



Proximal hyperspectral analysis in grape leaves for region and variety identification

Diniz Carvalho de Arruda^{1*}  Jorge Ricardo Ducati¹  Rosemary Hoff² 
Tássia Fraga Belloli¹  Adriane Brill Thum³ 

¹Universidade Federal do Rio Grande do Sul (UFRGS), Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, 91501-970, Porto Alegre, RS, Brasil. E-mail: dinizcarvalho28@outlook.com. *Corresponding author.

²Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Bento Gonçalves, RS, Brasil.

³Escola Politécnica, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brasil.

ABSTRACT: Reflectance measurements of plants of the same species can produce sets of data with differences between spectra, due to factors that can be external to the plant, like the environment where the plant grows, and to internal factors, for measurements of different varieties. This paper reports results of the analysis of radiometric measurements performed on leaves of vines of several grape varieties and on several sites. The objective of the research was, after the application of techniques of dimensionality reduction for the definition of the most relevant wavelengths, to evaluate four machine learning models applied to the observational sample aiming to discriminate classes of region and variety in vineyards. The tested machine learning classification models were Canonical Discrimination Analysis (CDA), Light Gradient Boosting Machine (LGBM), Random Forest (RF), and Support Vector Machine (SVM). From the results, we reported that the LGBM model obtained better accuracy in spectral discrimination by region, with a value the 0.93, followed by the RF model. Regarding the discrimination between grape varieties, these two models also achieved better results, with accuracies of 0.88 and 0.89. The wavelengths more relevant for discrimination were at ultraviolet, followed by those at blue and green spectral regions. This research pointed toward the importance of defining the wavelengths more relevant to the characterization of the reflectance spectra of leaves of grape varieties and revealed the effective capability of discriminating vineyards by their region or grape variety, using machine learning models.

Key words: vineyards, hyperspectral, spectroradiometer, machine learning.

Análise hiperespectral proximal em folhas de videiras para identificação de regiões e variedades

RESUMO: Medições de refletância de plantas da mesma espécie podem produzir conjuntos de dados com diferenças entre os espectros, devido a fatores que podem ser externos à planta, como o ambiente onde a planta cresce, e fatores internos, para medições com variedades de plantas. Este artigo reporta resultados da análise de medições por espectrorradiometria efetuadas em folhas de vinhas de variedades e em diferentes localidades. O objetivo desta pesquisa foi, após a aplicação de técnicas de redução de dimensionalidade para a definição dos comprimentos de onda mais relevantes, avaliar quatro modelos de aprendizado de máquina aplicados à amostra observacional visando discriminar classes de região e variedade. Os modelos de classificação de aprendizado de máquina testados foram Canonical Discrimination Analysis (CDA), Light Gradient Boosting Machine (LGBM), Random Forest (RF) e Support Vector Machine (SVM). A partir dos resultados, relatamos que o modelo LGBM obteve melhor acurácia na discriminação espectral por região, com valor de 0,93, seguido pelo modelo RF. Relativamente à discriminação entre castas, estes dois modelos também obtiveram melhores resultados, com acurácias de 0,88 e 0,89. Os comprimentos de onda mais importantes para as discriminações procuradas estiveram na região do ultravioleta, seguidos do azul e do verde. Este trabalho aponta para a importância de detectar os comprimentos de onda mais relevantes para a caracterização dos espectros de refletância das folhas de variedades de vinhas, e revela a capacidade efetiva de discriminar vinhedos por suas regiões ou variedades, usando modelos de aprendizado de máquina.

Palavras-chave: Vinhedos, hiperespectral, aprendizagem de máquina.

INTRODUCTION

The spectral response of vegetation expressed by its reflectance has been known to be a way to characterize different vegetal species, with applications in surveys and monitoring of forests, crops and other land uses (ZHANG et al., 2014; MIRZAEI et al., 2019).

Several studies have applied techniques of remote sensing for data acquisition, including satellite or aerial imagery and/or field or laboratory spectroradiometer. In the first cases, the spectral resolution, in general, tends to be moderate, and only the main spectral features are acquired; even with this limitation, classifications with significant accuracies have been

accomplished in studies on vineyards (KARAKIZI et al., 2016; MOGHIMI et al., 2020; SILVA & DUCATI, 2009) using conventional classification algorithms. In the latter cases, using a spectroradiometer extremely high spectral resolution can be attained, showing minute details of a spectrum, and allowing to detect subtle spectral features of vine leaves; these features express degrees or states of pigmentation, cell structure, and water content which, besides depending on intrinsic biological descriptors, can be influenced by environmental and geographical factors (CEROVIC et al., 2012; SMIT et al., 2010; THUM et al., 2020).

From this perspective, spectral data is valuable in studies focused on vine development in geographical contexts, since the high density of information carried by a high-resolution spectrum allows searching for differentiation between cultivars and from external influences caused by climate, soil, management, or other effects. Results from such studies are helpful to the characterization of viticultural regions aiming to distinguish themselves from other regions, contributing to the formation of a set of descriptors necessary to the attribution of a label of typicity of which AOC (Appellation d'Origine Contrôlée), IGT (Indicazione Geografica Tipica) or AVA (American Viticultural Area) are examples. Such characterizations, when coming from data of plant spectroscopy, have been achieved mainly using conventional classification algorithms (SILVA & DUCATI, 2009; KARAKIZI et al., 2016), but few results have been reported of applications of Machine Learning models which, with present computational resources, can outperform already existent classification methods (ANGUITA et al., 2010).

This paper reports the results from spectroradiometric field measurements performed on vineyards located in southern Brazil, where we investigated their potential to discriminate vines by their locations or by variety. Here, the location factor is dominated by environmental constraints (soils, climate), while the variety factor tends to be dominated by biological (genetic characteristics) constraints. Both factors have significant impacts on plant metabolism and development (WHITE, 2009), influencing leaf structure and chemical composition and, therefore, its reflectance spectrum (THUM et al., 2020). Specifically, the objectives of this research were: a) To discriminate vineyards by region and variety from leaf reflectance data; b) To select a technique to reduce the number of wavelengths necessary for the first objective; c) To select, from a selected set of Machine Learning techniques, the ones with the best performances in the classification process.

MATERIALS AND METHODS

Study area

As study areas, eight vineyards were selected in Rio Grande do Sul, which is the southernmost state in Brazil. These vineyards are distributed over a territory of about 500 km wide, on terrains of different types of rocks, and belong to the following wineries: a) Almadén Estate (W1) in Santana do Livramento, in the Campanha Gaúcha wine region, with sandstone-based soils from the Guará Formation (WILDNER et al. 2008); b) Boscato Winery in Nova Pádua, with two vineyards (W2 and W3, two kilometers apart) on acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas Formation (IBGE, 2018, ROSSETI et al. 2017); c) Chandon Estate (W4) in Encruzilhada do Sul, on the gneiss of the Arroio dos Ratos Gneissic Complex (WILDNER et al. 2008); d) Luiz Argenta Estate (W5) in Flores da Cunha, over acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas Formation (IBGE, 2018, ROSSETI et al. 2017); e) Miolo Winery in Bento Gonçalves (W6) in the Serra Gaúcha wine region, with soil on acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas Formation (IBGE, 2018, ROSSETI et al. 2017); f) Miolo Seival Estate (W7) in Candiota, in the wine-growing region of Campanha Gaúcha, whose soils are a transition between sandstone and claystone of the Rio Bonito and Palermo Formations (CAMOZZATO & LOPES, 2012); g) Terra Sul Winery (W8) in Pinheiro Machado, in the Serra do Sudeste wine region, with soils based on granitic rocks from the Pinheiro Machado Granitic-Gneissic Complex (WILDNER et al. 2008). From this description, it can be seen that the studied vineyards are over different soils, with varying amounts of sand, clay and organic matter. The balance of these soil components, meaning the variation in mineral content, play an important role in reflectance spectra, not only on the spectra of soils themselves (DEMATTÊ, 2002), but also on the spectra of vegetation growing on it (THUM et al. 2020), since many elements are important to plant metabolism; for example, CONRADIE (1981), SCHREINER et al (2006) and SCHREINER (2016) reported as elements like phosphorus potassium, calcium and magnesium move along vine tissues. It is known that different soils have different mineral availability to plant metabolism (WHITE, 2009), with an impact on leaf reflectance spectra (THUM et al. 2020). We note for the regions presently under study that iron availability (associated with clay content) changes greatly, possibly leading to significant changes on plant

reflectance spectra. As additional information, we briefly discuss the reason of dividing Boscato Estate in two parts (W2 and W3). From a previous investigation of this winery (THUM et al., 2020), it was reported that W2 (5.38 hectares) has elevations from 666 to 688m, and W3 (7.93 hectares) has elevations from 747 to 785m; in addition to the fact of W3 is at higher elevations, W3 displays steeper slopes. Furthermore, out of 21 measured agronomical parameters (data not presently shown), only 3 (P, Ca, Zn) had larger variability in W2; W2 is; therefore, much more homogeneous. Finally, measured soil profiles in W2 are deeper across that vineyard, what points for a possible reason of the larger variability of soil traits in W3, since shallower soils in a more rugged terrain would tend to put the surface in closer contact with deeper horizons and the bedrock, these two layers acting as mineral suppliers. This condition of soil diversity in terrains seating on the same bedrock provides an opportunity for assessing the limits of classification performances of the set of Machine Learning techniques to be presently tested. We also noted that estates W1 and W7 are located at areas covered by the “Campanha Gaúcha” viticultural region; W2, W3 and W5 are in the “Altos Montes” viticultural region; W4 and W8 are at the “Serra do Sudeste” viticultural region; and W6 is at the “Vale dos Vinhedos” viticultural region.

The distribution of these locations over the State’s territory is shown in figure 1.

As grape varieties or cultivars we selected twelve of those more commonly found in the chosen regions, which are: Cabernet Sauvignon (V1), Chardonnay (V2), Merlot (V3), Petit Verdot (V4), Pinot Grigio (V5), Pinot Noir (V6), Riesling Italic (V7) (also known as Welschriesling), Sauvignon Blanc (V8), Syrah (V9), Tannat (V10), Tempranillo (V11), and Viognier (V12). These twelve grape varieties are not present in all eight locations; for example, the Chandon Estate only has Pinot Noir, Chardonnay and Riesling Italic, and at Boscato only Cabernet Sauvignon and Merlot were measured. Detailed information on number of measurements is provided in table 1. The climate in all regions is subtropical with well-defined seasons; however, the Serra Gaúcha region tends to have summers with higher humidity. We visited in total seventy-eight vine parcels.

Leaf reflectance acquisition

Field spectroscopic measurements were performed with a Malvern Panalytical Spectral Devices (ASD, Westborough, MA, USA) FieldSpec® 3 spectroradiometer, which has spectral sensitivity between 350nm and 2500nm, using the Leaf Clip

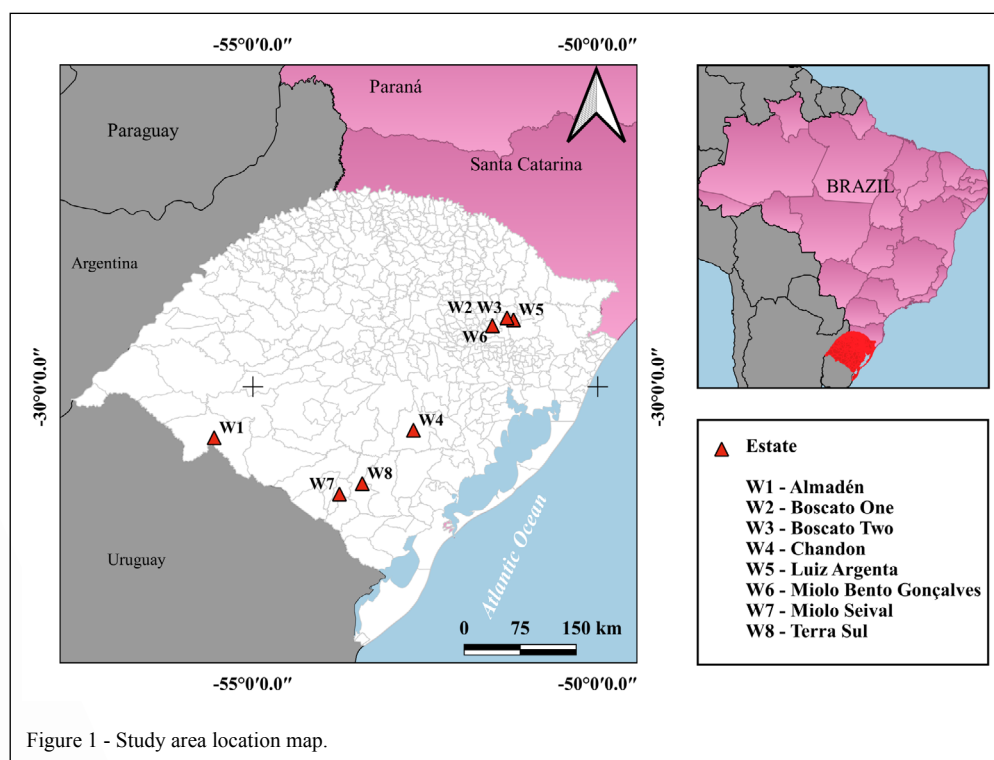


Figure 1 - Study area location map.

Table 1 - The number of measurements performed in the adaxial part of the leaves, *in situ* / *in vivo*, for each corresponding class.

Variety/ Estate	W1- Almadén	W2- Boscato One	W3- Boscato Two	W4- Chandon	W5- Luiz Argenta	W6- Miolo BG	W7-Miolo Seival	W8 - Terra Sul
V1- Cabernet Sauvignon	0	237	143	0	223	159	0	105
V2- Chardonnay	132	0	0	93	60	0	100	30
V3-Merlot	144	96	338	0	60	0	97	80
V4-Petit Verdot	0	0	0	0	0	0	0	30
V5- Pinot Grigio	0	0	0	0	0	0	0	30
V6- Pinot Noir	0	0	0	96	121	0	100	46
V7-Riesling Italic	132	0	0	48	61	0	0	0
V8 - Sauvignon Blanc	0	0	0	0	0	0	0	20
V9 - Syrah	0	0	0	0	121	0	0	0
V10 - Tannat	0	0	0	0	0	0	0	56
V11 - Tempranillo	0	0	0	0	0	0	0	30
V12 - Viognier	0	0	0	0	0	0	0	30

sensor. Field trips were performed in December 2018 and January 2019, since these dates correspond to a period in the phenological cycle where grape leaves are already well-developed, in the stage of growth and ripening of berries represented on the BBCH scale in the sub-stages 81 to 83 (LORENZ et al., 1995).

In each estate, we selected vine parcels with areas of about five hectares. At each parcel we chose rows centrally localized, at each row we selected four plants, and at each plant we measured four fully developed leaves at their adaxial sides. Calibration of the sensor, through optimization and measurement of the white reference plate of the Leaf Clip probe, was conducted before making the spectroradiometric readings. Every spectrum was recorded at one-nanometer intervals, resulting in 2151 reflectance values for the observed spectral domain (350 nm to 2500nm). The final sample had 3006 spectra corresponding to measurements of 1002 leaves (three spectra per leaf); however, the measurements used for the analyses were 2967 in total since 39 spectra were detected as being erroneous for several factors and were excluded.

Pre-processing of spectra

To mitigate the noise interference in the spectra, and to smooth the spectral breaks at the sensor's interfaces, we used the Savitzky-Golay filter and slice correction. The library packages used were SciPy, signal Filter, and Coefficients (VIRTANEN et al., 2020). Since high-resolution spectra tend to carry redundant information over neighboring wavelengths, a feature that tends to increase processing time of

classification tasks with no sizable gains, the next step was to decrease the number of wavelengths by means of two spectral reduction techniques applied to the database, which were: Spectrum Ratio (SR) and Kernel Principal Component Analysis (KPCA).

Spectrum Ratio (SR)

The SR technique was applied after a normalization procedure was performed on each original spectrum. Since in each acquisition the sensor can receive a particular influx of energy, recorded levels of reflectance can vary from one spectrum to another; that is, each spectrum comes from the acquisition of a certain amount of energy across the observed wavelength domain, implying in a specific area under the spectral curve. The SR technique consists in the direct comparison of two spectra at the same scale, and so, original spectra were transformed through a normalization procedure described elsewhere (PITHAN et al., 2021); we note that normalization is an operation that does not change the shape of any spectrum.

The Estates group had eight vineyards, so comparisons between them, by pairs, allowed twenty-eight combinations; for each estate, a mean spectrum was derived from all measurements, and this spectrum was divided by the mean spectrum of each other estate, an operation that, applied to all eight vineyards, resulted in twenty-eight "spectrum-ratios." The same procedure was followed for the Varieties group where, for twelve varieties, we obtained sixty-six possible "spectrum ratios". A typical "spectrum-ratio" has values around unity for

all wavelengths, except at those wavelengths where spectral differences between classes (in Estates or in Varieties) exist. In this sense, the technique reveals where differences between classes exist, knowledge to be used in classification tasks.

The spectra were subjected to non-parametric correlations tests for the whole spectral domain. First, a correlational test, the Spearman rank correlation model, was used to evaluate collinearity between the 2151 wavelengths. The coefficient of determination (R^2) was used to adjust the correlations for each wavelength. Wavelengths having statistical significance expressed by a p -value < 0.05 were selected. Additionally, and to address the level of statistical significance, the Kruskal-Wallis H test was used to assess the real differences between the sample groups. Levels of statistical significance, α (0.05), were determined to verify the difference in statistical distributions of the sub-groups internal to each main group (Estates and Varieties).

Kernel principal component analysis (KPCA)

KPCA, the second spectral dimension reduction technique, is a technique for transforming original data into components of uncorrelated variables, using Principal Component Analysis with extension Kernel in dimensionality reduction to create reliable compositions, since the determination of decision limits between classes is performed in a non-linear way (FAUVEL et al., 2009).

Hyperspectral classification

The classification of reflectance spectra was performed from both input techniques, KPCA and SR. Four Machine Learning (ML) algorithms were used in processes, developed in Python language using the Scikit-Learn package and using the libraries Pandas and NumPy for the preparation of matrix and tables. The four ML algorithms selected for the spectral classification process were: a) Canonical Discriminant Analysis (CDA), which is a multivariate analysis algorithm with a procedure for grouping individuals from a previously defined group into exclusive classes of a group of independent variables (LARK, 1995) estimated from an error matrix. A systematic classification of the questions that such a map is required to answer is proposed. In each case the utility of the map is best measured by a different subset of the components of accuracy. It follows that no one map will be optimal from the point of view of every user (given that the perfect map cannot be made; b) Random Forest (RF), a model tolerant of noisy data which evaluates correlations between

variables using a random vector. The RF performance is high in setting spectral reflectance measurements, because of its low sensitivity to outliers (FLETCHER & REDDY, 2016; HONG et al., 2019) Progeny 5160, and Progeny 5460; c) Support Vector Machine (SVM), a classifier that discriminates using separation hyper planes with support vectors, limiting the division area between the classes (MA & GUO, 2014); and d) Light Gradient Boosting Machine (LGBM), a gradient structure that uses learning algorithms on trees that grow vertically (FAN et al., 2019) e.g. irrigation scheduling design, agricultural water management, crop growth modeling and drought assessment. Nevertheless, reliable estimation of ETo is difficult when lack of complete or long-term meteorological data at the target station. This study evaluated the efficiency of a new tree-based soft computing model, Light Gradient Boosting Machine (LightGBM).

The training samples were selected at random from a data set with 70% ($n = 2077$) of reflectance spectra, with the remaining 30% ($n = 890$) being reserved for testing and validation of ML models. The quality of the validation procedure was evaluated by comparing some commonly used indicators of the performance of ML algorithms, such as Classification Accuracy, Area Under the ROC Curve (AUC), F1 Score, and Kappa, besides other parameters for validation metrics as Precision, Recall and Support. Finally, the wavelengths more relevant for the classifications were revealed through calculation of the Average Impact Magnitude parameter, using values from the SHAP library which allow identification of the more important features to the model, thus explaining the output of the machine learning model being studied.

RESULTS AND DISCUSSION

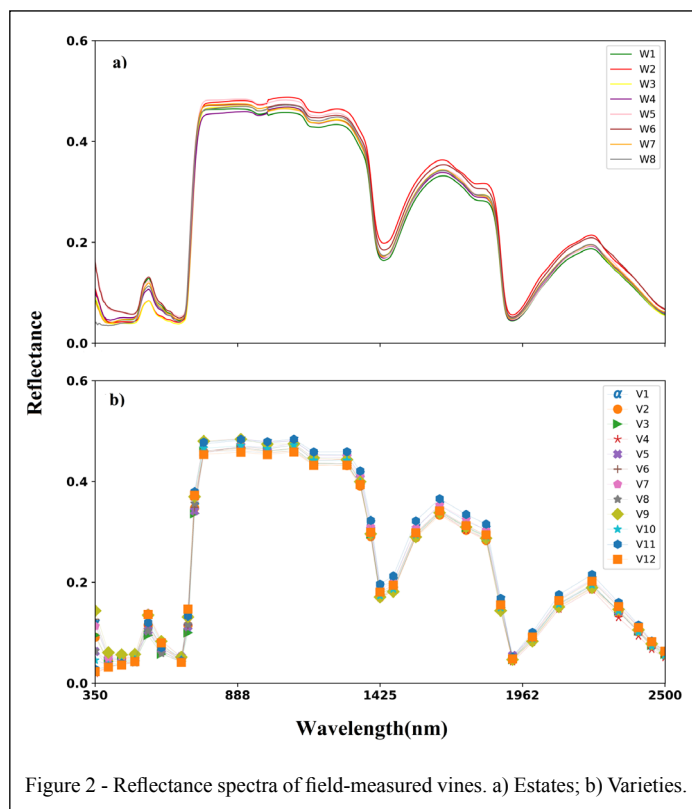
Average spectra for each Estate and each Variety classes are provided in figure 2. As expected, all spectra display the usual features typical of healthy vegetation, with subtle differences between classes which will be further discussed in what follows.

Results from the correlational Spearman test by coefficients are shown in figure 3, where in figures 3a and 3b R^2 values are presented. Values of R^2 as high as 0.6 were observed for the spectral ranges corresponding to the UV (350 to 399nm), NIR (780nm), and SWIR (1100 to 2300nm) for both groups. In the figures, areas next to the main diagonal have strong associations between their wavelengths, while coefficients with lower R^2 values, the darkest colors, indicate the low collinearity between wavelengths.

Figures 3c and 3d showed the p -values, where the wavelengths located at the main diagonal or nearby present determination coefficients above 0.9 and p -values < 0.05 , indicating statistical significance. After a correlational analysis has identified the spectral regions with low correlation (p -value < 0.05), fourteen wavelengths were selected as indicators of the most conspicuous spectral differences between the studied classes as revealed by the SR technique. These wavelengths were: 350nm; 358nm; 365nm; 467nm; 574nm; 705nm; 1350nm; 1410nm; 1420nm; 1723nm; 1850nm; 1894nm; 2306nm; and 2500nm.

Results from the non-parametric Kruskal-Wallis test for the fourteen wavelengths indicated significant differences ($P < 0.05$) at 365nm, 1350nm, 1420nm, 1850nm and 2306 nm at all Estates. The feasibility of spectral separability between classes within the Estates group has been previously reported, leading to the discrimination between vineyards located in different regions, a perception linked to the terroir concept expressing the soil-plant-climate-management relationship (CEMIN & DUCATI, 2011; THUM et al., 2020). In the Varieties group, the wavelengths 350nm, 358nm, and 574nm are the more suited to variety separation, while at 2500nm

little separation is achieved. These results; therefore, suggested that: a) variations either in region or in variety have a significant effect in the ultraviolet reflectance of vines (at 350nm, 358nm, and 365nm); b) concerning chlorophyll, these variations do not have a major effect on the 467nm band, and none at all at the 660nm band; c) a significant effect at near-infrared (NIR) bands was observed for region variation, and here it can be noted that in former studies a group of grape varieties was discriminated by hyperspectral sensors, pointing out the VIS and NIR spectral regions as crucial in the separability of vineyards (KARAKIZI et al., 2016; MIRZAEI et al., 2019) efficient and automated methods are required for the accurate detection of vegetation, crops and different crop varieties. To this end, we have designed, developed and evaluated an object-based classification framework towards the detection of vineyards, the vine canopy extraction and the vine variety discrimination from very high resolution multispectral data. A novel set of spectral, spatial and textural features, as well as rules, segmentation scales and a set of parameters are proposed based on object-based image analysis. The validation of the developed methodology was carried out on



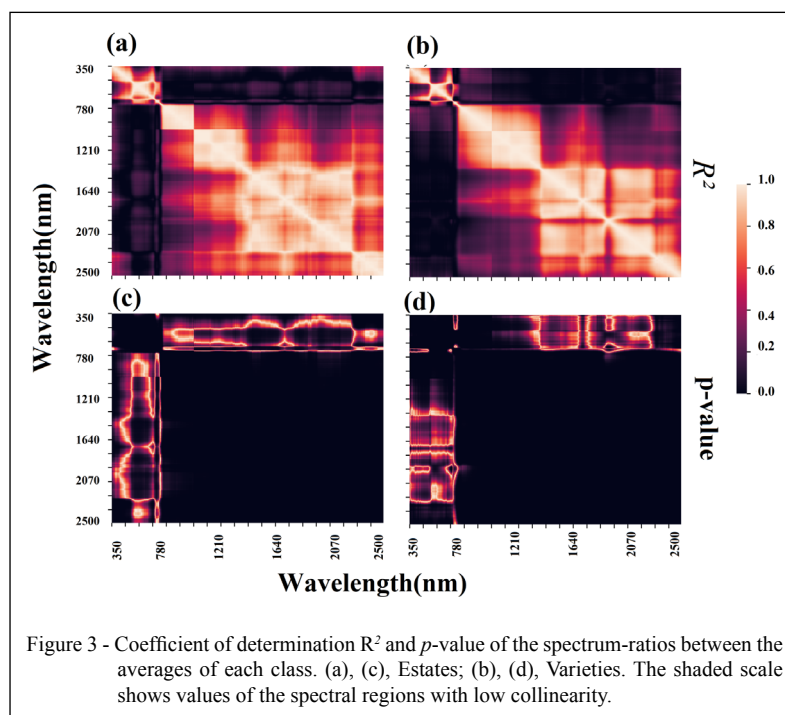


Figure 3 - Coefficient of determination R^2 and p -value of the spectrum-ratios between the averages of each class. (a), (c), Estates; (b), (d), Varieties. The shaded scale shows values of the spectral regions with low collinearity.

multitemporal WorldView-2 satellite data at four different viticulture regions in Greece. Concurrent in situ canopy reflectance observations were acquired from a portable spectroradiometer during the field campaigns. The performed quantitative evaluation indicated that the developed approach managed in all cases to detect vineyards with high completeness and correctness detection rates, i.e., over 89%. The vine canopy extraction methodology was validated with overall accuracy (OA; and d) the water absorption bands usually observed in vegetation (at 1450nm, 1950nm, and 2500nm) seem to have little importance on differentiation of vines induced by variation of region or variety.

The models' performance is presented in table 2. The highest predictive accuracies for classification are those of the LGBM algorithm, with a maximum accuracy range of 0.99. For both the Estate and Variety groups, the best performances were attained by LGBM, followed by RF. For the dimensionality reduction, the best performance came from the SR technique, but the KPCA method also yielded satisfactory results. Comparing KPCA and SR performances, the set of wavelengths extracted by SR showed an increase in performance from 0.91 to 0.93 (Estate) and 0.69 to 0.88 (Variety) using the LGBM algorithm and for RF accuracy raised from 0.74 to 0.92 (Estate) and from 0.45 to 0.89 (Variety).

The CDA and SVM algorithms did not perform well by KPCA but showed significant improvements in their metrics for discrimination by SR.

The spectral separation between classes internal to the groups (Estates or Varieties) is shown in figure 4, which displays the AUC values derived from the LGBM algorithm, the one with best performance, for both KPCA and SR. In this figure it is possible to assess the separability between classes by inspecting the relations between true or false positives; the more AUC values are near 1, the better the separation. Most AUC values were above 0.90, with the best fits to the discrimination being obtained by the SR method. For example, in figure 4, using as input data the set generated by the SR method, for the class W6 the AUC value was 0.95, while using KPCA we had $AUC = 0.90$; at the Varieties Group, for V8 we had $AUC = 0.99$ from SR and $AUC = 0.70$ from KPCA. Therefore, significant separability for both groups was achieved using the LGBM model with both reduction methods, with some advantage to SR.

The classification metrics (Figures 5a and 5b) presents the performance of each class through wavelengths extraction by SR. In figure 5a, the vineyards W4 and W6 obtained the smallest Recall (0.606 and 0.722) and F1-Score (0.684 and 0.765). With respect to separation between W2 and W3, which are 2km apart and on the same bedrock, inspection

Table 2 - Results obtained for spectral discrimination between the Leaf reflectance measured.

Class	Reduction	Model	Accuracy	AUC	F1	Kappa
Estate	KPCA	LGBM	0.91	0.99	0.91	0.89
		RF	0.74	0.97	0.71	0.69
		CDA	0.50	0.71	0.50	0.42
		SVM	0.48	0.00	0.39	0.38
	SR	LGBM	0.93	0.99	0.93	0.92
		RF	0.92	0.99	0.92	0.90
		CDA	0.92	0.99	0.92	0.91
		SVM	0.61	0.00	0.55	0.54
Variety	KPCA	LGBM	0.69	0.92	0.67	0.60
		RF	0.45	0.82	0.36	0.24
		CDA	0.17	0.53	0.16	0.07
		SVM	0.31	0.00	0.24	0.12
	SR	LGBM	0.88	0.98	0.88	0.86
		RF	0.89	0.98	0.89	0.87
		CDA	0.67	0.91	0.66	0.59
		SVM	0.41	0.00	0.36	0.24

of figure 5a reveals that classes W2 and W3 display similarity between True Positive and False Positive values, having AUC values near 1; therefore, these two classes show similar classification accuracies,

being nevertheless separable, what can be explained by the fact that, even if being on the same bedrock, they have different soil profiles, with a possible influence on plant development. It can be noted that

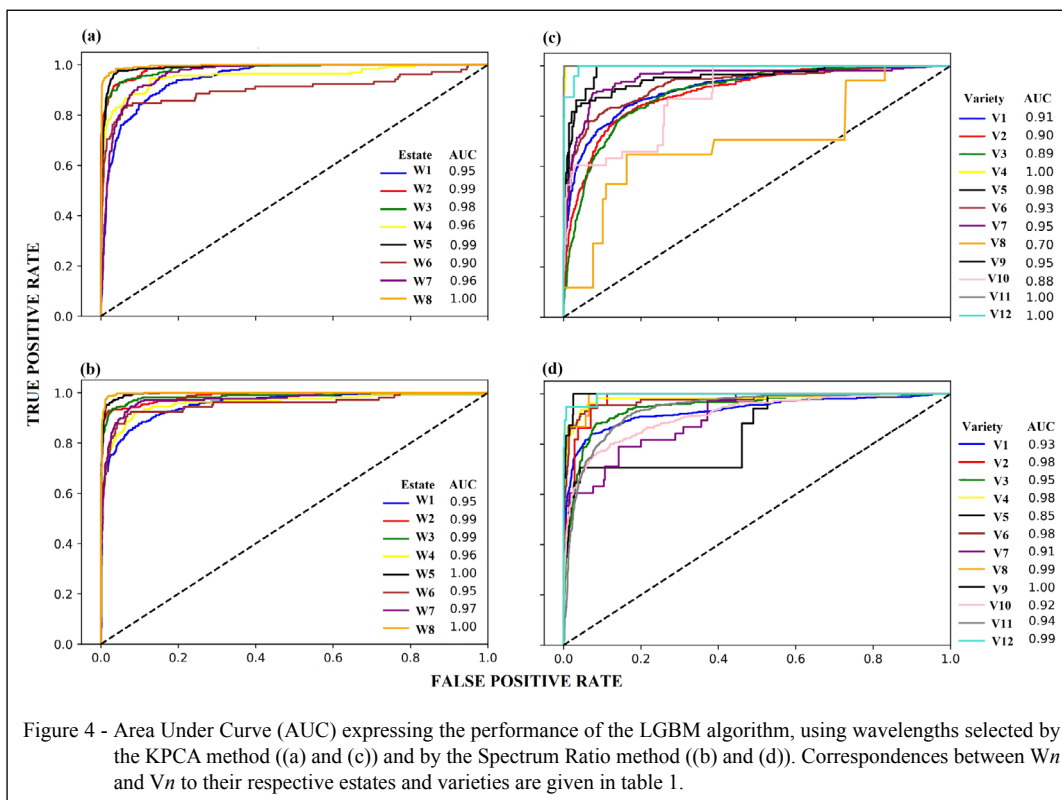


Figure 4 - Area Under Curve (AUC) expressing the performance of the LGBM algorithm, using wavelengths selected by the KPCA method ((a) and (c)) and by the Spectrum Ratio method ((b) and (d)). Correspondences between W_n and V_n to their respective estates and varieties are given in table 1.

W2 and W3 belong to the same owner and have the same management, what excludes differentiation due to anthropogenic factors. Still focusing on figure 5a, it can be seen that estates W1 and W7, both located at the Campanha Gaúcha viticultural region, are fairly separated, indicating non-negligible spectral differences; this fact, added to the one that W7 is on a transition of sandstone to clay, reinforces current perceptions that the presently established limits of this viticultural region are too wide, pointing to the future need of its division in more uniform territorial units. In figure 5b, the result of classification between varieties indicates for V7 and V8 the smallest Recall (0.667 and 0.444) and F1-Score (0.800 and 0.615). The lowest precision was shown by V3, with a value of 0.647. Estates W5 and W8 and varieties V1, V2, V6, V10, V11, and V12 obtained the best performances, all of them with values of F1-Score above 0.9. Furthermore, both groups obtained good discrimination accuracy, indicating the feasibility of spectral separability at leaf level.

Finally, the average Impact Magnitude of the wavelengths on the LGBM model using feature

extraction by the SR method is shown in figures 5c and 5d. The ultraviolet wavelengths (358nm, 574nm, and 365nm, in order of importance) presented a greater average impact magnitude for discrimination between Estates. The Variety classes displayed a similar average impact magnitude. The wavelengths in these spectral regions (green, blue, and ultraviolet) are important to detect changes in reflectance due to changes in pigment content (MERZLYAK et al., 1999), carotenoids (GITELSON et al., 2002), and anthocyanins (PROSHKIN et al., 2021)B, and C ranges (as additives to the main light at leaf level).

Two additional perceptions must be noted. The spectral differences between classes, especially those revealed in the fourteen wavelengths described above, are subtle, as reported elsewhere (DELALIEUX et al., 2007; ETTABAA & SALEM, 2017); in fact, taking as reference the usual range of reflectance values (from zero to unity), the conspicuous differences revealed by the spectrum-ratio technique are of the order of 10^{-4} or even smaller. Their detection is due to the extreme signal-to-noise ratio of the

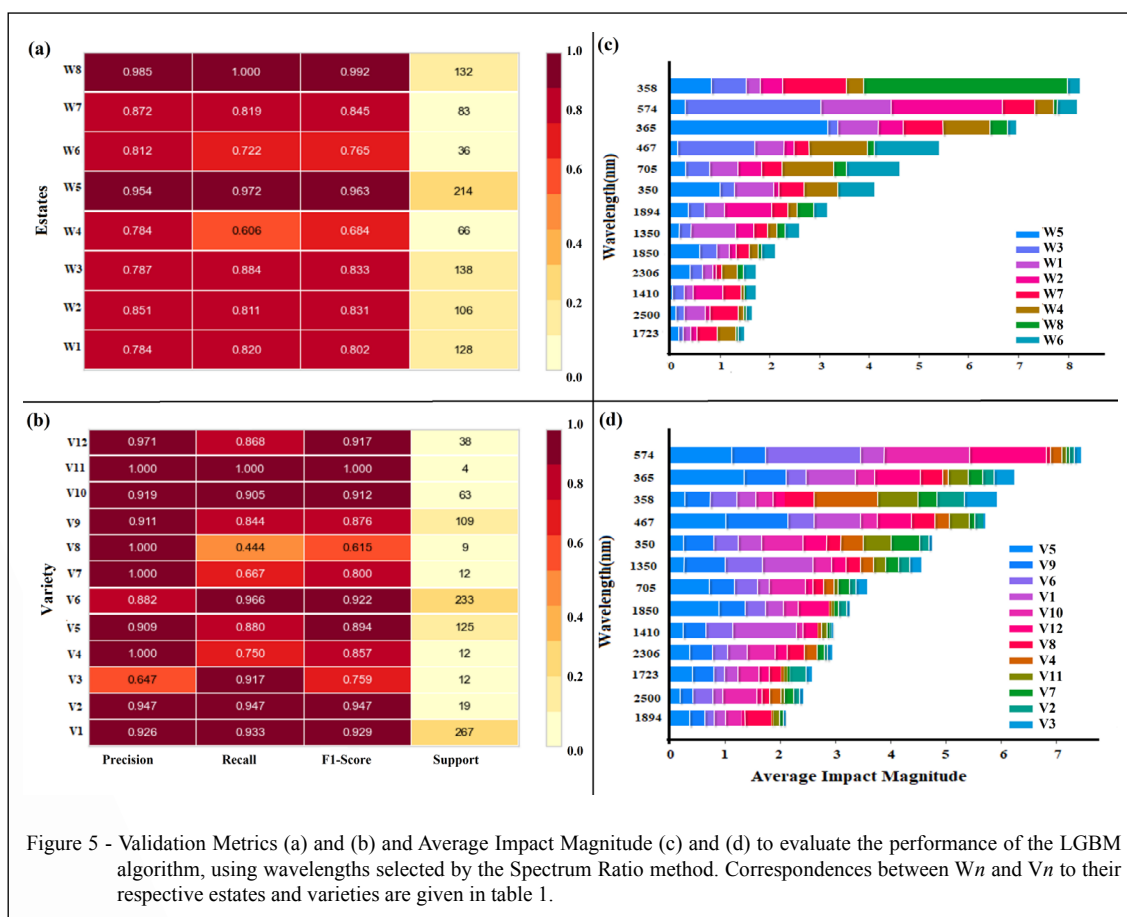


Figure 5 - Validation Metrics (a) and (b) and Average Impact Magnitude (c) and (d) to evaluate the performance of the LGBM algorithm, using wavelengths selected by the Spectrum Ratio method. Correspondences between W_n and V_n to their respective estates and varieties are given in table 1.

measurements taken with the equipment presently employed, leading to the significant detection of faint spectral features. A lengthy discussion of this point can be found at former research reported by our group (PITHAN et al, 2021). Finally, the results presented here do not suggest a capability, from our data and analysis, to separate between red and white grape varieties (classes V1, V3, V4, V6, V9, V10 and V11 are red grapes); however, it was reported by SILVA & DUCATI (2009) that, using ASTER satellite data, these two greater classes can be discriminated. This is intriguing, since the spectral resolution of ASTER images is much coarser. A possible explanation may come from the classification algorithm used on the images, the maximum likelihood, which was not presently used.

From these results, it seems that purely environmental variations (bedrock, climate) are not decisive to differentiation within the Estates group, since, for example, the Estates on volcanic rocks (W2, W3, W5 and W6), all of them with a more humid climate, do not form a separate group. This suggested that more complex processes are involved in the construction of reflectance spectra of vines (or of vegetation in general) confronted to environmental changes.

CONCLUSION

In this research, we investigated the potential of field hyperspectral leaf reflectance measurements to differentiate grape varieties and grape production regions. Our results have demonstrated that such separability is indeed possible, with significant accuracies. Acquiring spectral information about the vines *in situ*, without removal of leaves for laboratory analysis represents a gain both in costs and in logistical preparations. Due to its extreme signal-to-noise ratio, allowing the detection of subtle spectral features, the hyperspectral proximal sensor data presently used was a crucial tool in the detailing of faint leaf traits, making possible to discriminate grapevine varieties and the influence of environmental aspects. In this sense, our results can contribute to the comprehension of terroir issues related to regional variations, as discussed by VAN LEEUWEN & SEGUIN (2007). In fact, focusing on the presently demonstrated capability of spectrally separating regions, even when the bedrock is similar (being the cases of estates W2, W3, W5, and W6, all on volcanic acidic rocks), we saw that the geological similarity was not a confounding factor; these classes were fairly separated, suggesting that additional discriminating factors, like climate, also play a role

on plant development leading to specific spectral traits in leaf reflectance.

The wavelength extraction by the SR technique demonstrated advantages over the KPCA method when both were used for classification with the LGBM algorithm. This paper points towards the feasibility of the spectral discrimination of grapevines at leaf level, using a non-destructive method, for identification of vine varieties and their region, with applications valuable to the producer, allowing building a spectral library of grape wines.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance code 001, Award no. 8888248939/2019-1 and no. 88887.488339/2020-00.

DECLARATION OF CONFLICT OF INTEREST

We have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

All authors contributed equally for the conception and writing of the manuscript. All authors critically revised the manuscript and approved of the final version.

REFERENCES

- ANGUITA, D. et al. Model selection for support vector machines: Advantages and disadvantages of the Machine Learning Theory. In: **Proceedings of the International Joint Conference on Neural Networks**. 2010 Available from: <<https://ieeexplore.ieee.org/document/5596450/>> Accessed: Aug. 05, 2010. doi: 10.1109/IJCNN.2010.5596450.
- CEMIN, G.; DUCATI, J. R. Spectral Discrimination of Grape Varieties and a Search for Terroir Effects Using Remote Sensing. **Journal of Wine Research**, Mar. 2011. v.22, n.1, p.57–78. Available from: <<https://doi.org/10.1080/09571264.2011.550762>>. Accessed: Oct. 15, 2018. doi: 10.1080/09571264.2011.550762.
- CEROVIC, Z. G. et al. A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. **Physiologia Plantarum**, 6 Nov. 2012. v.146, n.3, p.251–260. Available from: <<https://doi.org/10.1111/j.1399-3054.2012.01639.x>>. Accessed: Nov. 15, 2019. doi: 10.1111/j.1399-3054.2012.01639.x.
- CAMOZZATO, E; LOPES, R. C. **Carta Geológica Hulha Negra**, Folha S H.22-Y-C-I. Estado do RS. Escala 1:100.000. 2012. CPRM, Porto Alegre. Available from: <<https://rigeo.cprm.gov.br/handle/doc/19253>>. Accessed: Nov. 15, 2021.
- CONRADIE, W. J. Seasonal Uptake of Nutrients by Chenin Blanc in Sand Culture: I. Phosphorus, Potassium, Calcium and Magnesium. **South African Journal of Enology and Viticulture**

- 2: 7–13. 1981. Available from: <<https://doi.org/10.21548/2-1-2403.9176>>. Accessed: Nov. 15, 2019. doi: 10.21548/2-1-2403.9176.
- DELALIEUX, S., et al. Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. **European Journal of Agronomy**, 2007(1), 130–143. Available from: <<https://doi.org/10.1016/J.EJA.2007.02.005>>. Accessed: Oct. 02, 2020. doi: 10.1016/J.EJA.2007.02.005.
- DEMATTE, J. A. M. Characterization and discrimination of soils by their reflected electromagnetic energy. **Pesquisa Agropecuária Brasileira**, 2002, p.1445–1458. Available from: <<https://doi.org/10.1590/S0100-204X2002001000013>>. Accessed: Oct. 02, 2020. doi: 10.1590/S0100-204X2002001000013.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Geologia. 2018. Available from: <http://geofp.ibge.gov.br/informacoes_ambientais/geologia/levantamento_geologico/vetores/escala_250_mil/>. Accessed: Jan. 02, 2020.
- ETTABAA, K. S, SALEM, M. B. Adaptive Progressive Band Selection for Dimensionality Reduction in Hyperspectral Images. **Journal of the Indian Society of Remote Sensing**. 2017. 46:2, 46(2), 157–167. Available from: <<https://doi.org/10.1007/S12524-017-0691-9>>. Accessed: Oct. 02, 2020. doi: 10.1007/S12524-017-0691-9.
- FAN, J. et al. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. **Agricultural Water Management**, Nov. 2019. v.225, n. August, p.105758. Available from: <<https://doi.org/10.1016/j.agwat.2019.105758>>. Accessed: Jan. 06, 2020. doi: 10.1016/j.agwat.2019.105758.
- FAUVEL, M. et al. Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data Over Urban Areas. **EURASIP Journal on Advances in Signal Processing**, 22 Dec. 2009. v.2009, n.1, p.783194. Available from: <<https://doi.org/10.1155/2009/783194>>. Accessed: Jan. 06, 2020. doi: 10.1155/2009/783194.
- FLETCHER, R. S.; REDDY, K. N. Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. **Computers and Electronics in Agriculture**, 1 Oct. 2016. v.128, p.199–206. Available from: <<https://doi.org/10.1016/j.compag.2016.09.004>>. Accessed: Jan. 06, 2020. doi: 10.1016/j.compag.2016.09.004.
- GITELSON, A. et al. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. **Photochemistry and Photobiology**, 2002. v.75, n.3, p.272. Available from: <[https://doi.org/10.1562/0031-8655\(2002\)075<0272:ACCIPL>2.0.CO;2](https://doi.org/10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2)>. Accessed: Jan. 25, 2018. doi: 10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2.
- HONG, Y. et al. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. **Science of the Total Environment**, Feb. 2019. v.651, p.1969–1982. Available from: <<https://doi.org/10.1016/j.scitotenv.2018.09.391>>. Accessed: Jan. 15, 2020. doi: 10.1016/j.scitotenv.2018.09.391.
- KARAKIZI, C. et al. Vineyard detection and vine variety discrimination from very high-resolution satellite data. **Remote Sensing**, 2016. v.8, n.3, p.1–25. Available from: <<https://doi.org/10.3390/rs8030235>>. Accessed: Feb. 15, 2020. doi: 10.3390/rs8030235.
- LARK, R. M. Components of accuracy of maps with special reference to discriminant analysis on remote sensor data. **International Journal of Remote Sensing**, 20 May. 1995. v.16, n.8, p.1461–1480. Available from: <<https://doi.org/10.1080/01431169508954488>>. Accessed: Feb. 15, 2020. doi: 10.1080/01431169508954488.
- LORENZ, D. H. 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 Jul. 1995. v.1, n.2, p.100–103. Available from: <<https://doi.org/10.1111/j.1755-0238.1995.tb00085.x>>. Accessed: Jan. 16, 2020. doi: 10.1111/j.1755-0238.1995.tb00085.x.
- MA, Y.; GUO, G. **Support Vector Machines Applications**. Cham: Springer International Publishing, 2014. v.9783319023. Available from: <<https://doi.org/10.1007/978-3-319-02300-7>>. Accessed: Feb. 10, 2019. doi: 10.1007/978-3-319-02300-7.
- MERZLYAK, M. N. et al. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. **Physiologia Plantarum**, May. 1999. v.106, n.1, p.135–141. Available from: <<https://doi.org/10.1034/j.1399-3054.1999.106119.x>>. Accessed: Feb. 25, 2019. doi: 10.1034/j.1399-3054.1999.106119.x.
- MIRZAEI, M. et al. Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. **International Journal of Applied Earth Observation and Geoinformation**, 2019. v.80, n. January, p.26–37. Available from: <<https://doi.org/10.1016/j.jag.2019.04.002>>. Accessed: Feb. 25, 2019. doi: 10.1016/j.jag.2019.04.002.
- MOGHIMI, A. et al. A novel machine learning approach to estimate grapevine leaf nitrogen concentration using aerial multispectral imagery. **Remote Sensing**, 26 Oct. 2020. v.12, n.21, p.1–21. Available from: <<https://doi.org/10.3390/rs12213515>>. Accessed: Nov. 25, 2020. doi: 10.3390/rs12213515.
- PITHAN, P. A. et al. Spectral characterization of fungal diseases downy mildew, powdery mildew, black-foot and Petri disease on *Vitis vinifera* leaves. **International Journal of Remote Sensing**, 3 Aug. 2021. v.42, n.15, p.5680–5697. Available from: <<https://doi.org/10.1080/01431161.2021.1929542>>. Accessed: Aug. 25, 2021. doi: 10.1080/01431161.2021.1929542.
- PROSHKIN, Y. A. et al. Assessment of ultraviolet impact on main pigment content in purple basil (*Ocimum basilicum* L.) by the spectrometric method and hyperspectral images analysis. **Applied Sciences (Switzerland)**, 2021. v. 11, n. 19. Available from: <<https://doi.org/10.3390/app11198804>>. Accessed: Feb. 25, 2019. doi: 10.3390/app11198804.
- RENZULLO, L. J.; BLANCHFIELD, A. L.; POWELL, K. S. A method of wavelength selection and spectral discrimination of hyperspectral reflectance spectrometry. **IEEE Transactions on Geoscience and Remote Sensing**, 2006. v.44, n.7, p.1986–1994. Available from: <<https://doi.org/10.1109/TGRS.2006.870441>>. Accessed: Feb. 25, 2019. doi: 10.1109/TGRS.2006.870441.
- ROSSETI, L.; et al. Lithostratigraphy and volcanology of the Serra Geral Group, Paraná-Etendeka Igneous Province in Southern Brazil:

- Towards a formal., **J. Volcanol. Geotherm. Res.** 2017. Available from: <<http://dx.doi.org/10.1016/j.jvolgeores.2017.05.008>>. Accessed: Oct. 25, 2019. doi: 10.1016/j.jvolgeores.2017.05.008.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, 1 Jul. 1964. v.36, n.8, p.1627–1639. Available from: <<https://doi.org/10.1021/ac60214a047>>. Accessed: Sep. 05, 2019. doi: 10.1021/ac60214a047.
- SILVA, P. R.; DUCATI, J. R. Spectral features of vineyards in south Brazil from ASTER imaging. **International Journal of Remote Sensing**, 23 nov. 2009. v.30, n.23, p.6085–6098. Available from: <<https://doi.org/10.1080/01431160902810612>>. Accessed: Sep. 05, 2017. doi: 10.1080/01431160902810612.
- SCHREINER, R. P. Nutrient Uptake and Distribution in Young Pinot Noir Grapevines over Two Seasons. **American Journal of Enology and Viticulture**. 2016. 67: p.436–448, 2016. Available from: <<https://www.ajevonline.org/content/67/4/436>>. Accessed: Sep. 05, 2018. doi: 10.5344/ajev.2016.16019.
- SCHREINER, R. P. et al. Nutrient Uptake and Distribution in a Mature ‘Pinot Noir’ Vineyard. **HortScience**, v.41, p.336–345, 2006. Available from: <<https://journals.ashs.org/hortsci/view/journals/hortsci/41/2/article-p336.xml>>. Accessed: Nov. 05, 2015. doi: 10.21273/HORTSCI.41.2.336.
- SMIT, J. L. et al. Vine signal extraction - an application of remote sensing in precision viticulture. **South African Journal of Enology and Viticulture**, v.31, n.2, p.65–74. 2010. Available from: <<https://doi.org/10.21548/31-2-1402>>. Accessed: Feb. 25, 2019. doi: 10.21548/31-2-1402.
- THUM, A. B. et al. The influence of mineral content on spectral features of vine leaves. **International Journal of Remote Sensing**, v.41, n.23, p.9161–9179, 2020. Available from: <<https://doi.org/10.1080/01431161.2020.1798547>>. Accessed: Dec. 5, 2020. doi: 10.1080/01431161.2020.1798547.
- VAN LEEUWEN, C.; SEGUIN, G. **The concept of terroir in viticulture**. 2007. Available from: <<http://dx.doi.org/10.1080/09571260600633135>>. Accessed: Oct. 15, 2015. doi: 10.1080/09571260600633135.
- VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, 3 Feb. 2020. v.17, n.3, p.261–272. Available from: <<https://doi.org/10.1038/s41592-019-0686-2>>. Accessed: Feb. 25, 2019. doi: 10.1038/s41592-019-0686-2.
- WHITE, R. E. **Understanding Vineyard Soils**. Oxford: Oxford University Press, 2009.
- WILDNER, W. et al. **Mapa Geológico do Estado do Rio Grande do Sul**, Escala 1:750.000 [Geological Map of the State of Rio Grande do Sul, Scale 1:750.000]. 2008. Porto Alegre, Brazil: CPRM. Available from: <<http://www.cprm.gov.br/publique/Geologia/Geologia-Basica/Cartografia-Geologica-Regional-624.html>>. Accessed: Feb. 25, 2019.
- ZHANG, C. et al. Separating Mangrove Species and Conditions Using Laboratory Hyperspectral Data: A Case Study of a Degraded Mangrove Forest of the Mexican Pacific. **Remote Sensing**, 25 Nov. 2014. v.6, n.12, p.11673–11688. Available from: <<https://doi.org/10.3390/rs61211673>>. Accessed: Dec. 02, 2021. doi: 10.3390/rs61211673.