A climate index proposal for the wine sector: a descriptive statistical approach

Aníbal Galindro[1], Ana Alexandra Marta-Costa[2], Adelaide Cerveira[3]; João Matias[4][a]

[1] University of Trás-os-Montes e Alto Douro (UTAD, www.utad.pt) and Centre for Transdisciplinary Development Studies (CETRAD), Portugal, anibalg@utad.pt.

[2] University of Trás-os-Montes e Alto Douro (UTAD, www.utad.pt) and Centre for Transdisciplinary Development Studies (CETRAD), Portugal, amarta@utad.pt

[3] University of Trás-os-Montes e Alto Douro (UTAD, www.utad.pt) and INESC-TEC, Portugal, cerveira@utad.pt

[4] University of Trás-os-Montes e Alto Douro (UTAD, www.utad.pt), and CMAT-UTAD, Portugal, j_matias@utad.pt [a]

[a] Corresponding author

Abstract. Understanding the role of the climate on the wine production is one of the major concerns of this sector since the environment usually determines the output of this industry. There are only a few previous studies that attempted to compile these environmental effects as an index, usually considering the temperature and the precipitation as their core variables. The present study suggests a new climate index which is based on descriptive statistics. Our index tries to mimic the target region characteristics and avoid the past studies premise of imposing previously conceived restrictions such as a fixed optimal climate. We then used yearly production and daily temperature data (1950-2016) from the Portuguese Minho wine region to test our proposed index and compare it with Ribéreau-Gayon and Peynaud (RGP, Ribéreau-Gayon et al., 2003) and Growing Degree-Days (GDD, Winkler et al., 1974) indexes. Our results showed that the newly proposed index may outperform the explanatory power of the other indexes and, in addition, may output interesting and unknown characteristics such as the different ideal temperatures regarding the studied region.

1 Introduction

The climate plays an important role on the wine production since its quality and quantity is strongly dependant on the inherent weather conditions as stated on Malheiro et al. [1] and Salinger et al. [2]. That strong bond made the yearly grape quality become a feature included in the study of the effects of the climate change [3,4]. Even though many individual climate factors make their contribution in the grape development and the following grape quality (*e.g.* solar radiation, wind, humidity, among others), the temperature and the water supply are among the most important [5].

The vines are capable to grow in a wide range of climatic conditions, nonetheless the most remarkable wine-growing regions are located between the 35th and the 50th parallels in the Northern Hemisphere and between the 30th and the 45th parallels in the Southern Hemisphere [6]. During the growing season, the extreme high temperatures can affect the grape quality and the vineyard overall production in many ways,

inducing high grape mortality or total failure of flavour ripening [7].

In the opposite way, low and frosting temperatures can also harm the grape flavour and colour due to the slower grape fermentation. In the worst-case scenario extreme cold can also cause the total wipe-out of the vineyard production. Summarily, Gladstones [8] suggests that the grape growing process usually requires mild and narrow ranged temperatures (avoiding the extreme heat or the extreme cold) and actually, according to Gladstones [9] and Jones [10] the most remarkable quality wines are associated with low frost damage and mild winters (January, February, March), early flowering, and warm springs (April, May, June) and optimal maturation associated with short range variability in the summer temperature (July, August, September).

The temperature has such a heavy impact on the grape growing process that the different vine stages can be predicted by simple single based temperature models [11]. The amount of the precipitation that a vineyard receives is also an important feature that determines the following grape quality and output. During the grape

maturation period a vineyard should only be exposed to a certain amount of precipitation since its excess may cause major quality issues, including disease susceptibility and rotten grapes. The lack of precipitation in every non-irrigated farm may also cause production problems such as shrinking berries, and of course vineyard mortality and low production levels [12]. Nonetheless, a moderate water deficit may reduce the grape yield but improve the grape quality [13,14]. This work gathers information about the climate impact on the wine production and presents an overview of the current bibliography about the existent wine-oriented climate indexes. Afterwards a new climate index based on descriptive statistical parameters is created and evaluated using a simple case study of the Portuguese Minho wine production. In order to test the proposed index, we compare its explanatory power upon the Ribéreau-Gayon and Peynaud (RGP) [15] and Growing Degree-Days (GDD) [16] indexes.

The article is divided in four sections. Besides the overview of environmental effects on the vineyard production, the next section introduces the notion of climate index, with a few examples from the available bibliography, and presents in more detail two indexes (GDD and RGP indexes). The New Index is described in section three. Finally, section four summarizes the main conclusions and suggestions for further research.

2 Climate Index

The weather conditions are roughly random and regarding the vineyard farms there is not much to do to avoid extremal conditions, apart from irrigated farms that can overcome dry and low precipitation periods. The effects of the weather conditions, particularly the temperature and the precipitation have been the subject of many studies [8,17-19] due to their strong link with good grape quality and the following premium wines. With such randomness brought into the wine sector and overall agricultural production there's also a substantial research on risk protection using financial assets such as weather derivatives as stated by Zara [20] or Leggio [21]. Those studies also consider weather indexes to determine the asset price and the possible hedging strategies.

The great majority of the bioclimatic indexes that are typically used to evaluate the wine production calculate the accumulated heat over the growing season (often since October 1st until April 30th). This growing season index originally stated on Winkler et al. [16] is usually called Growing Degree Days (GDD) or Winkler Index (WI) and it is used to describe a general climate and determine if it is suitable for wine production. Another remark about the WI is that the author defines five

possible temperature outcomes (1-very cold until 5-very hot).

The Huglin Index (HI), in Huglin [22], follows the same premise of WI, although, it differs in some features since it uses simultaneously the mean and maximum temperatures in the daytime temperature estimation. It is also calculated using a shorter 6-month period rather than seven. Furthermore, the HI incorporates a "length of the day" coefficient into their calculations which is self-explanatory.

The research by Stock et al. [23] showed that it is expected an overall northern hemisphere increase of 100 to 600 units in the HI until 2050. That fact may indicate a latitudinal shift of the grapevine cultivation in Europe, with new areas on the northern territories becoming more suitable to wine production opposed to the southern ones which may become inadequate due to the excessive heat.

Jones et al. [18] demonstrated that the simple average growing season temperatures on WI and HI are associated to quality vintages. The biologically effective degree-day index (BEDD) created by Gladstones [9] is also a heat sum index. One difference between HI and BEDD is the way that the daytime temperature is estimated. The BEDD index incorporates a new feature which is calculated based on the diurnal temperature range. This index is considered overly adjusted if the diurnal temperature range is greater than 13° C, and it is under adjusted when the temperature range is less than 10° C.

Other available indexes compile more than the cumulative temperature such as Branas et al. [24] which suggested an index that compiles the daily average temperature multiplied by the daily precipitation in order to estimate the risk of the downy mildew disease appearance. Ribéreau-Gayon et al. [15] created Ribéreau-Gayon and Peynaud bifactorial hydrothermal scale also known as the RGP index (RGP), even though it presents the similar cumulative temperature approach as its predecessors, the RGP also considers the daily precipitation and shifts the index cycle, starting at 1st April until 30th October. Fregoni [25] actually found a meaningful relationship between the RGP index and the Pinot Noir grape production. Similarly, Zara [20] also validated the RGP index utility for their hedging strategy development.

Other authors successfully connected different indexes such as Tonietto and Carbonneau [26]. They combined HI, the Dryness Index (DI) and the Cool Night Index (CI) to classify the climate stations in featured wine regions. Nemani et al. [27] also combined the WI and extreme event frequencies to highlight possible shifts and expansion/contraction of grapevine areas, while Fraga et al. [28] used the natural variables such as

minimum/maximum temperature and precipitation as straight-forward model regressors in a logistic model. Even though many features can be considered and included when the climate suitability is evaluated, the temperature grabs the first place as the most relevant and unavoidable (at least on viticulture).

3 A new climate index

In this section we will purpose a new index (NI), but in order to furtherly compare it to the available previous studies, we describe the aforementioned indexes GDD (WI) and RGP. These two indexes were selected since they use the temperature value, which is the core variable of the model of this work.
The specific formulation of GDD and RGP lays on the expression (1) and (2) respectively.

$$GDD = \sum_{i=1}^{n} max\{T_i - 10, 0\} \qquad (1)$$

$$RGP = \sum_{i=1}^{m} max\{T_i - 10, 0\} - P_i \qquad (2)$$

where *n* is equal to 214, corresponding to the number of days between April 1st and October 31th, and *m* is equal to 212 or 213, corresponding to the days between 1 October and 30 April, for each year.
Each index *i* corresponds to a single day, and T_i is the average daily temperature recorded during i^{th} day; P_i is the amount of rainfall in the ith day.

The RGP and GDD have quite similar expressions, in fact the RGP only shifts the analysed timeline and adds up the precipitation variable. One feature that immediately comes upfront is their lethargic nature since they both have a fixed reference temperature (10º C). According to Nemani et al. [27] the ideal temperature for grape growing changes across the whole production process, therefore, settling an average temperature (RGP and GDD) may generalize the calculations too much.
The main goal of this work is to find a more suitable way to consider the weather effects on the wine production and overcome some flaws pinpointed on the featured two indexes. The NI is suggested based on the descriptive statistics of the data, such as Kurtosis and Skewness which measures the tailedness and asymmetry of the distribution [29].

$$Kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{y})^4 / N}{s^4} \qquad (3)$$

$$Skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{y})^3 / N}{s^3} \qquad (4)$$

For univariate data $Y_1, Y_2, ..., Y_N$ the Kurtosis and Skewness values are given by equations (3) and (4) respectively, where s represents the standard deviation of the data, while $\bar{y}$ represents its mean value. Apart from the newly introduced parameters we also adopt a different timeline suggested by Gladstones [9] and Jones [10] in order to better define the vineyard lifecycle. Therefore, we consider three different periods of three months each (equation 5).

$$NI = \sum_{j=1}^{3} \left| \left| \frac{\sum_{i=1}^{N_j}\frac{(Y_{ji}-\bar{y}_j)^4}{N_j}}{s_j^4} - k_j \right| + \left| \frac{\sum_{i=1}^{N_j}\frac{(Y_{ji}-\bar{y}_j)^3}{N_j}}{s_j^3} - \gamma_j \right| + |\bar{y}_j - t_j| \right| \qquad (5)$$

Where *j* = 1 corresponds to the first time interval with N_1 days, the period between the 1st of January and 30th of March; *j* = 2 corresponds to the second time interval with N_2 days, the period between the 1st of April and 30th of June and, finally, *j* = 3 with N_3 days, corresponds to the period between the 1st of July and 30th of September. The parameter $\bar{y}_j$ is the average temperature for each period *j* and the parameters k_j, γ_j, t_j are, respectively, the optimal value of kurtosis, skewness and mean temperature for the correspondent period, which will be explained in the following.

The suggested index tries to mimic the descriptive statistical values of a target distribution. Therefore, to calculate these "optimal" values we fitted a polynomial single regression (equation 6).

$$Z = \beta_0 + \beta_1 X + \beta_2 X^2 + e \qquad (6)$$

The dependant variable *Z* corresponds to the production values (in litres) and the single explanatory variable *X* is the kurtosis, skewness, or the mean temperature values for each period *j* (*j* = 1,2,3).
Considering 67 values for variable *Z* of Minho's green wine production from 1950 until 2016, collected from CVRVV [30,31], nine different polynomial single regression were performed to obtain the three optimal parameters for each period and each variable.
We gathered the daily averaged temperature data and precipitation values from the E-OBS observational interpolated/gridded dataset, version 15 [32]. Despite a few limitations explained at Hofstra et al. [33], this dataset provides uninterrupted and homogeneous gridded fields of daily average temperatures and daily cumulative precipitation over Europe from 01-01-1950 until 31-12-2016, on an approximate 27 km spatial

resolution. The climatic data is extracted from a 16 square grid-box (41.12º to 42.12º latitude and 7.87º to 8.87º longitude) that covers the Minho territory. Considering those 16 values we calculate their total average which in the end samples a good approximation to the Minho's overall weather conditions.

First, we test our available data in the previously mentioned GDD (WI) and RGP indexes.

There are some interpretation difficulties associated with the polynomial regression. These problems can be surpassed performing algebraic manipulations, according to Stimson [34], and equation (6) can be reparametrized as equation (7):

$$Z = M + \beta_2(F - X)^2 \tag{7}$$

where M is the minimum/maximum value of Z (equation 8) while F is the minimizing (for convex, $\beta_2 > 0$) or the maximizing (for concave, $\beta_2 < 0$) value of Z (equation 9).

$$M = \beta_0 - \frac{\beta_1^2}{4\beta_2} + e \tag{8}$$

$$F = \frac{-\beta_1}{2\beta_2} \tag{9}$$

Table 1 displays the nine Polynomial regression results and the target F calculated as Stimson [34]. Each regression was performed individually settling the dependant variable as the yearly production upon of kurtosis k_j, skewness γ_j and mean temperature t_j for each period $j = 1,2,3$ corresponding to each previously mentioned trimester. The kurtosis values that maximize the output tend to be higher than three (heavier tails than the normal distribution) while skewness values should linger around zero (similar to the normal distribution).

Table 1. Polynomial regression results of Production (dependent-variable) upon kurtosis, skewness and mean temperature for each period

Variable X	β_1	β_2	F
k_1	76012408*	-5590239	6.7987
k_2	92450299*	-10763386*	4.2947
k_3	75220789*	-5819973*	6.4623
γ_1	-102728129*	432115254*	0.1189
γ_2	135225197*	165552672*	-0.4084
γ_3	-68206445	570990174*	0.0597
t_1	65840653*	-5500094*	5.9854
t_2	54500304*	-2988560*	9.1182
t_3	47406145*	-2033064*	11.6588

* Statistically significant results for a 5% level of significance. Source: Own elaboration with CVRVV [30,31] data.

There is not much to say about t_1, t_2, t_3 since they represent the target mean temperature values, although the increasing temperature $(t_3 > t_2 > t_1)$ displays the typical differences between Winter, Spring and Summer respectively. That is actually another concern, since the negative regression coefficient (β_2) illustrate that F accounts the value where the production is maximized. On the other hand, that doesn't apply to the skewness values where $\beta_2 > 0$, since F retrieves a minimum for γ_1, γ_2 and γ_3. Given that, we need to accommodate equation (5) for each situation and also ensure that the three variables contribute equally to the final result. To overcome that problem, we normalized each variable within [0,1] settling 1 as the major penalty to the overall result (Table 3) while 0 represents the optimal target. Since the parameter F absorbs different meanings we need to settle a set of rules for each F representation (while *x* denotes a single observation), presented in Table 2.

Table 2. Different boundaries per each polynomial regression result

	$x_{i.} = F - \omega$	$x_{i.} = F$	$x_{i.} = F + \omega$
F is a Maximum (A)	1	0	1
F is a Minimum (B)	0	1	0

Where:
$$\omega = \max\{|F_i - a_i|, |F_i - b_i|\}, \tag{10}$$
$$X_i = \{x_{i1}, \ldots, x_{in}\}, \quad a_i = \min\{X_i\}, \quad b_i = \max\{X_i\}$$

After gathering information about the F representation, we settle the normalization process accordingly (Table 3).

Table 3. Normalization scenarios per each F formulation

	Normalization Formula
Scenario A	$x'_{in} = \dfrac{|x_{in} - F|}{\omega}$
Scenario B	$x'_{in} = 1 - \dfrac{|x_{in} - F|}{\omega}$

The equation (5) only gives us the general sum and dispersion guideline when F represents a maximum. But technically each F can be a maximum (A) or a minimum (B) suggested value to place on equation (5), therefore the input calculation varies accordingly. Substituting each optimal $k_j; \gamma_j; t_j$ and the temperature values Y_{j_i} on equation (5) we gather a set of raw values that don't acquaint their F as a maximum/minimum

value. By labelling them properly according to Table 2 and applying the normalization on Table 3 the final index data series (67 observations, 1950-2016) is obtained.

To test our index, we performed an Ordinary Least Squares such as Fregoni [25] and Cossu et al. [35] opposed to other indexes already presented during this work (RGP, GDD). Equation (11) relates each index value (*D*) with the production (*Z*).

$$Z = \beta_0 + \beta_1 D + \varepsilon \qquad (11)$$

The Table 4 displays the OLS regression output coefficients for the three different indexes considering the Minho region production values as the dependent variable.

Table 4. OLS regression output for three different indexes

Index	β_0	β_1	R^2
RGP	1.421^8*	-1.651^3*	0.07545
GDD	482586786*	-231452*	0.251
NI	569342373*	-66472298*	0.3036

* Statistically significant results for a 5% level of significance.

The first remark that we've notice is the substantially low *R*-Squared values, that fact may be explained by the omission of many other explanatory variables that actually also affect the production output, such as the harvested land, the economic policies and market environment. Nonetheless, our suggested model manages to achieve a higher *R*-Squared upon the other two indexes. All three indexes reveal a negative coefficient β_1 indicating that increments on each index value may result in a decrease on the output variable (*Z*) – production as expected (since higher index values logically represent worse weather conditions). The straightforward interpretation of our suggested index (NI) is that the more the distributional weather values drift away from the settled as optimal, the smaller the production values should be. Other advantage of the NI against the other common indexes is that it doesn't settle arbitrary bounds since it is constructed to evaluate a specific production environment. Therefore, the optimal values $k_j; \gamma_j; t_j$ should vary upon different regions and adapt to different environment in order to become a versatile and suitable index. A well calibrated index may play an important role in production hedging calculations resembling Zara [20] or efficiency related works where the climate is quite responsible for apparent outliers and general data noise. Nonetheless, we present a quite complex formulation reinforced by the necessary set of rules upon the *F* values.

4 Conclusions

This empirical work tries to understand the role of the climate effects upon the wine industry production. After considering the available bibliography regarding the environment effect measures (climate indexes) we noticed that even though each existing index underpins and justifies its features in solid theoretical background, all of them used pre-formulated values that generalizes the index application apart from the production geographic position. Therefore, our main goal was to provide an index that can be adjusted individually per each wine production that we are actually studying. Our suggested index assembles descriptive statistics features such as the temperature mean, the kurtosis and skewness with optimal values calculated according to Stimson [34] polynomial regression approach. Regarding the Minho region and our model formulations, generally heavy-tailed and non-skewed (roughly normal) temperature distributions during the three period subsets are suggested in order to maximize target region output. In order to evaluate our index performance, we made an OLS regression such as Fregoni [25], Zara [20] and Cossu [35] to study if the suggested index actually does have better explanatory capabilities than the others. Using Minho's region wine production dataset our proposed index reveals himself statistically significant and also outperforms GDD and RGP, considering explanatory power parameters. For further development it is suggested to add more environmental parameters such as the precipitation to the model in order to increase his accuracy, the NI also computes the observation dispersion linearly on the objective function (OB), so, it might be interesting to test non-linear OB formulations. All inputs (temperature, kurtosis and skewness) were weighted equally, changing the formulation weights may also display important and relevant information since we might overvaluing/undervaluing the parameters. Such study may zoom even further geographic characteristics and find the core parameter for each specific farm. It would be also interesting to see a further NI application on vineyard hedging problems or efficiency-based studies.

References

1. A.C. Malheiro, J.A. Santos, H. Fraga, J.G. Pinto, Clim. Res. **43**,3, 163–177.(2010)
2. M.J. Salinger, M. Baldi, D. Grifoni, Int J Biometeorol. **59**,1799 (2015)
3. G.V. Jones, L.B. Webb, Journal of Wine Research **21**, 103-106 (2010)

4. B. Seguin, *Coup de chaud sur l'agriculture*. Editions Delachaux et Niestlé, Paris. (2010)

5. B.G. Coombe, Acta Horticulturae, **206**, 23-35 (1987)

6. C. Leeuwen, P. Darriet, Journal of Wine Economics, **11**,1,150-167 (2016)

7. M.G. Mullins, A. Bouquet, L.E. Williams, *Biology of the Grapevine*, Cambridge University Press, Great Britain, 239 (1992)

8. J. Gladstones, *Wine, Terroir and Climate Change* . Kent Town, South Australia: Wake field Press (2011)

9. J. Gladstones, *Viticulture and environment* Winetitles: Adelaide (1992)

10. G.V. Jones, Arbor Phaenol **42**, 3-7 (1999)

11. A. Parker, C.A.I. Garcia, C. van Leeuwen, I. Chuine, Australian Journal of Grape and Wine Research, **17**(2), 206-216 (2011)

12. T. Popovic, S. Mijovic, R. Pajovic, Agriculture & Forestry, **59**, 2, 137-145 (2013)

13 B. Basile, J. Marsal, M. Mata, X. Vallverdu, J. Bellvert, J. Girona, American Journal of Enology and Viticulture, **62**, 452-461 (2011)

14 D.S. Intrigliolo, D. Perez, D. Risco, A. Yeves, J.R. Castel, Irrigation Science 30, 339-349 (2012)

15 P. Ribéreau-Gayon, *Trattato di enologia*, Ed agricole, Milan (2003)

16. A.J. Winkler, W. Cook, W. Kliewer, L. Lider. *General Viticulture*, University of California Press, Berkeley (1974)

17 G.V. Jones, Acta Horticulturae (ISHS), **689**, 41-60 (2005a)

18 G.V. Jones, M.A. White, O.R. Cooper, K. Storchmann, **73**, 319-343 (2005b)

19 G.V. Jones, E. Duchene, D. Tomasi, J. Yuste, O. Braslavksa, H. Schultz, C. Martinez, S. Boso, F. Langellier, C. Perruchot, G. Guimberteau, GESCO 2005 (2005)

20 C. Zara, International Journal of Wine Business Research, **22**, 3, 222-237, (2010)

21. K.B. Leggio, Managerial Finance, **33**, 4, 246-252 (2007)

22 .P. Huglin, C R Acad. Agric. France, 1117-1126, (1978)

23. M. Stock, F.W. Gerstengarbe, T. Kartshall, P.C. Werner, Acta Hort, **689**, 29-39 (2005)

24. J. Branas, G. Bernon, L. Levadoux, *Éléments de viticulture générale*. Delmas, Montpellier, France (1946)

25. M. Fregoni, *Viticulture of Quality*, Editore Tecniche Nuove, Milan (2005)

26. J. Tonietto, A. Carbonneau, Agricultural and Forest Meteorology, **124**, 81-97 (2004)

27. R.R. Nemani, M.A. White, D.R. Cayan, G.V. Jones, S.W. Running, J.C. Coughlan, Climate Research, **19**(1), 25-34 (2001)

28. H. Fraga, A.C. Malheiro, J. Moutinho-Pereira, J.A. Santos, Agric. For. Meteorol. **185**, 26-36 (2014)

29. D.J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall /CRC (2011)

30. CVRVV, Produção Vinho Verde 1934-1998. Availa-ble at http://portal.vinhoverde.pt/pt/estatisticas#! (accessed at 15/08/2017) (2017)

31. CVRVV, Produção Vinho Verde e Vinho Regional Minho 1999-2016. Available at http://portal.vinhoverde.pt/pt/estatisticas#! (accessed 15/08/2017) (2017)

32. M.R. Haylock, N. Hofstra, A.M.G.T. Klein, E.J. Klok, P.D. Jones, M. New, J. Geophys. Res. **113** (2008)

33. N. Hofstra, M. Haylock, M. New, P.D. Jones, J. Geophys. Res. **114** (D21) (2009)

34. J.A. Stimson, E.G. Carmines, R.A. Zeller R.A. Sociological Methods and Research, **6**, 515-523, (1978)

35. A. Cossu, A. Bodini, E. Fronteddu, LInformatore Agrario, **21** (2004)