Statistical analysis of semi-natural areas and unsupervised climate clustering for the Alentejo region, Portugal

Filipa Inês Vilhena
ines.vilhena@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

June 2022

Abstract

Vineyards require specific ecological contexts for all stages of their life cycle, such as appropriate temperatures, sufficient water availability and some pest control, to maximize wine yield and quality. In Mediterranean Portugal, the region Alentejo is a known wine producer, with very high temperatures in the Summer and a worsening drought situation. Alentejo also contains multitudes of different regions; most of them derived from human settlement and land exploration, and some areas that are still semi-natural. Research published in 2021 showed a correlation between the landscape surrounding vineyards and pest outbreaks, in across 400 Spanish vineyards, suggesting that a landscape of semi-natural habitats may lower pests, decrease insecticide use and improve biodiversity. This thesis gathered 19 bioclimatic variables from the CHELSA dataset, as input for the ISO Cluster unsupervised algorithm, a variation of k-means, creating 10 clusters with specific climate characteristics. Those clusters were then integrated with a statistical description of semi-natural areas across the region, through land cover data from DGT — COS 2018. This project aims to provide an initial, descriptive statistical study of Alentejo, aiming, in future work, to analyse semi-natural habitats and vineyard productivity.
Keywords: Portugal; climate description; land cover; vine; GIS.

1. Introduction

Viticulture, the cultivation of grapes for the main purpose of producing wine, is a practice with a long history and still substantial throughout Europe. Particularly in the Mediterranean, wine is embedded in several countries' cultural identity, and grape cultivation is very significant. Two Portuguese regions, according to NUTS2 from the European Union, dedicate more than 8% of their agricultural area to vineyards, from cooperatives and big companies to small wineries.

This article is a conjoined effort made by members of Instituto Superior Técnico and the company NBI, regarding wine growing and its climatic and geographical context. NBI is a Portuguese environmental consultancy group, founded in the beginning of 2020. The group focuses on aiding companies improve their ecological impact, implementing different farming strategies to improve the surrounding environment, or to minimize prejudicial impact. NBI is currently partnered with CVRA, an organisation dedicated to certifying and protecting wines from a specific region in Portugal, Alentejo.

Alentejo, the blue zone in Figure 1, is the region on which this project will focus. Making up to one third of mainland Portugal's area, the climate is temperate, with dry, warm Summers and mild Winters. It is currently posing some challenges for agriculture, as it is the driest region in Portugal, and has some of the highest Summer temperatures. As the climate changes, the rising temperatures and less water availability may turn some areas inhospitable [1]. It becomes relevant to study new areas on which agricultural areas may be installed according to their ecological needs, for climate change adaptation [1, 2].

It is one of many wine-producing regions in Portugal, with its own wine classification label, *"DOC – Alentejo"*. DOC, or Label of Controlled Origin, refers to wines that grow on a specific geographical location, follow a rigorous production method and have a higher standard of quality.

In last years, an organic agricultural practice has been demonstrated to have positive effects, regarding changes in soil health, diversified biodiversity and higher product quality. In 2021, a study presented that the landscape surrounding vineyards may be also relevant: habitats classified as semi-natural were proven to decrease pests and improve biodiversity [3]. Therefore, it may be relevant to study surrounding land usage when managing a vineyard, which is the purpose of this project. This

Figure 1: .
Portuguese NUTS2 region classification in 2013, with subregions.

article will be the first step of a larger study, that intents to correlate these semi-natural areas with vine productivity, suggest a reclassification of the Alentejo region, and lead to a more resilient use of the land.

1.1. Objectives

This article's main objective is to propose new regions in the Alentejo area to assist decisions by future wine producers, such as the location of new vineyards or which grape varieties to plant. Through the ArcGIS software, those regions were defined according to climatic data, and by the presence of semi-natural habitats, such as nearby bodies of water or nearby native forests.

In future work, this project will integrate a larger research produced by NBI, that aims to associate climate and land cover data to vineyard productivity. By studying the presence of specific habitats and its impact on grape yield, wine producers may be advised to install vineyards near some habitats, and steer away from others.

2. Background

The ecological context of a plant rules its development. Variables related to the climate may affect the plant's health [4], therefore making their analysis of vital importance. The highest and lowest temperatures the vine experiences, as well as occurrence of frost or lack of precipitation, may prejudice its development [5, 6], which will in turn affect wine quality [7] and yield [8]. In a Mediterranean climate like Portugal, it has not been common to have extreme weather events like prolonged frost, snow or consistently very high temperatures; so far, Portugal has been recognized as one of the best countries for wine growing, along with others in the Mediterranean area, but climate change should be accounted for in future wine growing implementations.

In Alentejo, a region known for its high Summer temperatures, low precipitation and low water availability, wine producers have already established the best cultivars for specific zones, regarding their preferred temperature and precipitation values, altitude and soil; however the surroundings of vineyards are often overlooked.

In order to maximize yield, an absence of pests and diseases should also be sought for. One of the first studies of its kind, an article by Paredes et al [3] correlated habitats surrounding Spanish vineyards with pest outbreaks. Organic farming is an approach proved to be effective as a more natural pest control, by introducing animals that are natural enemies of common pests, or by planting specific plants that repels nefarious microbes and insects [9]. With the changing climate, pest outbreaks and fungal diseases may increase their impact on viticulture, and pesticides start to become less effective as pests grow more resistant [10].

The aforementioned article modelled landscape effects on pest infestations and insecticide applications, and concluded that vineyards that were surrounded with simplified landscapes, such as more vineyard area, had a significantly higher pest outbreak and need of insecticide application. On the other hand, vineyards surrounded by semi-natural habitats would have four times less outbreaks of the studied species, the European grapevine moth.

We proposed to study Alentejo's semi-natural habitats. Since there is not a developed research in what constitutes a semi-natural area, we had to analyse, one by one, land cover classes of Alentejo. The COS dataset provided us with 11 classifications, which the NBI team divided to 13 to clarify and distinguish some habitats. Variables that might prove beneficial to the environment of the vineyard were chosen to be Water, Riparian Area, Native Forest and Shrubland. Agricultural habitats, due to some practices by producers, negatively impact the soil with chemicals and contaminated animal waste. Urban areas and low vegetation areas do not provide enhanced biodiversity or any ecological benefit, and pastures, also due to wrongful management, strip the soil of vegetation and any other resource. Pine forests and eucalyptus, for their high water needs and for being an invasive species in Alentejo, were also considered non beneficial for vineyards. Native forests are mindful and appropriate of the Portuguese region, and water bodies provide a vital component for vine growth.

3. Datasets

The land cover dataset COS used was provided by DGT (*Directorate-General for the Territory*) [11] and collected by the NBI team. For the months of June, July, August, September and October 2018 the data was produced by DGT through orthophotography. For the remaining months of the year, the dataset collected information from other data sources available, like LUCAS (land cover and Coverage Area frame Survey) [12], CORINE Land Cover [13], and others. More information about data collection methods and specifications may be found on DGT's technical document for the dataset [14].

A supplemental dataset was utilized for water bodies, obtained from the European Commission's Joint Research Centre within the Copernicus Programme [15]. This data was also collected through satellite imagery, derived from the Landsat dataset [16]. The original dataset by DGT provided all bodies of water of Alentejo, despite some of them only being active on some months of the year, possibly due to lack of precipitation and high temperatures. This additional dataset provides an attribute to their data, "Occurrence", that indicates, in percentage, how much time of the year a specific water body exists. The Guadiana river, and the water harnessed by the Alqueva dam have occurrence values superior to 95, while small ponds that disappear during the warmest months have occurrences of 20 or 25%. This dataset allowed for a much accurate assessment of water impact in Alentejo.

The COS classified the area in Alentejo into 9 major classes. This classification is then divided into multiple subclassifications. The NBI team grouped many subclassifications and divided some classes, creating their own classification system with 13 categories, which is exhibited in Figure 2.

The climate set of data was collected from CHELSA (*Climatologies at high resolution for the earth's land surface areas*) [17], by the NBI team. CHELSA is a dataset that consists of "downscaled model output temperature and precipitation estimates" for a climatological period from 1979-2013 [18], using data collected with a 6-hour period from the ERA-Interim dataset.

This dataset has a 30 arcsec resolution. Arcsecond is an angular measurement, corresponding to $\frac{1}{3600}$ of a degree. Because of the spherical shape of the planet, arcsecond is a more precise way of describing the raster resolution. When "flattened" and converted to meters, 30 arcsec are approximately 1000 meters.

DGT Original Classification	Reclassification
Artificial territory	Urban area
Agriculture	Agriculture Vineyards and orchards Olive groves
Superficial water bodies	Water
Humid areas	Water
Forests	Forests Pine forest Eucalyptus
Pastures	Pastures
Agroforestry surfaces	Forests
Shrubland	Shrubland
Low Vegetation areas	Low vegetation areas

Figure 2: Land cover original and reclassified classifications.

The baseline for climate data are individual registries from local weather stations and observatories, that create a local report and analysis of weather. These, stations, however, have a limited geographical scope. Downscaled models of temperature or precipitation utilize climate records, which provides accurate readings for a finite number of locations, and interpolate it for a larger area using mathematical models.

ERA-Interim calculated that interpolation using specific models, which CHELSA reformulated using their own formulas, correcting possible biases.

This was used to calculate 19 bioclimatic variables, known as being biologically meaningful and used for ecological and climatic models [19]. A compiled table of these variables, with respective titles and units, was adapted from the official dataset source [18] and is presented in Figure 3.

Variables	Meaning	Unit
Bio_01	Mean annual air temperature	°C/10
Bio_02	Mean diurnal air temperature range	°C/10
Bio_03	Isothermality	°C/10
Bio_04	Temperature seasonality	°C/10
Bio_05	Mean daily maximum air temperature of the warmest month	°C/10
Bio_06	Mean daily maximum air temperature of the coldest month	°C/10
Bio_07	Annual range of air temperature	°C/10
Bio_08	Mean daily maximum air temperature of the wettest quarter	°C/10
Bio_09	Mean daily maximum air temperature of the driest quarter	°C/10
Bio_10	Mean daily maximum air temperature of the warmest quarter	°C/10
Bio_11	Mean daily maximum air temperature of the coldest quarter	°C/10
Bio_12	Annual precipitation ammount	kg/m^2
Bio_13	Precipitation amount of the wettest month	kg/m^2
Bio_14	Precipitation amount of the driest month	kg/m^2
Bio_15	Precipitation seasonality	kg/m^2
Bio_16	Mean monthly precipitation amount of the wettest quarter	kg/m^2
Bio_17	Mean monthly precipitation amount of the driest quarter	kg/m^2
Bio_18	Mean monthly precipitation amount of the warmest quarter	kg/m^2
Bio_19	Mean monthly precipitation amount of the coldest quarter	kg/m^2

Figure 3: Climate variables, respective meaning and units.

4. Implementation

To analyse Alentejo regarding its climate, a Principal Component Analysis was applied to 19 bioclimatic variables, gathered from the CHELSA dataset. The output for this PCA produced a text file containing the covariance matrix, correlation matrix, eigenvalues, and eigenvectors of the technique. Given the nature of these variables, and that some are derived from other ones in the dataset (e.g., $Bio07 = Bio05 - Bio06$), it's expected that the whole dataset is very correlated [20].

Two pairs of features were found to be identical ($|\rho| = 1$): the pair Bio09 and Bio10, and the pair Bio17 and Bio18. Considering the meaning of these variables in Fig. 3, this is not unexpected. As Portugal is a Mediterranean country, the driest quarter of the year, the Summer, is typically also the driest [21, 22], similar to other countries in the same Koppen climate classification [23]. We therefore proceeded to eliminate two of these features, Bio10 and Bio18.

Bio03, which refers to isothermality, is a measure of how temperature oscillates on a daily basis and its relation to an annual variation. We found its purpose to not be clear and easily explainable, and with its very low significance in the first three Principal Components, it was also not included in the final variable set.

Upon further examination of the correlation table, variables with correlation $|\rho| > 0.7$ were eliminated, keeping the one with higher eigenvalue, therefore relevance. $|\rho| > 0.7$ was tested as being a rigorous value for biologic variables, while still preserving a sufficiently high amount of information [20]. The remaining features were

- Bio01, average temperature for each month of the 1979-2013 period;
- Bio02, average diurnal temperature range ($T_{max} - T_{min}$) for each month of the year set;
- Bio04, standard deviation of the mean monthly temperatures;
- Bio07, temperature annual range, $Bio05 - Bio06$;
- Bio12, cumulative precipitation values of each year, summed and averaged;
- Bio13, maximum value between precipitation amounts for each year, averaged;
- Bio14, minimum value between precipitation amounts for each year, averaged;
- Bio15, standard deviation of precipitation amounts for each month;
- Bio16, mean monthly precipitation amount of the wettest quarter;
- Bio17, mean monthly precipitation amount of the driest quarter.

This final set of variables, the ones most relevant after cleaning, were used to create the final PCA. This PCA has also created a map output, describing the variables across the surface. The first PC is displayed in red, the second PC in green, and the third PC is displayed in blue, creating all colours in the RGB spectrum.

Figure 4: Output from Principal Components Analysis of the final ten bioclimatic variables, Alentejo Region. The three first components are pictured here in RGB format.

Since the objective is to create regions of this area of Portugal, we have applied a clustering algorithm. Given a set of observations, clustering algorithms intend to assemble such observations into groups, As there are no previously labelled groups a point could belong to, it is appropriate to apply an unsupervised machine learning method. The algorithm applied was ISO Cluster, an algorithm mainly utilized on ArcGIS. The ISO Cluster algorithm is a variation of K-Means, one of the most used techniques for unsupervised clustering. The algorithm that initializes K number of cluster centres (means), and assigns each sample to the nearest cluster. After the assigning step, the centroids are recalculated, samples are reassigned them to the nearest one, and this is done iteratively until convergence.

A key parameter of a clustering algorithm is the number of clusters to create. The optimal number is largely situational, depending on the context and goal of performing a clustering classification. For a general sense of the major zones, it might suffice to categorize into 5 classes; a more specific setting could benefit from 10 classes or more, depending on the importance of changes in temperature or precipitation. One degree Celsius of difference between two clusters, for example, might be too high and need additional classes in some situations.

In a general setting, to create too few clusters would be to misclassify the data, and to not provide as much information as the algorithm could. Too many would fall into a loss of significance over each cluster, as they would represent very small subsets of data and, at a maximum degree, would exist one cluster per observation. After gathering data about the ecological context of Alentejo, we decided to perform the algorithm with 5, 6, 7, 8, 9 and 10 clusters. Less than 5 would provide little information, too broad for the scope of the project, while more than 10 classes was found, by consultation with experts at the NBI company, to be too specific for the area under study.

Through ArcGIS Statistical Analysis tools, we were able to gather data to apply the ANOVA statistical test for cluster significance. Both tests calculate the mean and variance of populations, compare them and test the null hypothesis. In this setting, the null hypothesis states that cluster means are approximately identical, therefore the cluster is not robust enough to be relevant, and may be dissolved. If it's found that this hypothesis may be rejected, then there is statistical significance in the tested clusters. The chosen $\alpha = 0.05$ indicates that, for the null hypothesis to be rejected, a 95% or more Confidence Interval for the difference of means must be reached. The *p-value*, the value outputted by the statistical tests to prove the null hypothesis, must also verify *p-value* < 0.05 for it to be rejected [24, 25].

After gathering the number of observations (pixels), means and standard deviation values of the first PC, the ANOVA was applied. Through these gathered values, it revealed that all 10 clusters are statistically significant, with *p-values* $= 0.000$ for each calculated comparison.

The ANOVA, a widely used method in statistical analysis, is considered a valid assessment of significance [26, 27]. Possibly due to the large number of observations per population, small changes in means or standard deviation values were considered sufficiently relevant to justify the existence of all classes.

To try to capture small, and possibly relevant, microclimates in this area, and not do a too simplistic analysis, the 10 class clustering iteration was the one chosen to proceed with the statistical description, as seen on Figure 5.

Apart from the PCA analysis, this project focuses on intersecting COS and climate data. Having the clustered map, we aim to describe how habitats are

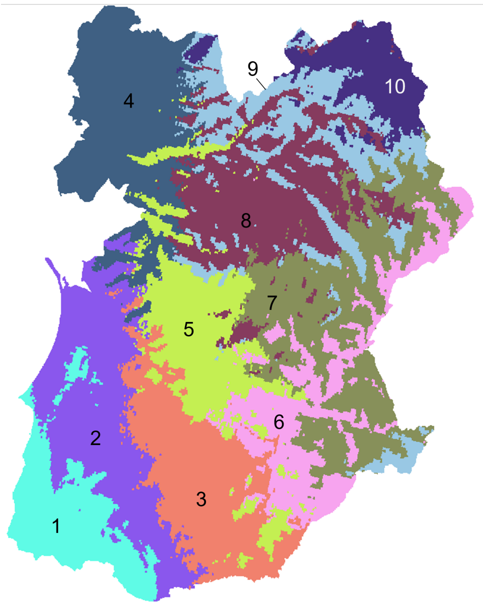

Figure 5: Clustered map of Alentejo, per bioclimatic variables.

scattered throughout Alentejo, therefore we aimed to intersect COS variables with each region. However, some habitats' impact is not limited exclusively to their area, but have a wider impact. The presence of some variables, such as water bodies, might completely alter their location, from the soil composition, plant health, to the habitats of pollinating insects and larger animals.

To cater for that, we created buffers. A buffer is an area created around all sides of a feature, utilized here to portray the reach of a habitat's influence. The decided area of influence was 300 meters, counting from the edges of the feature.

Figure 6: Example of buffer surrounding a water body.

Because cluster areas are important for future statistics, if there are polygons near each other with overlapping buffers, the buffer area in total would be miscounted. Overlapping areas of influence should count as a single one, but their areas would be summed, largely overstating the real value. They were therefore dissolved with the Dissolve Boundaries tool, which allowed for a join of

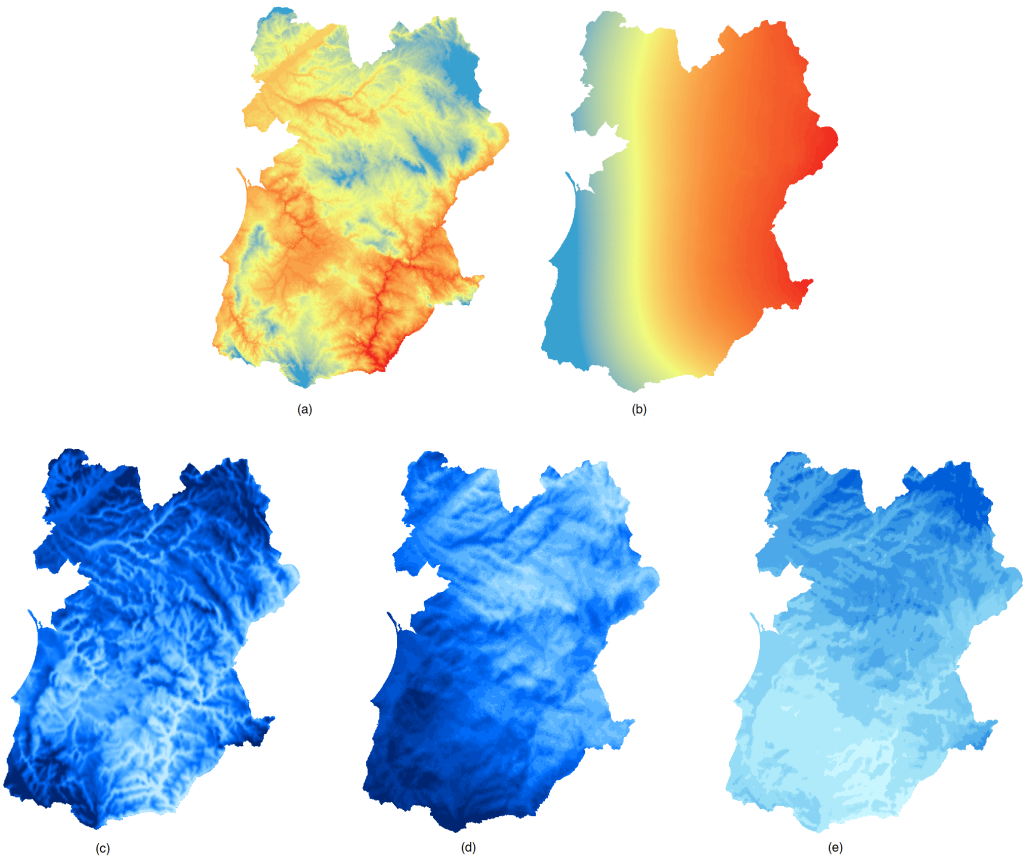

Figure 7: Visual representation of variables Bio01 (a), Bio07 (b), Bio12 (c), Bio15 (d) and Bio17 (e) across the Alentejo map. Darker colours indicate higher values.

overlapping buffers, flattening them and unifying their area.

For reasons stated earlier, these buffers were applied on specific variables: water bodies, native forest, riparian areas and scrubland.

An important detail of the riparian features is their possibly negligible area, because of their nature as usually thin streams of water, being length a more significant measure. On final calculations of area percentage on each cluster, riparian features would have a minimal presence, and their substantial relevance could be undervalued. To consider this issue, ArcGIS provides the Polygon to Centerline tool, particularly designed for hydrographic polygons. This tool creates lines to represent bodies of water, specifically through their centre. Creating centerlines allowed for a study of riparian zones' length, in meters, in final descriptive statistics, besides the area provided by polygons.

5. Results and Discussion

5.1. Climate description

With the PCA we have gained insight on what variables explain the most variability of the Alentejo climate dataset, so it is useful to picture some, as well as others of general interest, for future interpretation. In Figure 7 we depicted Bio01 (a), the annual mean temperature (meanings of bioclimatic variables in Figure 3) and Bio07 (b), deemed the variable with the highest weight on the first PC, which regards to the annual temperature range – the lowest temperature of the year subtracted from the highest temperature, for the 1979-2013 period.

We also present three variables related to precipitation, with special emphasis on Bio17, the variable with the largest weight of the second PC. To recall their meanings, Bio12 refers to Annual Precipitation, the mean amount of precipitation of this set of years; Bio15 is a measure of precipitation seasonality, meaning that higher values imply higher precipitation variability, and Bio17 is the mean precipitation of the driest quarter of the year, averaging the aforementioned years.

In the first image, (a), we see that the central and south areas of Alentejo present a higher annual mean temperature than the northeast. Topographical features are very well detailed here, with the Guadiana river and other small rivers presenting, interestingly, the highest annual mean temperatures. Bio07, (b), is displaying how different the highest temperature in the Summer is related to the lowest temperature in Winter. Near the coast the range is low, and it increases to the east, presenting higher variability. This map of temperature range is exhibiting the region's continentality, an observed climate condition that states that temperatures have smaller variations near a large water body, such as an ocean, and larger variations as one gets farther from the coast [28]. It is a known property and prevalent in countries with a large coastal area, such as Portugal. Near the ocean, temperatures are consistently temperate and with a low range; near the opposite border, shared with Spain, temperatures can vary greatly between the day and the night, and between summer and winter.

From Bio12, (c), we see that North Alentejo, as well as the southernmost area, has the highest precipitation rates, while the interior, as expected, presents less cumulative precipitation. The centre and south areas of the map present higher degrees of variability, meaning that monthly precipitation values can vary greatly, while the north, coincident with Portalegre district, is more stable throughout the year. Bio17 highlights how the south of this region can reach extremely low precipitation levels during the driest quarter, where drought is a common situation. The northernmost region showcased a more stable flux of precipitation throughout the year, one of the highest yearly precipitation levels, and the most humid dry season. We can then hypothesize that this area is more resilient to drought and favourable to species that may need higher water availability.

On Supplementary Materials, provided in the end of this article, we present some statistical analysis of certain bioclimatic variables per cluster. The continentality feature is observed comparing data of Cluster 1 and 10: the annual temperature range in Cluster 1 is 17.21 °C, while in Cluster 10 it is 26.34 °C. In a future implementation of a grape variety that needs constant temperate weather, coastal clusters present themselves as a better choice, temperature-wise.

The temperature values for Clusters 1, 2 and 4 are similar, with Cluster 1 exhibiting the least temperature range. It should be noted that, from the three clusters along the coast, Cluster 1 is the smallest: both Cluster 2 and 4 cover area that might be considered as belonging to the centre of Alentejo, less affected by continentality.

On the eastern border, Cluster 6, 7, 8, 9 and 10 present very similar levels of temperature range, with only one degree of difference between them. That is clearly consistent with their annual minimum and maximum temperature values. Cultivars appropriate for hot climates, which tend to produce sweeter and stronger wine, would be a suitable choice for new wine growing investments; if paired, however, with bodies of water or riparian areas, as rainfall values only seem consistent on Cluster 10.

Cluster 3 is the area with less precipitation amount on the driest quarter, and even in cumulative annual precipitation.

5.2. Habitat description

More than a climate description of Alentejo, in part already addressed in literature with respect to climate change challenges [29, 1], this project's motivation is to encompass that climatic description with land cover data, specifically with variables that are here hypothesized as semi-natural and positive.

This semi-natural variable distribution across the map is summarized in Figure 8. We have calculated the area percentage of each semi-natural habitat, per cluster, plus the riparian length in meters.

Cluster 3, and, to a lesser degree, 2 and 1, have considerably low precipitation values during the driest quarter, and precipitation amount is very inconsistent throughout the year. The three clusters have very substantial riparian length, but further research is needed to test if riparian areas might suffice for an extensive vineyard's water needs.

Cluster 1, which encompasses most of the Odemira municipality of Beja, is the second highest valued cluster in shrubland area, with 47.5%. From the article by Paredes et all [3], it might be beneficial to employ that shrubland area to provide more complex and diverse biodiversity for vineyards; however, it is the second less fertile in water availability, and, as seen earlier, also low on precipitation values. Therefore we may say that this specific area is not for cultivars with high water needs, which is one of the most important factors for vine development [30].

Cluster 10, which represents the Serra de S. Mamede Natural Park, is the cluster highest in shrubland area. Its water body percentage is low, and is the cluster with the second lowest value of riparian length. Despite these values, it is one of the clusters with higher precipitation values throughout the year, and its annual temperature range provides a comfortable interval for vineyards to have a proper dormancy period, where cold is a relevant factor. We may suggest this cluster, for all reasons described above, a potentially good candidate for new vineyards, proving the cultivars chosen are adequate for the high Summer temperatures.

Cluster 4 stands out as a seemingly good candidate for wine growing. It has a considerable riparian length and medium water availability; considerable precipitation amounts, which don't reach

Clusters	Riparian length (meter)	Native forest percentage	Shrubland percentage	Water percentage
Cluster 1	70 383	10,20%	47,50%	3,80%
Cluster 2	112 103	6,10%	11,70%	3,90%
Cluster 3	218 755	6,00%	21,60%	5,70%
Cluster 4	135 739	22,00%	22,00%	4,90%
Cluster 5	70 886	10,90%	6,60%	6,60%
Cluster 6	109 646	7,90%	11,60%	14,60%
Cluster 7	81 415	4,90%	9,00%	6,50%
Cluster 8	257 966	14,40%	7,50%	8,10%
Cluster 9	23 539	8,20%	17,10%	4,10%
Cluster 10	29 210	10,80%	52,60%	2,60%

Figure 8: Statistical description of each semi-natural habitat, per cluster.

critical levels during the warmest quarter; and has a medium precipitation variability. It is also fertile in native forests and shrubland, which are being hypothesized as being beneficial for vineyards. By land cover data, this area is highly urban. Further analysis of this area's land planning might prove beneficial, to utilize this area to its full potential.

6. Limitations

There is some lacking data that would have been beneficial to analyse, such as aquifers. When analysing merely water bodies at the surface, we may be wrongly estimating water availability, therefore it is suggested an aquifer analysis to complement this study.

In the COS dataset, the classification "Olive grove" was assumed to be referring to intensive olive groves, the most common in Alentejo. These are prejudicial for vineyards, due to specific management practices that worsen soil health and surrounding biodiversity. That classification was not explicit regarding that management, and some traditional olive groves might have been misclassified, and could be a positive presence on Alentejo's vineyards. A clarification regarding that specific class should be sought with DGT.

7. Conclusions

Climate suitability for wine growing is highly complex, and dependant on multiple factors. Per Gregory Jones, renowned viticulture expert and climatologist, the minimum temperature interval for proper wine growing is [-1, 18.9] in the growing period [31], which is, in the Northern Hemisphere, the Spring and Summer seasons. From temperature values alone, a producer would, in theory, be able to grow vines on all clusters defined in this project. Water availability is, however, a very important factor for vines, and also, proven by Paredes et al, their surroundings.

Cluster 4 and Cluster 10 are, from the data collected and analysed, good candidates for future wine growing, due to their medium to high rainfall values, appropriate temperatures (although for different cultivars), and very high percentage of semi-natural variables.

We have described the presence of four semi-natural variables in Alentejo, and, within that description, elaborate on the area's climate. NBI, partnering with CVRA, may deepen this work by obtaining productivity data from wine producers. Having a stable climate, productivity and semi-natural habitat presence may be correlated, to further discover relationships between them. A positive correlation would not only aid in controlling pests and lower insecticide applications, but it might incentivize general habitat conservation, which might bring benefits to Alentejo's climate, but also flora, fauna and population.

Acknowledgements

The author would like to thank her two professors, Prof. Ana Matos and Prof. Ana Subtil of Instituto Superior Técnico, for their guidance, and Prof. Hugo Rebelo and Francisco Marques of the NBI company, for their ecological and technical insights.

References

[1] Helder Fraga and Joao Santos. Climate change projections for the Portuguese viticulture using a multi-model ensemble. *Ciência Téc. Vitiv*, 27(1):39–48, 2012.

[2] D. Santillán, V. Sotés, A. Iglesias, and L. Garrote. Adapting viticulture to climate change in the mediterranean region: Evaluations accounting for spatial differences in the producers-climate interactions. 12.

[3] Daniel Paredes, Jay A. Rosenheim, Rebecca Chaplin-Kramer, Silvia Winter, and Daniel S. Karp. Landscape simplification increases vineyard pest outbreaks and insecticide use.

[4] Anna Dalla Marta, Daniele Grifoni, Marco Mancini, Paolo Storchi, Gaetano Zipoli, and Simone Orlandini. Analysis of the relationships between climate variability and grapevine phenology in the nobile di montepulciano wine production area. 148:657–666.

[5] Helder Fraga, Aureliano C. Malheiro, José Moutinho-Pereira, Rita M. Cardoso, Pedro M. M. Soares, Javier J. Cancela, Joaquim G. Pinto, and João A. Santos. Integrated analysis of climate, soil, topography and vegetative growth in iberian viticultural regions. 9(9):e108078.

[6] Gregory Jones and Gregory Goodrich. Influence of climate variability on wine regions in the western USA and on wine quality in the napa valley. 35:241–254.

[7] Gregory Jones, MA White, and O Cooper. Climate change and global wine quality. 85:504–+.

[8] Josep Camps and María Ramos. Grape harvest and yield responses to inter-annual changes in temperature and precipitation in an area of north-east spain with a mediterranean climate. 56:853–64.

[9] Daniel Paredes, Daniel S. Karp, Rebecca Chaplin-Kramer, Emilio Benítez, and Mercedes Campos. Natural habitat increases natural pest control in olive groves: economic implications. Publisher: Springer Verlag.

[10] Fred Gould, Zachary S. Brown, and Jennifer Kuzma. Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? 360(6390):728–732.

[11] Registo nacional de dados geográficos - Direção-Geral do Território.

[12] Land use and coverage area frame survey (LUCAS) —| land & water | food and agriculture organization of the united nations | land & water | food and agriculture organization of the united nations.

[13] EEA), European Environment Agency. CLC 2018 — copernicus land monitoring service.

[14] DGT. Especificações técnicas da carta de uso e ocupação do solo (COS) de portugal continental para 2018 - DGT.

[15] Jean-François Pekel, Andrew Cottam, Noel Gorelick, and Alan S. Belward. High-resolution mapping of global surface water and its long-term changes. 540(7633):418–422.

[16] European Commission's Joint Research Centre. Global surface water – data access.

[17] Dirk Nikolaus Karger, Olaf Conrad, Jürgen Böhner, Tobias Kawohl, Holger Kreft, Rodrigo Wilber Soria-Auza, Niklaus E. Zimmermann, H. Peter Linder, and Michael Kessler. Data from: Climatologies at high resolution for the earth's land surface areas, 2018.

[18] Dirk Nikolaus Karger and Niklaus E. Zimmermann. CHELSA climate downloads—technical description.

[19] Bioclimatic variables — WorldClim 1 documentation.

[20] Carsten F. Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell, and Sven Lautenbach. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. 36(1):27–46.

[21] Carla Mora and Gonçalo Vieira. The climate of portugal. In Gonçalo Vieira, José Luís Zêzere, and Carla Mora, editors, *Landscapes and Landforms of Portugal*, pages 33–46. Springer International Publishing. Series Title: World Geomorphological Landscapes.

[22] Pedro Miranda, M. Fátima Coelho, António Tomé, Maria Valente, Anabela Carvalho, Carlos Pires, H.O. Pires, and Vanda Pires. 20th century portuguese climate and climate scenarios in climate change in portugal: Scenarios, impacts and adaptation. pages 27–83.

[23] IPMA. Clima Normais em Portugal.

[24] Andrew F. Siegel and Michael R. Wagner. Chapter 15 - ANOVA: Testing for differences among many samples and much more. In Andrew F. Siegel and Michael R. Wagner, editors, *Practical Business Statistics (Eighth Edition)*, pages 485–510. Academic Press.

[25] A.K. Md. Ehsanes Saleh, Mohammad Arashi, R. A. Saleh, and Mina Norouzirad. Analysis of variance (ANOVA). In *Rank-Based Methods for Shrinkage and Selection*, pages 149–190. John Wiley & Sons, Ltd.

[26] Mario Peña, Mariela Cerrada, Ximena Alvarez, Diana Jadán, Pablo Lucero, Barragán Milton, Rodrigo Guamán, and René-Vinicio Sánchez.

Feature engineering based on ANOVA, cluster validity assessment and KNN for fault diagnosis in bearings. 34(6):3451–3462.

[27] T. Calinski and L. C. A. Corsten. Clustering means in ANOVA by simultaneous testing. 41(1):39.

[28] N. P. Molchanova, A. V. Letuchy, S. V. Morozova, K. S. Kondakov, and N. A. Shcherbakova. The influence of the degree of climate continentality on the productivity of agricultural production. 1010(1):012156. Publisher: IOP Publishing.

[29] H. Fraga, J. A. Santos, A. C. Malheiro, A. A. Oliveira, J. Moutinho-Pereira, and G. V. Jones. Climatic suitability of Portuguese grapevine varieties and climate change adaptation. *International Journal of Climatology*, 36(1):1–12, January 2016.

[30] Henry Ojeda, Alain Deloire, and Alain CARBONNEAU. Influence of water deficits on grape berry growth. 40:141–145.

[31] Gregory Jones. Climate, grapes, and wine.

Supplementary Materials