ESTIMATION OF DEGREE BRIX IN GRAPES BY PROXIMAL HYPERSPECTRAL SENSING AND NANOSATELLITE IMAGERY THROUGH THE RANDOM FOREST REGRESSOR

Authors: Diniz Carvalho de ARRUDA, Jorge Ricardo DUCATI*

Remote Sensing Center, Universidade Federal do Rio Grande do Sul, Av. Bento Goncalves 9500, CEP 91501-970, Porto Alegre RS, Brazil

*Corresponding author: *ducati@if.ufrgs.br*

Abstract:

Context and purpose of the study

The assessment of physiological parameters in vineyards can be done by direct measurements or by remote, indirect methods. The latter option frequently yields useful data, and development of methods and techniques that make them possible is worthwhile. One of the parameters most looked for to define the quality status of a vineyard is the degree Brix of its grapes, a quantity usually determined by direct measurement. However, other ways may be possible, and presently Brix estimations in vineyards using as data sources field radiometry, localized Brix measurements and satellite imagery are reported.

Material and methods

The investigation was developed in a commercial vineyard in south Brazil at two stages of the 2017/2018 vegetative cycle. Brix degree was measured twice: using a spectroradiometer which measured reflectance from 350nm to 2500nm, and a refractometer. Brix estimates were derived using a machine learning model, the Random Forest Regression (RFR) algorithm, applied on data from images of PlanetScope satellites.

Results

Results produced coefficients of correlation between observed and predicted degrees Brix as high as 0.89. Analysis of an importance parameter, the Gini index, suggested that spectral data at ultraviolet, visible, and near-infrared wavelengths and the vegetation indices TGI and NDVI are the most important variables used for the predictive model. This methodology is potentially useful for the derivation of vineyard quality parameters at situations when specific vineyard conditions, as rugged terrain and large variations in soils, turn direct measurements a difficult task.

Keywords: degree Brix, hyperspectral data, Random Forest Regression

1. Introduction

Grape producers are in constant search for fruit excellency, and factors that have to be in close monitoring include soil fertility, vine susceptibility to pathogen attacks, and effects of humidity, temperature, and exposure to sunlight. These procedures are closely related to precision viticulture, and one of the techniques employed in such approaches are those of remote sensing, supporting the decision-making of producers with the help of remotely acquired data carrying information about the vineyard. Medium and high-resolution images are used to delineate vineyards, and by spectral indices it is possible to detect stresses related to water status, diseases, and soil characteristics (Martín et al. 2007; Power et al. 2019).

One of the most searched parameters related to fruit quality is the berry sugar content, which is expressed by the Brix degree (Brix), which is commonly estimated at the vineyard by devices called refractometers, a direct measurement performed on berries (Ferrer et al. 2019); however, alternative techniques may eventually prove as more practical, for example in occasions where Brix have to be known in remote vineyards (Bonilla et al. 2015; Poblete-Echeverría et al. 2017), or in very large surfaces, especially if in this last case the vineyard extends over areas with rugged or inhomogeneous soils, where remote sensing techniques may be useful (Hall et al. 2003; Fredes et al. 2021). Such indirect determination or prediction of needed parameters and its validation by comparison with direct measurements can be done by machine learning models like the Random Forest Regressor (RFR), a model tolerant to noisy data which evaluates correlations between variables using a random vector (Breiman 2001). The RFR performance is high in setting spectral reflectance measurements, because of its low sensitivity to outliers (Chea et al. 2022). Hence, the aim of this study is to estimate grape Brix degree in a commercial vineyard by hyperspectral proximal sensing and nanosatellite imagery by applying an RFR machine learning model, using as supporting data direct Brix measurements.

2. Material and Methods

Study area and data

The study area was the Luiz Argenta Winery, a producer of fine wines located at the northeastern part of State of Rio Grande do Sul, Brazil. Six Cabernet Sauvignon plots were selected with espalier trellis, rootstocks Paulsen 1103, conventional viticultural treatments, row spacing of 2.8 meters, with 1.45 meters between plants. At each plot four plants located at central rows were selected and marked. The experiment was conducted in two stages of the grape ripening cycle.

Input data sets

Field spectroradiometric measurements were performed in December 15, 2017 and February 27, 2018, dates that coincide respectively with the beginning of the veraison and with the phenolic ripeness. The equipment used was a FieldSpec® 3 spectroradiometer from ASD – Analytical Spectral Devices, which performs reflectance measurements in the 350nm to 2500nm domain, covering the ultraviolet (UV) and the near and mid-infrared (NIR, SWIR1 and 2). Measurements with the Leaf Clip accessory were taken in four leaves per plant, in four points for each leaf, on the adaxial part of the selected leaves, opposite to the clusters under full light, in the middle third of the canopy, on branches closest to the main trunk, for four plants per plot. Calibrations were performed and calculations were done to produce a single, average spectrum per plant. At the end of each observing run 24 reflectance spectra were available for analysis. Corrections for filtering noisy data and reflective plateau shifts were performed using Savitzky-Golay filters, Jump Correction and Normalize function.

Brix measurements with a refractometer were made at February 8, 2018 at temperatures between 23°C to 30°C. Four grape clusters were chosen at lower parts of the selected plants, and at each cluster the Brix was measured at four berries at its lower parts. The resulting data set had twenty-four Brix values.

For satellite data, two images from the Dove-R sensor aboard PlanetScope nanosatellites were acquired at December 16, 2017 and February 28, 2018. The product level 3B, Surface Reflectance (SR), provides images with radiometric and atmospheric corrections, projected to plane coordinates, with spatial resolution of about three meters. Four bands were available: blue (465nm to 515nm), green (547nm to 583nm), red (650nm to 680nm), and NIR (845nm to 885 nm. After the treatment of the images, a pixel sampling procedure was done for the creation of the grape Brix prediction models. The selection of pixel samples in the images happened with the creation of a buffer with an area of 9 m² around the midway point of the plants selected for this study.

Vegetation indices were entered as input data into the estimation models, being calculated from the measured reflectance spectra: Normalized Difference Vegetation Index (NDVI), Triangular Greenness Index (TGI) and Visible Atmospherically Resistant Index (VARI).

Random Forest Regression

For prediction the non-parametric algorithm Random Forest Regressor (RFR) was used. The main parameters of the model are estimators as tree depth, scores per bagging, and Out of Bagging (OOB). The selection of the input features by the RFR algorithm occurs by applying the importance criterion for the variables, the Gini Index, a multivariate process to the variable that best divides the nodes randomly.

From the measured Brix values a table was compiled with the descriptive statistics for each of the six vine plots, including the maxima, average and minima values, besides coefficients of variation. The predicted Brix was estimated both from leaf hyperspectral reflectance spectra and from PlanetScope imagery, using the Random Forest Regression algorithm; the data base formed by the average values for each plot was used for two input groups: a) Brix *x* Hyperspectral Reflectance (HR input dataset) and b) Brix *x* Surface Reflectance from PlanetScope and Vegetation Indices (SR set). For RFR parameters n_estimators = [50,100,250,500] were used; these values express the maximal number of trees used at models to predict the Brix values from each sensor. Several estimators were tested to assess the performance stability of the model predictions, expressed by the following metrics: Mean Square Error (MSE); Root Mean Square Error (RMSE); Coefficient of Determination (R^2); Adjusted R^2 (adj_R^2); and Out of Bagging (OOB). After metrics analysis and comparison, the RFR models with 500 decision trees were chosen for spacialization using satellite data and the vegetation indices (NDVI, TGI and VARI) for the two acquisition dates, and finally maps for predicted Brix were produced.

3. Results and Discussion

Measurements of grape Brix (Figure 1) showed a coefficient of variation of about 5%-6%, except for plot 16a which presented the largest variation, around 9%. The sugar content showed similar values for all plots, an expected behavior since a single variety was studied; the highest coefficient of variation which was found for plot 16a, was probably due to end-of-cycle diseases observed in some plants, bunches, and leaves. Peak Brix values were found at 19a plot; this parcel was the one with the smallest density of plants and canopies, presenting smaller bunches, being on stony soil with exposed surface. These high variations in Brix can be understood by the fact that at this stage of the grape ripening cycle it is normal for the leaves to present distinct chlorophyll values.

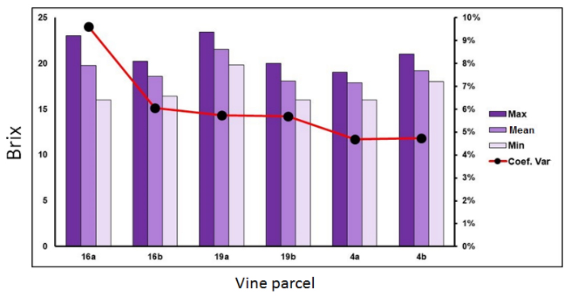

Figure 1: Descriptive statistics for grape Brix readings, refractometer data measured *in situ*.

The prediction metrics of the models from both HR and SR input data sets are presented in Table 1, with number of estimators from 50 to 500. Predictions with the proximal sensor input data (HR) for the first date (1st) presented R^2 values above 0.86 independently of the number of estimators, and for the second date (2nd) R^2 values were between 0.82 and 0.85. Still for HR, the OOB metric shows, with 100 estimators, a value of 0.03 for 1st and 0.19 for 2nd with 500 estimators, which were the best scores among all models for input from leaf hyperspectral reflectance. Keeping with Table 1, the results of the metrics for the input data from surface reflectance (SR) show that in both collection dates R^2 values were in the range between 0.84 and 0.87. The model with 500 estimators obtained the lowest OBB scores, with values of 0.01(1st) and 0.04 (2nd).

Table 1. Metrics for evaluating grape Brix estimates for hyperspectral data (Hyperspectral Reflectance, HR) and PlanetScope (Surface Reflectance and Indices, SR). Mean Square Error (MSE); Root Mean Square Error (RMSE); Coefficient of Determination (R^2); Adjusted R^2 (adj_R^2); Out of Bagging (OOB).

Parameters	Estimators	MSE		RMSE		R^2		adj_R²		OOB	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
HR	50	0.28	0.32	0.53	0.57	0.86	0.84	1,00	1,00	0.11	0.28
	100	0.27	0.36	0.52	0.6	0.87	0.82	1,00	1,00	0.03	0.2
	250	0.25	0.33	0.5	0.57	0.87	0.84	1,00	1,00	0.05	0.3
	500	0.26	0.3	0.51	0.55	0.87	0.85	1,00	1,00	0.08	0.19
SR	50	0.29	0.26	0.54	0.51	0.86	0.87	0.79	0.87	0.06	0.06
	100	0.21	0.25	0.46	0.5	0.89	0.88	0.85	0.82	0.14	0.05
	250	0.24	0.28	0.49	0.52	0.88	0.86	0.83	0.81	0.08	0.06
	500	0.27	0.26	0.52	0.51	0.86	0.87	0.81	0.82	0.01	0.04

The importance of the bands of each sensor in the estimation was expressed by the Gini index. For hyperspectral data the most important bands on the first date were concentrated in the spectral region of the ultraviolet, red, at near-infrared and at some areas in the SWIR2. In the second collection the model selected wavelengths of the ultraviolet, blue, and SWIR1. For satellite data plus vegetation indices the Gini index was also used as indicator of importance, and indices VARI, NDVI, and TGI showed their importance to the spectral modeling. The NDVI showed relevance in both acquisition dates. TGI Index was only relevant in the first acquisition date and band B3 in the second collection. TGI is considered as a valuable indicator at more advanced phenological stages, at the onset of changes in leaf color and decrease of photosynthetic.

Figure 2 presents the Brix estimate for the six vine plots. As the cycle progresses and grape ripeness approaches, it is possible to identify an increase in Brix (hatchings in dark violet shades). From the models it is suggested that rightness rate is high; even at coarser satellite resolutions it seems to be possible to assess grape quality, as it is suggested by Fredes et al. (2021).

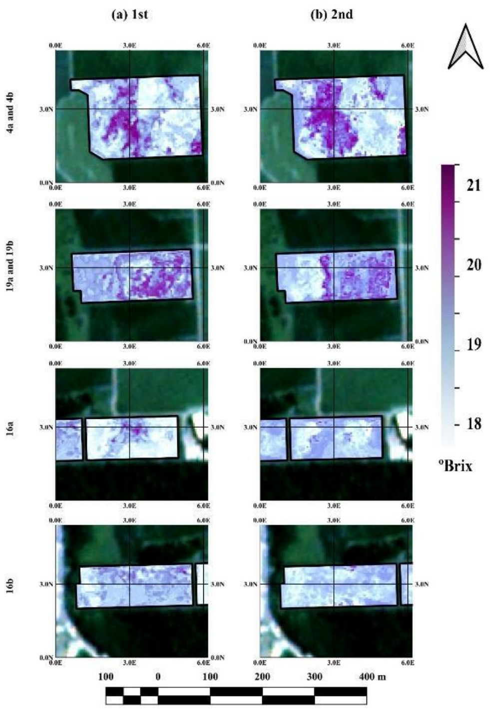

Figure 2: Brix estimates to the six studied vineyards. (a) December 2017; (b) February 2018.

4. Conclusions

The results suggest that the derivations made using the Random Forest Regressor were significant, either based in proximal or satellite data; this perception is supported by analysis of the metrics evaluating the predictions for both dates. The measurements of Brix are linked to the vegetative state of the plant and the exchange processes between plant and its environment, involving an analysis more complex than only considering light reflection by leaves. This is important to set up a better understanding between spatial factors and the development and monitoring of vine vegetative vigor.

The importance and utility of high-resolution sensors like Dove-R was highlighted, as well as that using nanosatellite and proximal derived products will help address the impacts of management, weather, soil, and climate on vineyard vegetative growth and consequently the quality of its grapes. For future studies, the development of a sampling grid with a higher quantity of plants, plots, and varieties will help to understand the identity of the spectral signatures of the vines and to seek new relationships with other grape quality and quantity parameters.

5. Literature cited

BONILLA I., de TODA F. M., MARTÍNEZ-CASASNOVAS J. A. 2015. Vine vigor, yield and grape quality assessment by airborne remote sensing over three years: Analysis of unexpected relationships in cv. Tempranillo. Spanish Journal of Agricultural Research, 13(2), e0903.

BREIMAN, L. 2001. Random forests. Machine Learning, 45(1), 5–32.

CHEA C., SAENGPRACHATANARUG K., POSOM J., SAIKAEW K., WONGPHATI M., TAIRA, E. 2022. Optimal models under multiple resource types for brix content prediction in sugarcane fields using machine learning. Remote Sens. Appl. Soc. Environ. 26, 100718.

FERRER M., ECHEVERRÍA G., PEREYRA G., GONZALEZ-NEVES G., PAN D., MIRÁS-AVALOS J. M. 2019. Mapping vineyard vigor using airborne remote sensing: relations with yield, berry composition and sanitary status under humid climate conditions. Precision Agriculture 21(1), 178–197.

FREDES S.N., RUIZ L.A., RECIO J.A. 2021. Modeling °Brix and pH in wine grapes from satellite images in Colchagua Valley, Chile. Agriculture, 11(8), 697.

HALL A., LOUIS J., LAMB D. 2003. Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images. Computers and Geosciences, 29(7), 813–822.

MARTÍN P., ZARCO-TEJADA P. J., GONZÁLEZ M. R., BERJÓN A. 2007. Using hyperspectral remote sensing to map grape quality in “Tempranillo” vineyards affected by iron deficiency chlorosis. Vitis - Journal of Grapevine Research, 46(1), 7–14.

POBLETE-ECHEVERRÍA C., OLMEDO G. F., INGRAM B., BARDEEN M. 2017. Remote sensing detection and segmentation of vine canopy in ultra-high spatial resolution RGB imagery obtained from unmanned aerial vehicle (UAV): A case study in a commercial vineyard. Remote Sensing 2017, 9(3), 268.

POWER A., TRUONG V. K., CHAPMAN J., COZZOLINO D. 2019. From the laboratory to the vineyard-Evolution of the measurement of grape composition using NIR spectroscopy towards high-throughput analysis. High-Throughput 2019, *8*(4).