

Grapevine Yield Big-data for Soil and Climate Zoning. A case study in Languedoc-Roussillon, France

Hugo Fernandez-Mena^{1,2,*}, Nicolas Guilpart³, Philippe Lagacherie⁴, Renan Le Roux⁵, Nina Graveline⁶, Jean-Marc Touzard⁶, Hervé Hannin⁷, and Christian Gary¹

¹UMR ABSys – Montpellier SupAgro; INRAE; CIRAD; U. Montpellier – Montpellier, France

²UMR EMMAH – INRAE; U. Avignon – Avignon, France

³UMR Agronomie – AgroParisTech ; INRAE; Université Paris-Saclay – Thiverval-Grignon, France

⁴UMR LISAH – INRAE; CIRAD; IRD – Montpellier, France

⁵U. AgroClim – INRAE – Avignon, France

⁶UMR Innovation – INRAE; CIRAD; Montpellier SupAgro; U. Montpellier – Montpellier, France

⁷UMR MOISA – INRAE; CIRAD; Montpellier SupAgro; U. Montpellier – Montpellier, France

Abstract. New winegrower and resource datasets appear to be a great opportunity to understand which are the environmental factors involved in grapevine yield spatially. Such analysis can help regional label managers and winegrowers for the conception of local adaptation strategies to climate change, reducing yield gaps. In the present study, we aggregated yield a big dataset obtained from *Pays d'Oc* winegrowers ($n = 96677$) between 2010 and 2018 at the municipality level ($n = 606$), located in the Languedoc-Roussillon region, in the South of France. We used a backward stepwise model selection process using linear mixed-effect models to discriminate and select significant indicators capable of estimating grapevine yield at the municipality level, these include: Soil Available Water Capacity (SAWC), soil pH, Huglin Index, the Climate Dryness Index, the number of Very Hot Days and Days of Frost. We then determined spatial zones by creating clusters of municipalities with similar soil and climate characteristics. The seven zones presented two marked yield levels. Yet, all zones had municipalities with both high yield and high yield gaps. On each zone, grapevine yield was found to be driven by a combination of climate and soil factors, rather than just by a single environmental factor. Environmental factors at this scale largely explained yield variability across the municipalities, but they were not performant in terms of annual yield prediction. Further research is required on the interactions between environmental factors, plant material and farming practices.

1 Introduction

In viticulture, zoning is based on the idea of ‘terroir’, which is the relationship between climate, soil, vine compartments, and farming practices to ensure wine quality and typicity [1]. Terroir was often based on traditional knowledge to delimitate zones, however, as some authors suggest, terroir studies should be more unbiased by using precision agriculture techniques to measure environmental data [2,3]. For so, authors started using climate indicators for zoning, like in the MCC system (Multicriteria Climatic Classification) [4]. Recent studies combined both climate and soil indicators for

viticultural zoning [5-7]. Although some studies have integrating grapevine yield data [8], no methods exist for directly classifying soil and climate indicators related to grapevine yield.

Compared to other crops, grapevine yield has been historically overlooked, assuming a trade-off between yield and wine quality [9]. Most European labels set limits on grapevine yields within the framework of geographical indications, but in many vineyards, producers cannot reach the maximum authorized yield. Historical grapevine yields in France have stagnated since the 1980s [10], and

* Corresponding Author: hugo.fernandez-mena@supagro.fr

environmental and management drivers may be causing a “vineyard decline” [11]. Climate change is also expected to have a particularly negative impact on grape yield in warm, dry winegrowing regions such as the Mediterranean [12]. Little is known about grapevine yields at large scales, with only recent studies analyzing water-limited grapevine yield gaps in the Barossa and Eden valleys [13].

In Languedoc-Roussillon, the *Pays d’Oc* Protected Geographical Indication (PGI) is particularly concerned about the long-term stability of its production levels, as yield is a main driver of individual wine estate performance and its long-term sustainability. All grapevine producers under the *Pays d’Oc* PGI quality label are subject to the same maximum yield requirements established by the label and most years, numerous producers are far from the yield limit. The PGI label supports exploring avenues for stabilizing yields at satisfactory individual and collective levels. Knowledge about environmental factors involved in grapevine yield will help to create potential adaptation measures to reduce yield gaps.

The objective of this study was to identify spatial zones within the wine-growing region of ‘Languedoc-Roussillon’ (South of France, south-eastern part of the French Occitanie region) based on climate and soil indicators linked to grapevine yield. To achieve this, we collected data from grapevine producers under the *Pays d’Oc* PGI quality label for a period of 9 years. The data was aggregated at the municipality level and we calculated the soil and climate indicators that influence grapevine yield at this level. We used scientific literature to determine the indicators that had a significant impact on grapevine yield at the municipality level and clustered the zones presenting similar indicators. This facilitated the identification of environmental resources. Our hypothesis was: (i) different combinations of climate and soil can result in different yield levels, (ii) the same yield level can be achieved with different combinations of climate and soils, and (iii) some grape varieties are preferentially cultivated in specific combinations of climate and soils with higher associated yields.

2 Selection of indicators by data mining

Grapevine yield data were obtained from harvest customs declaration data provided by producers under the *Pays d’Oc* PGI in the former Languedoc-Roussillon region. The dataset contains a total of 96,667 yield data for a period of 9 years (from 2010 to 2018), 58 grapevine varieties and 606 municipalities. Yield data was aggregated on a yearly basis at municipality level for all grapevine varieties and vineyards, resulting in 4455 annual municipality yield data. The PGI label sets a maximum red and white wine production limit at 90 hl·ha⁻¹·year⁻¹.

In our database, the vineyard cultivated area and wine volume as declared by the *Pays d’Oc* label were aggregated on the municipalities presenting a mean of 65.3 hl·ha⁻¹·year⁻¹, a median of 67.03 hl·ha⁻¹·year⁻¹. To fill the yield gap to reach label’s maximum yield requirements for red and white wines (i.e. 90 hl·ha⁻¹·

year⁻¹) in all municipalities, a total volume of 684 318 hl·year⁻¹ could be expected.

The estimation of average grapevine yield at the municipality level between 2010 and 2018 revealed localised yield gaps in numerous municipalities (Fig. 1). Temporally, no declining trend was observed within this time frame, although some years had lower yields, in particular 2010 and 2017, which were linked to severe drought conditions.

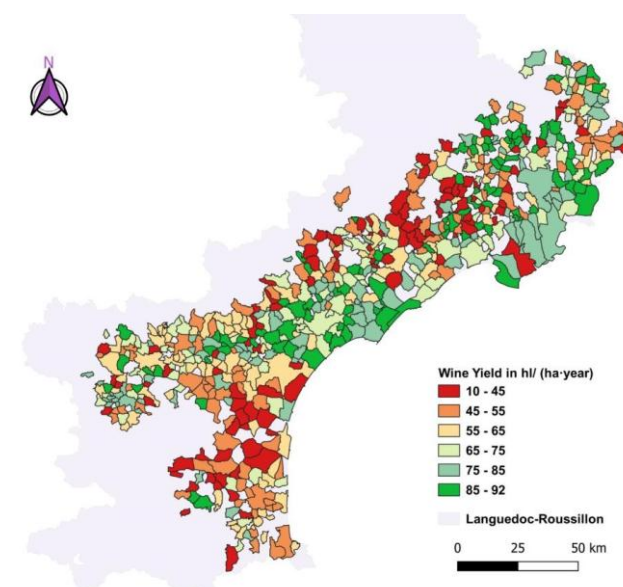


Figure 1. Municipality average grapevine yield (weighted by their area for all grapevine varieties) in Languedoc-Roussillon of *Pays d’Oc* PGI labelled wines between 2010 and 2018 ($n = 606$ municipalities). Top left: histogram of the displayed data for all municipalities and 9 years between 2010 and 2018 ($n = 4455$) averaged from individual yield declarations ($n = 96677$).

Climate and soil data were aggregated at the municipality level. Average grapevine yield (hl·ha⁻¹·year⁻¹) was calculated as the area-weighted average of yields over all grapevine varieties in each municipality. SAWC map raster was intersected with the municipalities of the region studied and an area weighted average of SAWC was calculated for each municipality [14]. We calculated the municipality soil pH as the average pH of several soil layers at depths of 0-5, 5-15, 15-30 and 30-60 cm [15]. The SAFRAN-Météo France weather data in 8x8km grid were aggregated at municipality level using the nearest neighbor method [16].

After identifying and calculating an initial set of climate and soil indicators relevant to grapevine yield, we used a backward stepwise model selection process using linear mixed-effect models to discriminate and select the statistically significant indicators [17,18] capable of estimating grapevine yield at the municipality level. The indicators tested and selected are presented in Table 1.

We selected a mixed model that maximised AIC and BIC performance, for which 6 of the 10 calculated indicators proved to have a significant effect on the annual grapevine yield of the municipalities ($n = 4455$). This method obtained a low marginal R^2 (0.112), thus showing low potential for annual yield prediction. Yet, the same predictors proved to be more relevant for the prediction of average grapevine yield for the whole period ($n = 606$), with a marginal R^2 of 0.546 and a conditional R^2 of 0.627. The variables that were found to have a significant effect on grapevine yield at the municipality level were, in order of increasing significance: soil available water capacity, climatic dryness index, Huglin Index, days of frost, soil pH and the very hot days. Despite their theoretical impact on grape yield, four indicators were excluded from our model, presented in Table 1.

Table 1. Soil and Climate indicators as Yield predictors.

Variable name	Units	Average	p-value selection or exclusion
Grapevine yield			
Municipality average grapevine yield	hl·ha ⁻¹ ·year ⁻¹	67.03	
Soil			
Soil Available Water Capacity	mm	88.79	***
Soil pH	-	7.91	**
Climate			
Climatic Dryness Index	mm	-221.6	***
Huglin Index	°C	2270	***
Days of Frost	days	16.76	***
Frequency of Late Frost	days	0.01	excluded
Severity of Late Frost	°C	0.0002	excluded
Hot Days	days	31.83	excluded
Very Hot Days	days	1.8	**
Severity of Heat Stress	°C	2.16	excluded

Based on these indicators, we clustered the municipalities with similar soil and climate, helping us to create groups of municipalities, hereon, referred as zones. For the clustering, we used a combination of principal components analysis and ascendant hierarchical classification [19].

3 Zoning results

The principal components analysis and ascendant hierarchical classification helped to define seven clusters of municipalities using the selected soil and climate indicators. Each of those clusters represents an agroecological zone with similar soil and climate characteristics that favours or constrains grapevine yield. The zones are spatially displayed in Fig. 2 and their soil and climate characteristics are explained as follows.

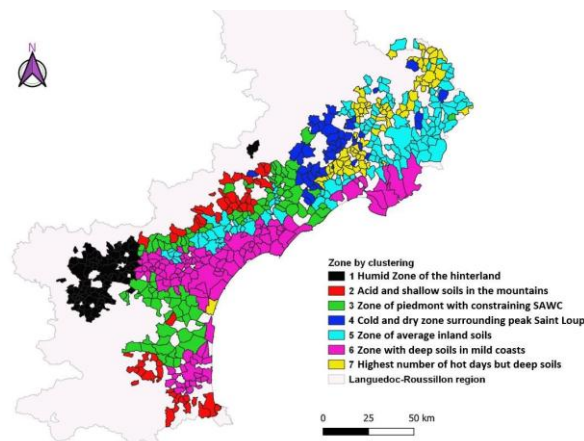


Figure 2. Soil and climate zones related to grapevine yield at the municipality level in Languedoc-Roussillon.

Zone 1 is the ‘Humid zone of the hinterland’ and has the coolest temperatures due to its distance from the Mediterranean coast. As a consequence of having the lowest Climate Dryness Index (around -150 mm) and number of Very Hot Days (from 0 to 2), this region benefits from high grapevine yield. Its main constraint is its Huglin Index, which is the lowest, with 300-400 degree-days less than other zones. The Soil Available Water Capacity is relatively high (from 70 to 100 mm) and Days of Frost are average (from 10 to 20).

Zone 2 is the ‘Zone with acid and shallow soils in the mountains’, which is the only one with an acid soil pH (ranging from 5 to 7.5), which constrains grapevine yield. This zone also has the lowest soil available water capacity (from 50 to 80 mm) and a low Huglin Index. The rest of the variables are average. The municipalities of this zone are located at the highest elevations with municipalities in the southern (Pyrenees mountains) and northern (Caroux Mountains) areas of the region.

Zone 3 is the ‘Zone of piedmont with constraining SAWC’. It has low temperature-related variables, similar to those in Zone 2, but municipalities in this zone have alkaline soil pH (from 7 to 8.3). Water-related indicators are also not very favourable, although is significantly higher than in Zone 2. The municipalities of this zone are located at mid-elevation and in the piedmont areas of the region.

Zone 4 is the ‘Cold and dry zone surrounding Pic St Loup’. This zone is constrained by numerous Days of Frost, but high temperatures in summer. This zone is also constrained by low water availability from rainfall and low Soil Available Water Capacity. The municipalities of this zone are located in high areas surrounding the peak Saint-Loup (north of Montpellier).

Zone 5 is the ‘Zone of average inland soils’. It comprises relatively average soils, although Soil Available Water Capacity is very variable. The region is

constrained by a high Climatic Dryness Index and the highest Huglin Index. The municipalities of this zone are mainly located on the inland plains in the central and eastern parts of the region.

Zone 6 is the ‘Zone with deep soils in mild coasts’. It comprises the best soils (highest Soil Available Water Capacity), compensating for having the highest water deficit (highest Climatic Dryness Index) in the region. Extreme temperatures are rare in this zone due to the proximity of the sea.

Zone 7 is the ‘Highest number of very hot days but deep soils’. It is subject to the most extreme temperatures with the highest level of Very Hot Days and many Days of Frost. In contrast, water availability is favourable due to deep soils (high Soil Available Water Capacity) and lower Climatic Dryness Index. The municipalities of this zone are located on several inland plains in the eastern part of the region.

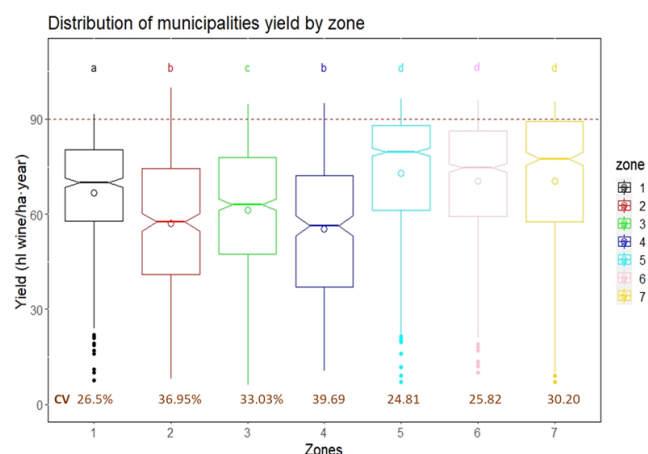


Figure 3. Distributions of average municipality grapevine yields ($n=606$) in $\text{hl}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$ in each of the seven clustered zones in Languedoc-Roussillon between 2010 and 2018. The boxplots represent the distribution in quartiles with median lines. Circles represent the mean and filled dots are outliers. Letters correspond to Tukey’s range test for comparison of means. The dashed red line corresponds to the $90 \text{ hl}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$ maximum label yield used for yield gap calculation. The percentage in brown corresponds to the coefficient of variation over time for each zone.

Depending on their average yield distribution, the clustered zones can be divided into two main groups (Fig. 3):

1. The group with the lowest yields, ranging from 50 to $60 \text{ hl}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$. This group corresponds to municipalities in Zones 2, 3 and 4. Within this group, Zone 3 has a significantly higher yield.
2. The group with the highest yields, ranging from 65 to $80 \text{ hl}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$, which corresponds to the municipalities of Zones 5, 6 and 7.

Municipalities in Zone 1 show an intermediate yield gap between the two above-described groups. We observed a high variation in yield levels depending on the zones as shown by the coefficient of variation in Fig. 3. Although over time there was not a significant tendency towards lower yields in the zones, the zones with low yields (i.e., Zones 2, 3 and 4) drastically reduced their yields in occasional years. Zones 5 and 6 account for the highest cultivated areas (i.e., 20000 to 25000 ha) and Zones 4 and 2 for the lowest (i.e., 1000 to 2500 ha).

4 Conclusion and perspectives

The current study proposes a methodology for selecting theoretical climate and soil factors that could significantly impact grapevine yield at the municipality level. By conducting research at this scale, it is possible to acquire further insights into winegrowing landscape characteristics that could facilitate future studies on vineyard management practices. Our analysis revealed six pertinent factors that accounted for grapevine yield at $R^2 = 0.546$, which only partially explained grapevine yield. Future research should consider a longer yield and the time-span of the climate database to enhance the precision of indicator selection.

We opted to employ clustering to aid in the analysis of the types of municipality with comparable soil and climate characteristics. The decision to apply this zoning approach was also driven by the fact that it offers a foundation for creating R&D recommendations. Based on statistical clustering carried out with the guidance of regional wine label experts, we divided the Languedoc-Roussillon region into seven distinctive zones that had two contrasting yield gap levels linked to distinct indicator combinations related to the limitation of grapevine yield. For each zone, we ascertained the extent to which pedoclimatic factors could account for the variability. Understanding the limiting factors associated with each zone could assist local experts in implementing adaptation measures to prevent or limit grapevine yield loss.

In this study, we demonstrated that environmental factors at this scale could account for a small portion of the annual variability of yield but a significant portion (>50%) of the average yield over time. Further research is necessary to examine the interactions between plant material and farming practices within each zone, as they may also play a crucial role in grapevine yield gaps at the regional scale.

References

1. van Leeuwen, C., Bois, B. *E3S Web of Conferences*. EDP Sciences, Zaragoza, Spain, 01044 (2018)
2. Bonfante, A., Brillante, L.. *Oeno One* **56**, 375-388 (2022)
3. Brillante, L., Bonfante, A., Bramley, R.G.V., Tardaguila, J., Priori, S. *Front. Earth Sci.* **8** (2020)

4. Tonietto, J., Carbonneau, A., Agric. For. Meteorol. **124**, 81-97 (2004)
5. Fraga, H., Malheiro, A.C., Moutinho-Pereira, J., Cardoso, R.M., Soares, P.M., Cancela, J.J., Pinto, J.G., Santos, J.A. PLoS One **9** (2014)
6. Cardoso, A.S., Alonso, J., Rodrigues, A.S., Araújo-Paredes, C., Mendes, S., Valín, M.I., Appl. Geogr. **107**, 51-62 (2019)
7. Bramley, R. g. v., Ouzman, J., Aust. J. Grape Wine Res. **28**, 196-207 (2022)
8. Bramley, R.G.V., Ouzman, J., Trought, M.C.T., OENO One **54**, 903-917 (2020)
9. Poni, S., Gatti, M., Palliotti, A., Dai, Z., Duchêne, E., Truong, T.-T., Ferrara, G., Matarrese, A.M.S., Gallotta, A., Bellincontro, A., Sci. Hortic. **234**, 445-462 (2018)
10. Schauburger, B., Ben-Ari, T., Makowski, D., Kato, T., Kato, H., Ciais, P. Scientific Reports **8**, 1-12 (2018)
11. Riou, C., Agostini, D., Aigrain, P., Barthe, M., des Robert, M.-L., Gervais, J.-P., Jobard, E., Lurton, L., Moncomble, D., Prêtet-Lataste, C.. *BIO Web of Conferences*, EDP Sciences, 01040 (2016)
12. Ollat, N., van Leeuwen, C., de Cortazar Atauri, I.G., Touzard, J.-M.. Oeno One **51**, 59-60 (2017)
13. Bonada, M., Phogat, V., Collins, C., Petrie, P.R., Sadras, V.O. bioRxiv (2022)
14. Styc, Q., Lagacherie, P. Soil Syst. **3**, 34 (2019)
15. Vaysse, K., Lagacherie, P., Geoderma Reg. **4**, 20-30 (2015)
16. Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.-M., Int. J. Climatol. **30**, 1627-1644 (2010)
17. Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M. Mixed effects models and extensions in ecology with R, Statistics for Biology and Health, (2009)
18. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., nlme R package version 3.1-152 (2021)
19. Lê, S., Josse, J., Husson, F.J. Stat. Softw. **25**, 1-18 (2008)