of extreme rainfall events on individual perception of climate change, in: 2023 International Symposium on Nonlinear Theory and Its Applications NOLTA2023, 2023, pp. 202–205.

[50] Eleonora Vitanza, Giovanna Maria Dimitri, Chiara Mocenni, A multi-modal machine learning approach to detect extreme rainfall events in Sicily, Sci. Rep. 13 (1) (2023) 6196.

[51] G. Voronoi, Nouvelles applications des paramètres continus à la théorie des formes quadratiques, J. Reine Angew. Math. (1907).

[52] Alexandre M.J.-C. Wadoux, Budiman Minasny, Alex B. McBratney, Machine learning for digital soil mapping: applications, challenges and suggested solutions, Earth-Sci. Rev. 210 (2020) 103359.

[53] Alexandre M.J.-C. Wadoux, Alessandro Samuel-Rosa, Laura Poggio, Vera Leatitia Mulder, A note on knowledge discovery and machine learning in digital soil mapping, Eur. J. Soil Sci. 71 (2) (2020) 133–136.

[54] Michael A. White, et al., Extreme heat reduces and shifts United States premium wine production in the 21st century, Proc. Natl. Acad. Sci. 103 (30) (2006) 11217–11222.

[55] A. Winkler, F. Cook, W.M. Kliewer, L.A. Lider, General Viticolture, Uiversity of California Press, 1974.

[56] Xueqin Zhu, et al., A model-based assessment of adaptation options for chianti wine production in Tuscany (Italy) under climate change, Reg. Environ. Change 16 (2016) 85.