A better understanding of the climate effect on anthocyanin accumulation in grapes using a machine learning approach.

Girault Gnanguenon Guesse[1], Patrice Loisel[1], Bénedicte Fontez[1], Nadine Hilgert[1], Thierry Simmonneau[2*]

[1]MISTEA, Université Montpellier, INRAE, Institut Agro, Montpellier, France
[2]LEPSE, Université Montpellier, INRAE, Institut Agro, Montpellier, France

***Corresponding author:** thierry.simonneau@inrae.fr

Keywords: Machine learning, Anthocyanin, Temperature, Irradiance, SpiceFP.

Abstract

Our knowledge of the climate effect on the quality of harvest remains very incomplete. Original experiments were carried out on Syrah to study the combined effects of normal or high air temperature and varying degrees of exposure of the berries to the sun. Berries were sampled and analysed throughout the ripening period. Several quality characteristics were determined, including anthocyanin content. The objective of the experiments was to investigate which climatic determinants were most important for anthocyanin accumulation. We developed the SpiceFP procedure to explore the possible joint influence of two to three temporal predictors (here temperature and irradiance, observed over time with a very thin discretization step) on the variations of a scalar response variable (here a grape berry quality variable as an example). Particular attention was paid to the interpretability of the results. Analysis of the data using SpiceFP identified a negative impact of morning combinations of low irradiance and high temperature. This impact slightly varied depending on night temperature.

Introduction

The current climate changes are directly threatening the delicate balance of grape attributes that is sought at harvest to make good wines. The ripening period is moved forward to midsummer, at a time when radiation and air temperature peak at their maximum. In this context, the implementation of corrective practices becomes critical. Unfortunately, our knowledge of the climate effect on the quality of different grape varieties remains very incomplete to guide these choices.

Numerous experiments combining various climatic scenarios, different grape varieties and different types of management are being carried out to establish benchmarks and highlight generic behaviour laws with respect to climate change or cultural practices. In particular, the European INNOVINE project (2012-2016) brought together several research teams from the Montpellier centre specialised in ecophysiology, oenology and data analysis. During this project, original experiments were carried out on Syrah vines to study the combined effects of normal or high air temperature and varying degrees of exposure of the berries to the sun. Berries subjected to these different conditions were sampled and analysed throughout the ripening period. Several quality characteristics were determined, including the anthocyanin content. The objective of the experiments was to investigate which climatic determinants were most important for the accumulation of anthocyanin in the berries and at what time of the day and in the developmental cycle these climatic determinants had the greatest impact. The primary aims were to better understand and predict the impact of climatic conditions on anthocyanin accumulation. Several practices, known for their effect on exposure to irradiance, such as position of the clusters in the row and leaf removal, were also compared.

This article describes the method developed to identify which combined levels of the two key climatic variables, temperature and irradiance, had the greatest impact on anthocyanin accumulation and when it is most influential during the day.

Materials and methods

Data were collected during an experiment conducted in a vineyard of Institut Agro Montpellier in 2014 (Syrah vines). The aim was to study the influence of the micro-climate (temperature, irradiance) measured at the grape level on the anthocyanin content of the berries (Zhu et al., 2016). Experts in viticulture consider that the accumulation of chemical compounds affecting the quality of the grape berry is jointly influenced by these explanatory variables. This assumption is reinforced by previous results (Tarara et al., 2008) showing that the anthocyanin composition of Merlot grapes is influenced by a complex combined effect of berry temperature and solar irradiation.

The experimental design consisted of: 3 rows × 2 sun exposures (East or West side of the row) × 16 vine stocks per row, resulting in a potential number of 96 'statistical individuals'. The Ferari Index FI_i were measured weekly for each individual *i*. This is a non-destructive proxy of anthocyanin content of the bunches (Agati et al., 2007). Some individual temperature records had missing data, which were imputed via MissForest approach (Stekhoven & Bühlmann, 2011) when it concerned less than 20% of the total time of observation, otherwise the individual was removed. The final sample size was n=79 statistical individuals with complete observation on temperature, irradiation and Ferari index measurements. Ferari index was determined weekly from veraison to maturity on one same bunch for each individual with Multiplex Force A apparatus (Ghozlen, 2010). This study focused on the data observed during the third week where the increase in Ferari index was the largest. Individuals were classified in either delayed or advanced group depending on their Ferari index value (FI lower or higher than 0.4) at the beginning of this 3rd week.

We modelled the increase in Ferari index, for each group, in function of combined intervals of temperature and irradiation values using the SpiceFP approach (Gnanguenon Guesse, 2021), and the associated R-package (Gnanguenon Guesse & al., 2021). SpiceFP is based on a transformation of temporal variables into categorical variables by defining joint modalities, from which a collection of multiple regression models is derived. The regressors are the frequencies associated with the joint class intervals. Selection of class intervals and related regression coefficients are performed through a generalised linear constraint regression (mixed between sparsity and fused values for regression parameter values).

We applied a logarithm transformation for irradiance values. The input parameters for SpiceFP were the numbers of class intervals (for temperature: 9, 11, 13, 15, 17, 19, 21; for irradiance: 9, 12, 15, 18, 21, 24, 27), the power used in the irradiance values transformation (0.0025, 0.0039, 0.0060, 0.0094, 0.0147, 0.0229, 0.0357, 0.0556, 0.0868, 0.1353), and strength of the ratio between sparsity and fused constraints (1/200, 1/100, 1/50, 1/25, 1/12.5, 1/6.25, 1, 6.25, 12.5, 25). SpiceFP allowed us to compare 490 000 models and to select among them the best one according to AIC (Akaike Information Criteria) or BIC (Bayesian Information Criteria).

We deepened our study on the delayed group by looking at the effect of the night temperature from dawn to midnight (Rient & al., 2014) on the Ferari index. Two subsets were created respecting the natural distribution of the individual temperatures into two night temperature conditions: cool night (less than 21,75°C) and warm night (higher than 21,75°C).

Results

Detection using SpiceFP

The application of SpiceFP on the group of advanced individuals did not show significant results, which suggests no influence of the temperature-irradiance couple during the 3rd week in this group.

The main results were observed in the group of delayed individuals. During this phase of strong increase in anthocyanin content (week 3), high temperatures combined with low morning irradiances had a negative impact that delayed anthocyanin accumulation. This observation was reinforced with an analysis on both subsets created with respect to the two night temperature conditions: cool night and warm night, as illustrated in Figure 1 below.

For each line, the graph on the left represents the non-zero coefficients identified by SpiceFP and associated with the selected model (AIC). The frequencies of the pairs (temperature, irradiance) corresponding to the selected model are given in the middle graph, these are the regressors of the model. The graph on the right indicates the goodness of fit obtained with the selected model.

This negative impact is generally observed for temperature levels that increase with the irradiance level, drawing an oblique separation line in the temperature-irradiance interval matrix in Figure 2. This separation varies with

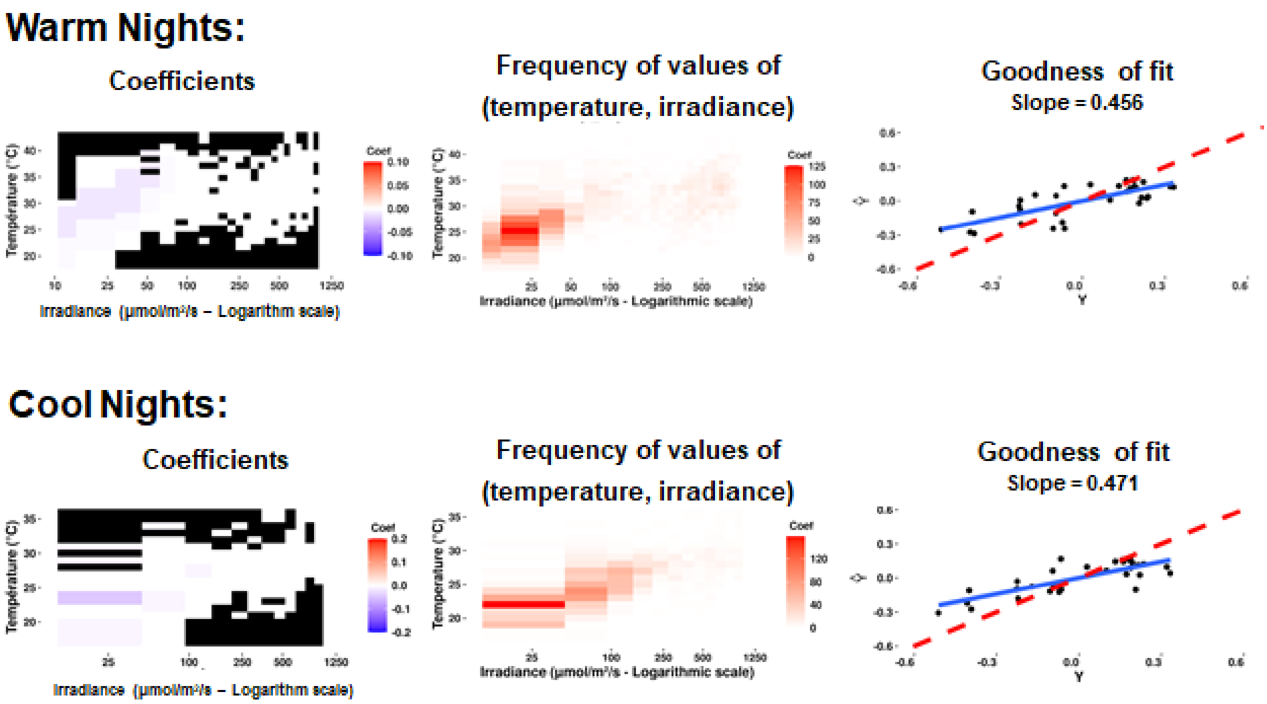

Figure 1. Results of SpiceFP on the group of delayed individuals (FI<0.4) divided into two sub-groups according to either cool or warm night temperature conditions.

the temperature of the night. These results show the importance of combinations of irradiance-temperature values and the need for further studies to analyse the possible decoupling of temperature and irradiance.

Influence model: representation of favourable climatic conditions

From the SpiceFP results (not all are presented here), above a straight line in the (T; log(I)) plane, negative coefficients are found for the best solutions with respect to AIC or BIC criteria. In some very limited cases, positive coefficients are present below this line. Following these two remarks, we proposed the following model Δ Ferari Index = β_- Nb(Temperature>T and I_0 < Irradiance < I_1) + β_+ Nb(Temperature<T or Irradiance>I_1) where Δ Ferari Index is the increase in Ferari index during the selected week and T is the temperature on the straight line : $T = T_0 + \theta \log(I/I_0)$ and $\beta_- < 0 < \beta_+$. The model is illustrated in the following Figure 2:

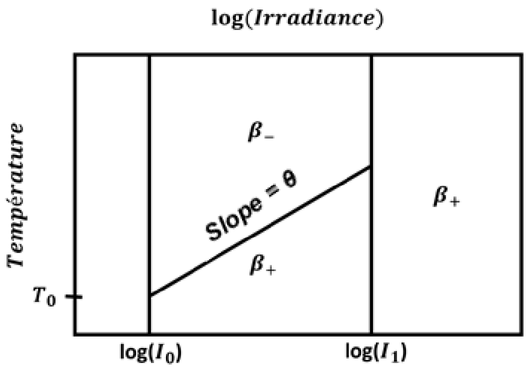

Figure 2. Model based on the oblique separation line.

T stands for temperature, *I* for irradiance and β for the regression coefficient (- and + for respectively a negative and a positive value of β).

The parameters of the model were estimated with an optimization criterion under these constraints $\beta_- < 0 < \beta_+$, $\theta > 0$ and I0 < I1. It led to the following estimation:

Table 1. Legend of table 1 here, above the table (Times New Roman, 11, bold).

	β_-	β_+	T_0 (°C)	I_0 (µmol m^{-2} s^{-1})	θ (°C)	I_1	Crit/N
Delayed (32)	-.00043	.00504	24.42	42.65	7.77	infinity	.02596
Advanced (45)	-.06019	.00375	26.52	139.5	25.68	infinity	.02727

In each case, I_1 is rejected at infinity, so the associated condition is not operational. The negative value of β above the separation line is confirmed. The novelty is the detection of a positive value below this same line. This positive value is only questioned in the case of the advanced individuals, but there are few individuals.

Discussion

The results depend on the experimental conditions that enabled the observed values for temperature and irradiance. When looking at the distribution of these values (Figure 1, middle graphs), we note many areas without observations and two areas/peaks where the observed values are concentrated (unbalanced design).
An interesting result produced by SpiceFP is the difference between the delayed and advanced individuals, the latter being less sensitive to the temperature-irradiance couple. It could be interesting to compare with the sugar level in the berries that is a usual indicator of technological maturity. As this (destructive) measure was not available, the "advanced" or " delayed" states were defined on the basis of the observed values of the Ferari index. Even if this measure of ripening state is not the most reliable/frequent one, our results suggest that it makes sense. One hypothesis to explain the difference in sensitivity between the two states is that the colour of the berry could be due to an acceleration of its physiological cycle, including the accumulation of anthocyanins regardless of climatic factors. The advanced individuals may have a faster physiological cycle and have reached a state that is less sensitive to climatic factors. The decrease in sensitivity during maturity could be due to a slowing down of the accumulation determined by the "end" of a physiological process (similar to the end of a leaf growth which becomes insensitive to temperature). Another possibility can be a greater difficulty in detecting the effects of climatic factors for the so-called advanced individuals.

Conclusion

The use of SpiceFP combined with a linear separation model identified a negative impact of morning combinations of low irradiance (below 100 or 45 µmol m^{-2} s^{-1} depending on the advanced-delayed state) and high temperature (above 25°C). These combinations correlated with a delay in the early phase of anthocyanin accumulation. For individuals with a low Ferari index at the beginning of week 3 (FI<0.4), a linear separation was identified between more and less favourable conditions, that was clearer for relatively warm nights (between 22°C and 26°C).
Overall, our results are consistent with the literature (Fernandes & al., 2015) which suggests that, in natural environments, high temperatures over a long period of time have a negative impact on anthocyanin biosynthetic pathway (Spayd & al, 2002; Downey, 2006) whereas low temperature associated with high irradiance favour anthocyanin accumulation (Cohen & al, 2008).

References

Agati G., Meyer S., Matteini P. & Cerovic Z. G. (2007). Assessment of anthocyanins in grape (vitis vinifera l.) berries using a noninvasive chlorophyll fluorescence method. *Journal of agricultural and food chemistry*, 55(4), 1053–1061. https://doi.org/10.1021/jf062956k

Ben Ghozlen N., Cerovic Z., Germain C., Toutain S. & Latouche G. (2010). Non-destructive optical monitoring of grape maturation by proximal sensing. *Sensors (Basel, Switzerland)*, 10, 10040–68. https://doi.org/10.3390/s101110040

Cohen S., Tarara J. & Kennedy J. (2008). Assessing the impact of temperature on grape phenolic metabolism. Analytica chimica acta, 621, 57–67. https://doi.org/10.1016/j.aca.2007.11.029

Downey, M., Dokoozlian, N. & Krstic, M. (2006). Cultural practice and environmental impacts on the flavonoid composition of grapes and wine : A review of recent research. *American Journal of Enology and Viticulture*, 57 :257–268.

Fernandes de Oliveira, A., Mercenaro, L., Del Caro, A., Luca, P. & Nieddu, G. (2015). Distinctive anthocyanin accumulation responses to temperature and natural uv radiation of two field-grown vitis vinifera l. cultivars. *Molecules*, 20. https://doi.org/10.3390/molecules20022061

Gnanguenon Guesse, G. (2021). Modélisation et visualisation des liens entre cinétiques de variables agro-environnementales et qualité des produits dans une approche parcimonieuse et structurée. *PhD thesis*, *Université Montpellier, ED I2S Biostatistique.*

Gnanguenon Guesse G., Loisel, P., Fontez, B. & Hilgert, N. (2021). SpiceFP: Sparse Method to Identify Joint Effects of Functional Predictors. R package version 0.1.0. https://CRAN.R-project.org/package=SpiceFP

Rienth M., Torregrosa L., Kelly M. T., Luchaire N., Pellegrino A., Grimplet J. & Romieu C. (2014).. Is transcriptomic regulation of berry development more important at night than during the day? *PLoS ONE*, 9(2). https://doi.org/10.1371/journal.pone.0088844

Spayd S., Tarara J., Mee D. & Ferguson J. C. (2002). Separation of sunlight and temperature effects on the composition of vitis vinifera cv. merlot berries. *American Journal of Enology and Viticulture*, 53, 171–182.

Stekhoven, D. J. & Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.

Tarara J.M., Lee J., Spayd S. E. & Scagel C. F. (2008). Berry temperature and solar radiation alter acylation, proportion, and concentration of anthocyanin in merlot grapes. *American Journal of Enology and Viticulture*, 59(3), 235–247.