

# Linking viticultural climatic indices to grape phenology in the South Tyrolean Alps

Andrea Balotti<sup>1,a</sup>, Simon Tscholl<sup>1</sup> and Lukas Egarter Vigl<sup>1</sup>

<sup>1</sup>Eurac Research, Institute for Alpine Environment, 39100 Bolzano, Italy

**Abstract.** Climate indices based on heat accumulation, e.g. the Winkler index, are widely used to define the climatic niches for vines. In this study, we investigate how a combined use of high-resolution (25 m) climate index maps and phenology records from 30 vineyards in South Tyrol (in the Italian Alps) can help to (i) assess viticultural suitability across a mountainous landscape and (ii) estimate the timing of physiological processes of Pinot Noir development (ripening and must weight) at various sites across the region. First, the best interpolation method is chosen (from multiple linear regression (MLR), regression kriging or support vector regression) to create maps of climate indices averaged over the time period 1991-2010. Second, correlation is calculated between the timing of phenological stages of Pinot Noir for the year 2017 and various climate indicators, such as temperature-based indices (Winkler, Huglin, Biologically Active Degree Days, Cool Night, Fregoni) and average (GST), minimum and maximum temperature over the growing season. The MLR method is shown to yield the best interpolations of the climate indices across the complex terrain of the study area. The Winkler and GST indices correlate most precisely with the late-season phenological events of the study sites, and are thus the most predictive. These findings demonstrate the potential of climatic maps to effectively define suitable areas for grape growing and estimate ripening dates in South Tyrol.

## 1 Introduction

Unlike other famous Italian wine regions, such as Tuscany or Veneto, which are characterized mainly by hilly topography, South Tyrol is characterized by complex topography over a large elevation range. On the one hand, this is an advantage because many different climatic niches are available for grape growing. On the other hand, the amount of available cultivation area in each niche is limited. The latter is especially true in South Tyrol, where lower elevations are widely used for other cultures, like orchards, and higher elevations are often characterized by conditions that are too harsh for viticulture.

Viticultural suitability is strongly influenced by temperature [1], and thus having accurate maps of temperature-related indices is the first step in obtaining realistic climatic zoning of an area.

The complex topography of mountainous landscapes, however, represents a large challenge when spatializing environmental variables. In recent years, different efforts have been made to develop gridded temperature datasets specifically designed for complex topography. However, these maps are usually in spatiotemporal resolutions that are too coarse for effective use in small-scale viticulture studies. Thus, in the present study we decided to perform an in-house interpolation of temperature-based indices to create a set of maps that are suitable for defining climatic zones for this purpose.

High-resolution maps (25 m) of climatic indices are interpolated using different geostatistical approaches (multiple linear regression, regression kriging, support vector regression) and data from the official national meteorological service from 1991-2010. The best interpolation method is selected to create maps of climate indices that define average climate zones in South Tyrol for this period.

Using data from meteorological stations installed in 30 different vineyard plots, correlation is calculated between various temperature-based indices and the timing of phenological stages of Pinot Noir for the year 2017. The goal is to find the best explanatory variables to estimate ripening date at any specific location.

The 30 vineyard stations were installed as part of the transnational EU-funded “REBECKA” project (Interreg V-A IT-AT:ITAT1002, duration: 2017-2019) between South Tyrol and Carinthia (Austria). The goal of the project is to assess the parameters that influence the vines of Pinot Noir in these Alpine regions and how climate change can impact them. One of the aims of the project is to monitor the phenology of grapevines and the chemical characteristics of grapes, and to correlate them with temperature-based indices.

This work presents the first results of the project, focusing on South Tyrol and on the parameters that have the largest influence on viticulture.

## 2 Material and methods

### 2.1 Study area

South Tyrol is located in Northern Italy, in the central-eastern Alps. From a topographic point of view, South Tyrol has a large elevation gradient, with a range from 200 to 3900 m, but with most of the territory situated between 1000 and 2500 m. This mountainous area is crossed by three interconnected main valleys that are relatively wide and with different orientation. The vineyards are distributed along these three main valleys from 200 to 1300 m. Due to its position in the middle of the Alps, the climate is influenced by both continental and Mediterranean factors, but in general conditions typical of subcontinental and continental areas can be found.

If we look at the climate in South Tyrol, and more generally in the central-eastern Alps, we find a warming trend in the annual mean temperature from 1980 to 2000, which subsequently flattens out.

Seasonal temperature trends are less consistent: spring and summer have seen increased mean temperatures, while winter and autumn have been much more variable. Additionally, hot days (maximum temperature above 25 °C) are occurring more frequently, while frost days (maximum temperature below 0 °C), on the other hand, show a decreasing trend [3].

Wind fields are complex and driven by general circulation, pressure gradient and slope. Pressure driven winds from the north, called Föhn, have an important role in northwestern areas of South Tyrol, where the driest part of the study area is located.

Precipitation in South Tyrol is driven mainly by mesoscale systems and, during summer, also by convective activity [3]. There are no discernable trends in climatic analysis, yet there exist differences in the spatial distribution of precipitation in South Tyrol.

### 2.2 Datasets

The values of dependent variables (i.e. climatic parameters) were obtained by processing the data from existing meteorological stations in South Tyrol. We have collected the daily value of minimum and maximum temperatures from 1990 to 2017 from a pool of 82 stations, all below 2000 m. Using the package *ClimClass* [4] in *R environment*, we have calculated various climate indices, for each year and station. The calculated climate indices and their abbreviations are reported in Table 1.

The years between 1991 and 2010 are the reference period over which climate averages are computed. This timespan was selected because it aligns with the majority of the WMO climatological “standard normal” period (1981-2010) [5]. The climate indices are also calculated for 2017 to find a correlation with phenological stages.

Of the pool of 82 stations, three were used as final validation points, whereas the remaining stations were used as training points for the models. The number of training stations with meaningful data each year varies from 44 to 58 (if we consider the period 1991-2010) and

gives a sample density of 0.23-0.30 stations per square kilometer.

**Table 1.** Acronyms of climate indices used.

Climate index name	Abbr.
Winkler index	WI
Huglin index	HI
Biologically Active Degree Days	BEDD
Cool Night Index	CNI
Fregoni’s Simplified index	FSI
Growing Season Temperature	GST
Maximum Growing Season Temperature	GSTmax
Minimum Growing Season Temperature	GSTmin

### 2.3 Interpolation methods

In order to map climate indices across the landscape, data from meteorological stations must be interpolated using the most appropriate method, either multiple linear regression (MLR), regression kriging (RK), or support vector regression (SVR).

The MLR assumes the existence of a relationship between a dependent variable and its (independent) predictors that can be explained using a linear function [6]. Using independent predictors doesn’t mean that all of them are statistically significant to describe the dependent variable. For each dependent variable, the statistically significant predictors were assessed using the R package *leaps* and the method for cross-validating all the possible models described by James et al. [7]. The statistic used to select the best predictors is the BIC value. This method retains only the predictors that are statistically significant for the model.

The RK method combines a linear regression and a variogram analysis of the linear regression residuals [8], [9]. The variogram is used to construct a distance-based model for the residuals that are later interpolated with ordinary kriging, and then added to the linear regression interpolation. If the variogram of residuals has no spatial correlation, theoretically, the RK is composed only by the linear part as a normal linear regression. But in practice, it is still possible to model the residuals as a constant value added to the linear regression (nugget model). Our interpolation script is constructed to first try a variogram fit with an exponential model and, in case of failure, to adopt a nugget model. In this way, it is always possible to construct a model for the residual.

The SVR method uses a non-linear regression formula [10]. Since the predictor selection method used in MLR (and consequently in RK) is developed for linear regression, the same method is not applicable for SVR and a specific predictor selection algorithm should be developed. In literature, this topic is still debated (see [11], [12]), so we decided to retain all of the predictors for SVR.

### 2.4 Predictors

The initial predictors we had selected for the interpolation process were the same used in Le Roux et al [13]: elevation, slope, aspect (North-South and West-East

components), longitude, latitude, and potential solar radiation and direct solar radiation duration, both summed over the vegetation period (April 1<sup>st</sup> – October 31<sup>st</sup>). All these predictors are computed directly in R except for potential solar radiation and solar radiation duration, which are created within ArcGIS using the *Area Solar Radiation* tool and its default parameters.

Because vineyards in South Tyrol cover a relatively small area (only 0.2 hectares on average), we have decided to use a resolution of 25 m as a good compromise between computational time and accurately assessing spatial variability. Aspect was processed in order to remove circular geometry and avoid problems in computation [13]. In particular, we have split North-South and West-East components, computing sine and cosine of the aspect angle respectively. Longitude and latitude are UTM coordinates expressed in meters. Potential solar radiation and direct solar radiation duration over vegetation period, were computed first at 50 m of resolution and then resampled to 25 m using bilinear interpolation in order to speed up computation.

Before using these predictors in the interpolation process, a correlation analysis is necessary to detect signs of multicollinearity. In general, the predictors don't show signs of strong correlation that could influence the regression, but potential solar radiation shows a moderate correlation with elevation, irradiance duration and slope with values of 0.63, 0.44 and 0.43 respectively. Moreover, slope has a correlation of 0.44 with elevation. We have therefore decided to apply a principal components analysis on predictors in order to reduce the multicollinearity effects as suggested in Hengl et al [8]. As expected, the new predictors show a negligible correlation and are thus used in the interpolation process.

Climate conditions above a certain elevation become too cold for effective viticulture. We therefore decided to restrict all subsequent computations to elevations below 2000 meters.

## 2.5 Phenology data

In 2017, in the scope of the “REBECKA” project, we installed 30 meteorological stations: 27 in vineyards with Pinot Noir vines and 3 near meteorological stations of the South Tyrolean provincial network to assess the measurement quality of our stations. Our meteorological stations are equipped with two temperature sensors: one for the standard 2-m air temperature measurement, and one positioned at 0.9 m, at the base of the canopy. In addition, there are solar radiation sensors, and soil temperature and soil water potential sensors installed at 40 cm below the soil surface.

Due to time constraints, it was not possible to start the measurements at the beginning of the vegetation period (April 1<sup>st</sup>), so a gap filling procedure has been adopted to reconstruct minimum and maximum 2-m air temperature until that date. To do this, a linear regression model is trained between the data of each “REBECKA” station and the data of the nearest provincial station. The training process uses the data from the whole period where the “REBECKA” station was recording. The model is then

used to estimate the missing data of our stations from the measured temperature of the respective nearest provincial station. The performance of the model is validated via a 5-fold cross-validation, which is repeated 3 times. This process proves quite effective for most stations. The highest error rate is around 1.7 °C, average mean absolute error is 0.86, average  $r^2$  is 0.97 and 0.98 is the average Pearson's correlation. The result of this gap-filling procedure is a table with daily measurements of maximum, minimum and average temperature starting from April 1<sup>st</sup> for every “REBECKA” station.

For each site with a meteorological station, surveys to detect the day-of-year (DOY) of phenological stages are performed. In particular, DOY of budburst, flowering and ripening (respectively the level 8, 65, 81 on the BBCH scale [14]), as well as DOY of reaching 16, 17 and 18 °KMW for must weight are assessed.

## 3 Data analysis

### 3.1 Interpolation results

To compare the performance of interpolation methods, we have computed RMSE and MAE statistics, as well their difference, averaged over the years 1991-2010 (Figure 1). Since the indices have different units of measurement, statistics are normalized using the range of observed values used to compute the statistics. The results show that, looking at the model training, the SVR is the best method for temperature-sum indices (WI, HI, BEDD) and RK is best for pure temperature indices (CNI, GST). But if we look at the statistics of validation points, MLR and RK show the best performance, as reported in Table 2.



**Figure 1.** Average performance of models over training and validation points. Statistics are normalized to permit comparison between indices.

Pearson's correlation is computed and values are reported in Table 2. However, three of the twenty years analyzed have observed values for only two validation points instead of the three we have chosen. In these cases, Pearson's correlation is equal to one and we discarded these points from the average correlation value computation. Using only three points for validation may underestimate the true quality of the interpolations, but this choice was required due to the low density of available station data.

**Table 2.** Normalized RMSE and Pearson’s correlation average over 1991-2010. Correlation average is computed only over years where all three validation points are available.

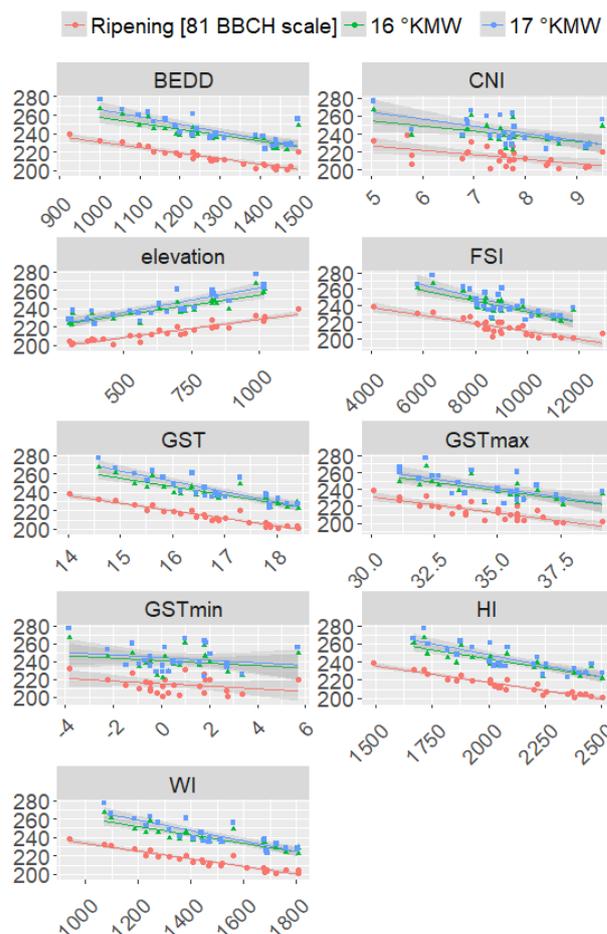
Index	Method	Normalized RMSE	Pearson’s correlation
BEDD	MLR	38.81	0.98
	RK	39.47	0.98
	SVR	124	-0.96
CNI	MLR	60.1	0.82
	RK	50.78	0.84
GST	SVR	345.79	-0.76
	MLR	32.07	0.98
	RK	31.68	0.98
HI	SVR	133.3	-0.95
	MLR	39.26	0.94
	RK	39.25	0.94
WI	SVR	159.84	-0.93
	MLR	35.36	0.98
	RK	35.57	0.98
	SVR	191.7	-0.94

### 3.2 Phenology analysis

To be able to predict the suitability of an area for viticulture, it is necessary to know the relationship between climate and the phenological and chemical characteristics of the vines and grapes. This relationship is quantified via a linear regression between climate indices, calculated from the gap-filled time series of the “REBECKA” stations, and measured phenological and chemical properties of the grapes. The sampling of grapes for the assessment of chemical characteristics is done in the same plots where the “REBECKA” stations are located. These values are used to evaluate different climate indices in predicting the day of ripening for vines. Once the relation for the most accurate index is known, it is possible to spatialize the estimated ripening date using only an interpolated index map. Knowledge about the estimated timing of ripening can then be used in assessing the suitability of an area for viticulture.

The climate indices evaluated here are temperature-sum indices (BEDD, HI, WI) and average temperature (GST, GSTmin, GSTmax). Since they are computed from April 1<sup>st</sup>, we decided to focus our attention on end-of-season phenological stages, in particular ripening (defined as level 81 in BBCH scale), and when must weight reaches 16-17 degrees on the KMW scale.

From Figure 2 it is possible to recognize trends: higher temperatures, or temperature-sums, are linked to earlier phenological developments. We have also tried to fit linear models to the distributions. As indicated in Table 3, a statistically significant linear relationship ( $p$ -value < 0.001) exists for almost all of the indices. Looking at  $r^2$  values, models with WI and GST are equally effective in explaining most of the variability for



**Figure 2.** Correlation between day-of-year of phenological stages versus climatic indices and elevation. The lines represent a linear relationship between points and grey area is the 0.95 confidence interval.

all phenology stages, but other indices perform worse. It must also be noted that linear regression explains ripening far better than the KMW stages.

### 4 Conclusions

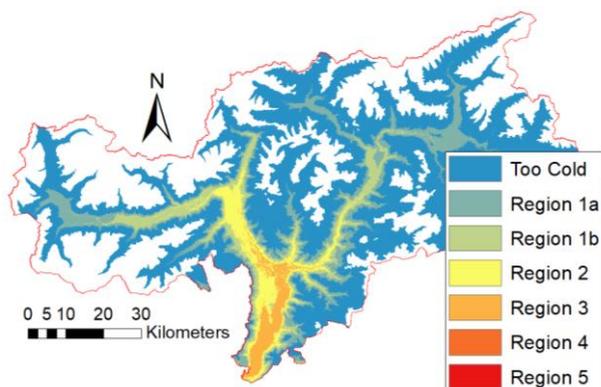
From the interpolation statistics we can state that MLR can be used to create high-resolution maps for climate indices with a significant level of accuracy. Despite RK and MLR performing similarly, the decision to select the MLR method is supported by two facts. First, computational time for a single map using MLR is four times faster than RK. Second, RK requires the model to fit over a variogram and automating this process does not always leads to satisfying results because our script can’t detect all the possible variogram patterns and fits the model accordingly. The average map with interpolated Winkler zoning over 1991-2010 is shown in Figure 3.

From the phenology analysis, we have found that WI and GST are good indicators for estimating later phenological stages. It is now possible to apply the linear model of phenological stages to the interpolated average maps in order to obtain interpolated phenology maps. For example, Figure 4 represents the map of estimated DOY for ripening, obtained from 1991-2010 Winkler index average.

**Table 3.** Linear model estimation between phenology stages and climatic indices. Only statistically significant models (p-value < 0.001) are reported.

Phenology stages	Index	Adj. r <sup>2</sup>	Residual std. dev.	p-value
Ripening	WI	0.91	3.09	9.39E-15
	GST	0.91	3.15	1.50E-14
	HI	0.89	3.37	8.19E-14
	elevation	0.88	3.55	3.05E-13
	BEDD	0.8	4.6	2.09E-10
	FSI	0.71	5.49	1.79E-08
	GSTmax	0.66	5.99	1.65E-07
16 °KMW	WI	0.76	5.98	1.64E-07
	GST	0.76	5.98	1.64E-07
	elevation	0.73	6.34	5.13E-07
	HI	0.71	6.56	9.97E-07
	BEDD	0.62	7.47	1.22E-05
	FSI	0.61	7.57	1.58E-05
17 °KMW	GSTmax	0.47	8.85	3.47E-04
	WI	0.74	7.31	2.15E-08
	GST	0.74	7.35	2.48E-08
	elevation	0.71	7.72	7.75E-08
	HI	0.67	8.26	3.75E-07
	BEDD	0.61	8.92	2.27E-06
	FSI	0.51	10.02	3.54E-05

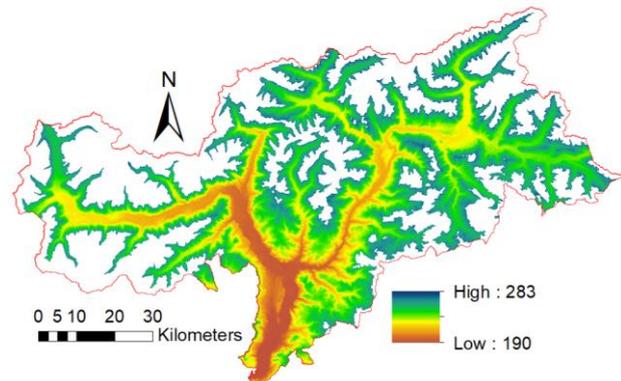
The “REBECKA” project will collect data until the end of 2019, providing two full years worth of data beginning January 1<sup>st</sup>. Further efforts have already begun to move the interpolation from a yearly basis (vegetation-season) to a daily basis. This will allow us to better evaluate the earlier phenological stages using specific phenology models: models that are usually based on daily temperature sums.



**Figure 3.** Winkler index average of years 1991-2010

In addition, the effects of climate change may play an important role in the upwards or downwards shift of the climatic niches for specific varieties [1]. This includes the risk, in the long-term, that suitable areas currently inside DOC zones in South Tyrol, will shift outside of them. Comparing the climate zoning computed over 1991-2010

with the average over a suitable climatic period (at least 10 years, e.g. 2011-2020) will provide information about a possible shift of the climatic niches suitable for viticulture.



**Figure 4.** Estimated ripening day-of-year using a linear model with Winkler index as predictor.

## References

- G. V. Jones, “Climate, Grapes, And Wine: Structure And Suitability In A Changing Climate,” *Acta Horticulturae* (931) (2012): 19–28, doi:10.17660/ActaHortic.2012.931.1.
- C. Frei, “Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances,” *International Journal of Climatology* 34(5) (2014): 1585–1605, doi:10.1002/joc.3786.
- Zentralanstalt für Meteorologie und Geodynamik, Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto, and Institut für Interdisziplinäre Gebirgsforschung, *Il Clima del Tirolo - AltoAdige - Bellunese* (2015).
- E. Eccel, E. Cordano, and G. Toller, *ClimClass: Climate Classification according to Several Indices* (2016).
- World Meteorological Organization, *Guidelines on the Calculation of Climate Normals* (Geneva, Switzerland: World Meteorological Organization, 2017).
- M. H. Kutner, ed., *Applied linear statistical models*, 5th ed. (Boston: McGraw-Hill Irwin, 2005).
- G. James et al., eds., *An introduction to statistical learning: with applications in R* (New York: Springer, 2013).
- T. Hengl et al., *A practical guide to geostatistical mapping of environmental variables* (Luxembourg: Publications Office, 2007).
- T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter, “About regression-kriging: From equations to case studies,” *Computers & Geosciences* 33(10) (2007): 1301–1315, doi:10.1016/j.cageo.2007.05.001.
- S. R. Gunn, *Support Vector Machines for Classification and Regression*, 1998.
- A. J. Miller, *Subset selection in regression*, 2nd ed. (Boca Raton: Chapman & Hall/CRC, 2002).
- M. H. Nguyen and F. de la Torre, “Optimal feature selection for support vector machines,” *Pattern Recognition* 43(3) (2010): 584–591, doi:10.1016/j.patcog.2009.09.003.
- R. Le Roux et al., “Comparison of two fine scale spatial models for mapping temperatures inside winegrowing areas,” *Agricultural and Forest Meteorology* 247 (2017): 159–169, doi:10.1016/j.agrformet.2017.07.020.
- D. H. Lorenz et al., “Growth Stages of the Grapevine: Phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. *vinifera*)—Codes and descriptions according to the extended BBCH scale†,” *Australian Journal of Grape and Wine Research* 1(2) (1995): 100–103, doi:10.1111/j.1755-0238.1995.tb00085.x.