Ciência Rural, Santa Maria, v.54:07, e20220558, 2024
http://doi.org/10.1590/0103-8478cr20220558
ISSNe 1678-4596
RURAL ENGINEERING

Hyperspectral data analysis for chlorophyll content derivation in vineyards

Diniz Carvalho de Arruda[1*] **Jorge Ricardo Ducati**[1] **Pâmela Aude Pithan**[1] **Adriane Brill Thum**[2] **Rosemary Hoff**[3]

[1]Programa de Pós-graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul (UFRGS), 91501-970, Porto Alegre, RS, Brasil. E-mail: dinizcarvalho28@outlook.com. *Corresponding author.
[2]Escola Politécnica, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brasil.
[3]Embrapa Uva e Vinho, Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Bento Gonçalves, RS, Brasil.

ABSTRACT: Quality and yield of a vineyard are related to canopy biomass and leaf vigor, and proximal techniques have been used as alternatives to conventional methods to estimate these parameters. Knowledge on chlorophyll content is crucial to plant health assessments. However, chlorophyll indices can also be extracted from reflectance spectra obtained for an ample range of applications. In this perspective, relations between chlorophyll indices obtained by direct measurements and derived from field radiometry were investigated, with the aim to assess the accuracy of predicted chlorophyll content. The investigation was performed on Cabernet Sauvignon vines, being based on direct chlorophyll surveys, vine leaf spectroradiometry and the derivation of Hyperspectral Vegetation Indices (HVIs), with data acquisition being performed on two stages of the vegetative cycle. Direct chlorophyll data was compared with predicted indices using two machine learning algorithms: Partial Least-Squares Regression (PLSR) and Random Forest Regressor (RFR), using data from reflectance spectra and derived HVIs. The higher correlations between measurements and predictions were obtained for *Chl a* and *Chl a/Chl b* modeled by the RFR algorithm, with R^2 values as high as 0.8 and Root Mean Squared Errors as low as 0.093. With respect to HVIs, the Photochemical Reflectance Index (PRI) calculated for the second acquisition date, corresponding to leaves reaching senescence was the one which produced the highest percentage of prediction explanations. This study can bring a significant contribution to the development of non-invasive techniques to vine monitoring.
Key words: hyperspectral, vineyards, partial least-squares regression, random forest regressor.

Análise de dados hiperespectrais para derivação do teor de clorofila em videiras

RESUMO: A qualidade e a produtividade de um vinhedo estão relacionadas com a biomassa do dossel e o vigor foliar, e técnicas de sensoriamento próximo têm sido utilizadas como alternativas aos métodos convencionais para estimar esses parâmetros. O conhecimento do teor de clorofila é fundamental para as avaliações fitossanitárias. No entanto, índices de clorofila também podem ser extraídos de espectros de refletância obtidos para uma ampla gama de aplicações. Nesta perspectiva, foram investigadas as relações entre os índices de clorofila obtidos por medidas diretas e derivados de radiometria de campo, com o objetivo de avaliar a acurácia do teor de clorofila previsto. A investigação foi realizada em plantas da variedade Cabernet Sauvignon, baseando-se em levantamentos diretos de clorofila, espectrorradiometria foliar e na derivação de Índices de Vegetação Hiperespectrais (HVIs), sendo a aquisição de dados realizada em duas fases do ciclo vegetativo. Os resultados das estimativas mostraram que os maiores coeficientes de determinação expressando a correlação entre medições e predições foram obtidas para *Chl a* e *Chl a/Chl b* modeladas pelo algoritmo RFR, com valores de R^2 tão altos quanto 0,8 e erros quadráticos médios tão baixos quanto 0,093. Com relação aos HVIs, o Photochemical Reflectance Index (PRI) calculado para a segunda data de aquisição, correspondente às folhas que atingiram a senescência, foi o que produziu o maior percentual de explicações de predição. Em conclusão, sugere-se que este estudo pode trazer uma contribuição significativa para o desenvolvimento de técnicas não invasivas de monitoramento de vinhedos.
Palavras-chave: hiperespectral, vinhedos, partial least-squares regression, random forest regressor.

INTRODUCTION

Quality and yield of a vineyard are related to canopy biomass and leaf vigor. Along the vegetative cycle, leaf vigor is an important indicator of plant health status (BERGSTRÄSSER et al., 2015; LACAR et al., 2002). In this perspective, remote detection techniques have been applied in studies based on plant spectral patterns, aiming to analyze vegetative development, phenological dynamics, management practices and a diversity of stresses due to biotic and abiotic attacks, either focused in vineyards (LOGGENBERG et al., 2018; THUM et al., 2020) or in other plant species (ZHANG et al., 2013).

One of the most important parameters for the assessment of the vegetative conditions of plants is chlorophyll content; for example, chlorophyll levels are important indicators to the monitoring of nitrogen content in leaves (ARGENTA et al., 2004). From the various chlorophyll types, chlorophyll *a* and chlorophyll *b* are fundamental constituents of the photosynthetic apparatus in most plant species; chlorophyll *a* is essential in photochemistry, while chlorophyll *b* is necessary for stabilizing the major

Received 10.06.22 Approved 08.30.23 Returned by the author 01.29.24
CR-2022-0558.R1
Editors: Leandro Souza da Silva Alberto Cargnelutti Filho

light-harvesting chlorophyll-binding proteins, the ratio of chlorophyll *a* to *b* being in the range 2.5 to 5 (TANAKA & TANAKA, 2011). This ratio generally changes during the vegetative cycle, one of the main factors for change being the available illumination, which has variations as the vegetative season advances (SESTAK, 1963; BENERAGAMA & GOTO, 2011) with a trend towards smaller ratios. Determination of chlorophyll levels are conventionally done by laboratory techniques applied on field-collected samples; however, as the characteristic plant green shades are due to the reflected light after interaction of illuminating radiation with leaf photosynthesizing pigments, chlorophyll amounts can be estimated by non-destructive methods (FASSNACHT et al., 2015; ORDÓÑEZ et al., 2018; STEELE et al., 2008) like reflectance analysis at hyperspectral resolutions, allowing to map with improved performance the spectral properties of plant leaves at visible and near infrared wavelengths, studying color changes, hemispherical reflectance and subtle variations in leaf tissues (ZHAO et al., 2014). Chlorophyll levels are also sensitive to water stress and to soil type (MITRA et al., 2018). Data from plant spectroscopy can be used to predictive models, supplying estimations of plant physiological and morphological traits as an alternative to conventional methods. In terms of remote sensing, the use of hyperspectral sensors to the detection of variations in leaf pigmentation is an improvement compared to multispectral data, as more detailed information becomes available.

Ample use has been made of spectroscopy to estimate vine descriptors (POWER et al., 2019). However, in hyperspectral data the large number of spectral bands tends to have a negative impact on the metrics expressing the performance of estimating models, a problem which is addressed by the use of dimensionality reduction techniques (LOGGENBERG et al., 2018; SAHEB ETTABAA & BEN SALEM, 2017) where machine learning algorithms perform a crucial role. Through machine learning models to data analysis, it is possible to regularize and reduce the number of wavelengths necessary to build structured spectral libraries. Models Partial Least Squared Regression (PLSR) and Random Forest Regression (RFR) are examples of robust algorithms to the characterization and analysis of spectral data, dimensionality reduction and parameters prediction using non-invasive methods (CHENG & SUN, 2017; EL-HENDAWY et al., 2019; KAWAMURA et al., 2017).

Furthermore, the arrival of new methods for data acquisition by *in situ* proximal remote sensing increased the potential to monitoring plant phenological dynamics during the growing cycle. Therefore, the objectives of this study were: a) to analyze the relationships between chlorophyll parameters and plant reflectance at visible (VIS, from 380nm to 700nm), near infrared (NIR, from 700nm to 1400nm) and short-wave infrared (SWIR, from 1400nm to 2500nm) wavelengths, derived from hyperspectral proximal data measured at leaf level, in commercial vineyards; b) to compare the performance of two machine learning models, PLRS and RFR for prediction of chlorophyll parameters; and c) to reveal the wavelengths more relevant to these tasks. It was expected that this study would bring a significant contribution to the development of non-invasive techniques to vine monitoring, contributing to vineyard management by allowing fast, low-cost, real-time interventions by the producer.

MATERIALS AND METHODS

Study area

As study area the Luiz Argenta Winery was chosen, due to its easy access and favorable topography. This estate is in a viticultural region called Vinhos dos Altos Montes (High Hills Wines), a geographical denomination ("Indicação de Procedência") located in north-east of Rio Grande do Sul State in south Brazil. Coordinates are 29º 01' 23.37" S and 51º11'02.23" O, being at a larger wine region called "Serra Gaúcha". The area with vines covers about forty-eight hectares with several *Vitis vinifera* grape varieties, with focus in the production of quality wines. All measurements were performed during the 2017/2018 season.

The number of grape varieties selected for this study was limited to only one variety, therefore avoiding possible confusing factors arising from intrinsic differences between varieties. Due to the relatively large number of parcels of Cabernet Sauvignon scattered along the terrain, allowing a diversity of local environmental conditions, this variety was chosen for the study; it is important to note that the choice of variety is not a crucial factor for this study, which could be performed with another variety, the criterion being to be widely available across the study area. Specific parcels of Cabernet Sauvignon were selected considering ease of access, topography, uniformity, and availability of information on soils, which in this case stay over basaltic to dacitic/ryolitic volcanic flows (acidic terms), which are the prevailing geological unit at Serra Gaúcha viticultural region. Six vine plots were studied, and following the estate use

they were called 4a, 4b, 16a, 16b, 19a, and 19b. Vines were planted in trellis driving system, on Paulsen 1103 rootstocks, distance between rows were 2.8m following east-west orientation (plots 4a, 4b, 16a, 16b) or north-south (19a, 19b), and distance between plants were 1.45m. These vineyards had conventional management with treatments on an approximate weekly basis. All plants used for the study were marked prior to the beginning of measurements.

To follow the evolution of spectral behavior of plants, measurements were performed in phases 81 and 83 of the vegetative cycle according to the BBCH scale (LORENZ et al., 1995) The first data acquisition was done in December 16, 2017, during the stage of the phenological cycle known as *véraison*, meaning the phase during which the berries acquire dark pigmentation. The second acquisition took place on February 27, 2018, during the final ripening and harvest.

Data acquisition and treatment

Chlorophyll and radiometric in situ measurements

All measurements in this study took place during a four-hour interval between 10AM and 2PM, to ensure uniformity of plant conditions leading to uniformity in acquired data; reasons for this protocol are presented in more detail in THUM et al. (2020), added to the fact that earlier measurements tend to be done over humid leaves, affecting leaf spectral features. A sample of twenty-four plants was selected, being four plants per parcel, located in the two central rows of each vine plot.

Chlorophyll and radiometric data were acquired in succession, beginning with transmittance measurements at 635nm, 660nm and 880nm using a Falker CFL1030 (Porto Alegre, Brazil) chlorophyll meter (SCHLICHTING et al., 2015), providing chlorophyll *a*, *b* and total chlorophyll content mediated by calibration using a white reference. These acquisitions were followed, for the same leaf, by spectroradiometric measurements using a Malvern Panalytical Spectral Devices (ASD, Westborough, MA, USA) FieldSpec® 3 spectroradiometer, which has spectral sensitivity between 350nm and 2500nm. A typical spectrum provides 2151 reflectance values between 0.0 and 1.0 at intervals of one nanometer, a calibration being made through measurements of a reference plate taken at regular time intervals. For each plant fours leaves were measured, and each leaf was measured twice, in different points of the adaxial face. All measurements were made with an attached Leaf Clip probe, which carries an internal halogen light source and an internal reference plate of Spectralon® (Labsphere, Inc., North Sutton, NH, USA). Every 15 minutes calibrations with white reference and optimization were performed, following the protocols described by PITHAN et al. (2021) and THUM et al. (2020). The total number of spectroradiometric measurement was 192 (24 plants *x* four leaves, twice); of these, 12 acquisitions were discarded for various reasons, and the final radiometric data base had 180 measurements.

All reflectance field data was recorded in ASD format, and managed in computer environment in Python language, where proprietary codes were created to spectra treatment, helped by public libraries. Exploratory data analysis was done using Panda's libraries. A frequent issue when acquiring and analyzing spectra collected by the equipment used in this study comes from the fact that data acquisition is done by a succession of three sensors, where the first one operates between 350nm and 1000nm, the second from 100 nm to 1800nm, with the third sensor being sensitive from 1800nm to 2500nm. Due to differences in sensitivity between these three sensors, the raw spectra shows discontinuities in the measured reflectance, seen as steps, at 1000nm and 1800nm. This issue is managed through smoothing procedures, and presently routines from the Specdal library with the jump correction function were used; likewise, noisy lines were smoothed by applying the Savitzky-Golay filter treatment.

Hyperspectral vegetation indices

Hyperspectral Vegetation Indices (HVIs) are defined using reflectance values at wavelengths selected according to customized, specific purposes, being therefore tuned to express characteristic plant metabolical functions. Using only a few selected wavelengths, HVIs, may avoid a hyperspectral data redundancy problem through the use of the more informative wavelengths, which are sensitive to plant characteristics such as cellular structure and biochemical and physiological processes. Many HVIs have been defined in the literature, and in this study we used 19 HVIs, calculated from the use of specific, discrete wavelengths, which were: Anthocyanin Reflectance Index 1 (ARI1) (GITELSON et al, 2001); Anthocyanin Reflectance Index 2 (ARI2) (GITELSON et al, 2001); Cellulose Absorption Index (CAI) (NAGLER et al., 2003); Chlorophyll Absorption in Reflectance Index (CARI); Carotenoid Reflectance Index 1 (CRI1) (GITELSON et al., 2002); Carotenoid Reflectance Index 2 (CRI2) (GITELSON et al., 2002); Leaf Water Vegetation Index 2 (LWVI-2) (GALVÃO et al., 2005); Modified Chlorophyll Absorption in Reflectance

Index (MCARI) (YANG et al., 2006); Normalized Difference Nitrogen Index (NDNI) (SERRANO et al., 2002); Normalized Difference Vegetation Index (NDVI) (ROUSE et al., 1974); Normalized Difference Water Index (NDWI) (GAO, 1996)); Photochemical Reflectance Index (PRI) (PEÑUELAS et al, 1995); Pigment Specific Normalized Difference 1 (PSND1) (BLACKBURN, 1998); Pigment Specific Normalized Difference 2 (PSND2) (BLACKBURN, 1998); Plant Senescence Reflectance Index (PSRI) (MERZLYAK et al., 1999); Plant Senescence Reflectance Index 2 (PSR2) (MERZLYAK et al., 1999); Structure Insensitive Pigment Index (SIPI) (PEÑUELAS et al., 1995); Vogelmann Red Edge 1 (VOG1) (VOGELMANN et al., 2007); and Water Index (WI) (PEÑUELAS et al., 1997).

Modeling process and prediction assessment

In this process we generated three data sets and applied the spectral dimension reduction process for each one. The three data sets were composed by: i) 2151 variables (only reflectance bands), ii) 19 HVIs, and iii) 2170 variables (reflectance bands + HVIs). The prediction response variable is represented by the chlorophyll parameters *Chl a*, *Chl b*, *Chl a* + *Chl b*, and *Chl a/Chl b*). We conducted the normalization process for each dataset with the Normalize function, to equalize the descriptors variation scale. The machine learning models used for the prediction analysis were the Partial Least Squares Regression (PLSR) and the Random Forest Regressor (RFR).

The PLSR is employed in sensor calibration and in spectral analysis associated with infrared and hyperspectral spectroscopy. It is a linear model, easy to fit and which presents low computational complexity (CHENG & SUN, 2017), being effective for selecting wavelengths employing score coefficients (MIRZAEI et al., 2019). Leaf traits and soil properties have been measured non-invasively through PLSR factors, for selection of spectral variables and estimation of physical-chemical parameters (EL-HENDAWY et al., 2019; THUM et al., 2020; ZHANG et al., 2017). The problems of high collinearities among wavelengths were solved by maximizing the covariance between measured and predicted (VISCARRA ROSSEL et al., 2006). Moreover, many authors used this model for reducing the dimensionality of wavelengths both at leaf and canopy levels (ABBASI et al., 2020; MIRZAEI et al., 2019).

The second algorithm, Random Forest Regressor, is a method based on predictions made in decision trees (estimators) randomly selected (BREIMAN, 2001). Empirically, it is possible to select the number of estimators, as tree depth, to develop the sample bagging, then extracting an Out of Bag (OOB) percentage (PALMER et al., 2007). To the final model an average of the results of the individual iterations was done. The selection of the wavelengths with greater impact was made from the importance of variables, calculated by the Gini index (NEMBRINI et al., 2018), which is a way, in the RFR model, to address the size of the relation between the wavelengths and the measured parameters.

Summing up the steps described above, the methodological approach used is this study was as follows, being also presented in figure 1:

1. In each one of the six vineyards, four plants were selected, and at each plant we selected a full-developed leaf, opposed to a grape cluster, under full solar illumination, located in the mean third of canopy, in a branch near the main vine trunk; being six parcels, a total of twenty-four leaves were selected.
2. Chlorophyll and radiometric measurements were performed at two phases of the vegetative cycle. These acquisitions produced, for each leaf and phase, a set of chlorophyll indices and a spectrum with 2151 reflectance values.
3. Using reflectance values for the required wavelengths, nineteen different hyperspectral vegetation indices (HVI) were calculated for each leaf.
4. As one of our aims were to predict chlorophyll concentrations from radiometric data, three sets of input data were assembled to be applied by the two models to be tested (RFR and PLSR). These three data sets were composed by: i) reflectance bands, with 2151 variables, Reflectance input data set, ii) hyperspectral vegetation indices, with 19 variables, HVI input data set; and iii) reflectance bands + HVIs, with 2170 variables.
5. Observed and predicted chlorophyll values were compared, and correlation accuracies were expressed, for each set of input data, by the following parameters: coefficient of determination (R^2), coefficient of determination with cross validation (R^2 (CV)), Root Mean Squared Error (RMSE), Root Mean Squared Error with cross validation (RMSE (CV)).

Validation of results was made using the method known as k-fold cross-validation. In this application, the data set is divided in k subsets, known as *folds*, and the model is trained and evaluated k times. At each iteration, a different fold is used as a test set, while the remaining sets are used for training; therefore, each fold is used as a test set only once. After k iterations, the performances from each

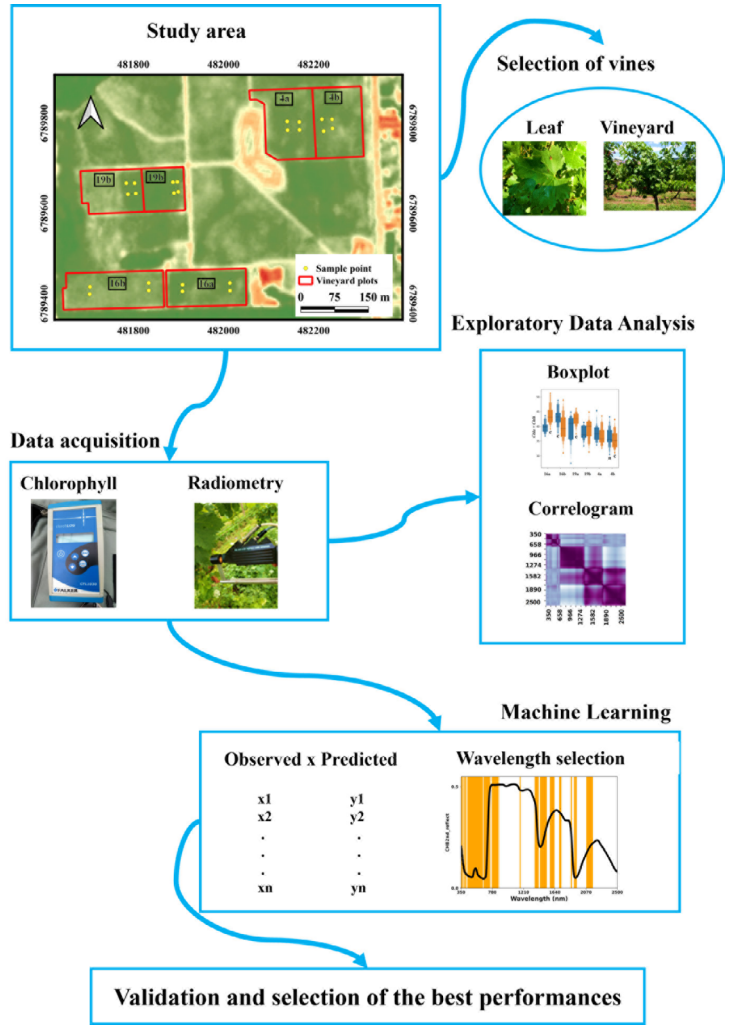

Figure 1 - Flowchart showing the main steps of this study.

fold are combined to produce an average estimate of the model performance. This method in general provides a more accurate estimate of the model performance, when compared to a conventional, single division of the data base in training and testing sub sets. This study was made from 180 spectroradiometric acquisitions, and validation was performed with the cross-validation machine learning algorithm, using five *k*-folds; in this case, the algorithm made five combinations with groups of 36 acquisitions (180/5), testing and combining these five combinations. Processing was done in Python, using library packages Pandas, Numpy, Scipy, Sk-learn and Matplotlib.

RESULTS AND DISCUSSION

Chlorophyll parameters

Variations in chlorophyll levels at the studied Cabernet Sauvignon parcels are shown in figure 2. Variations in *Chl a* were in general larger than variations in *Chl b*, with parcel 16a presenting the largest variation in *Chl a* at the first acquisition date, while parcels 4b and 16b showed large variations in *Chl a* at the second acquisition date. Levels of *Chl b* increased in the second acquisition date, a trend reported by BENEGARAMA & GOTO (2011).

Concerning ratios between chlorophylls, it can be observed in figure 2 that the ratio *Chl a/Chl b*

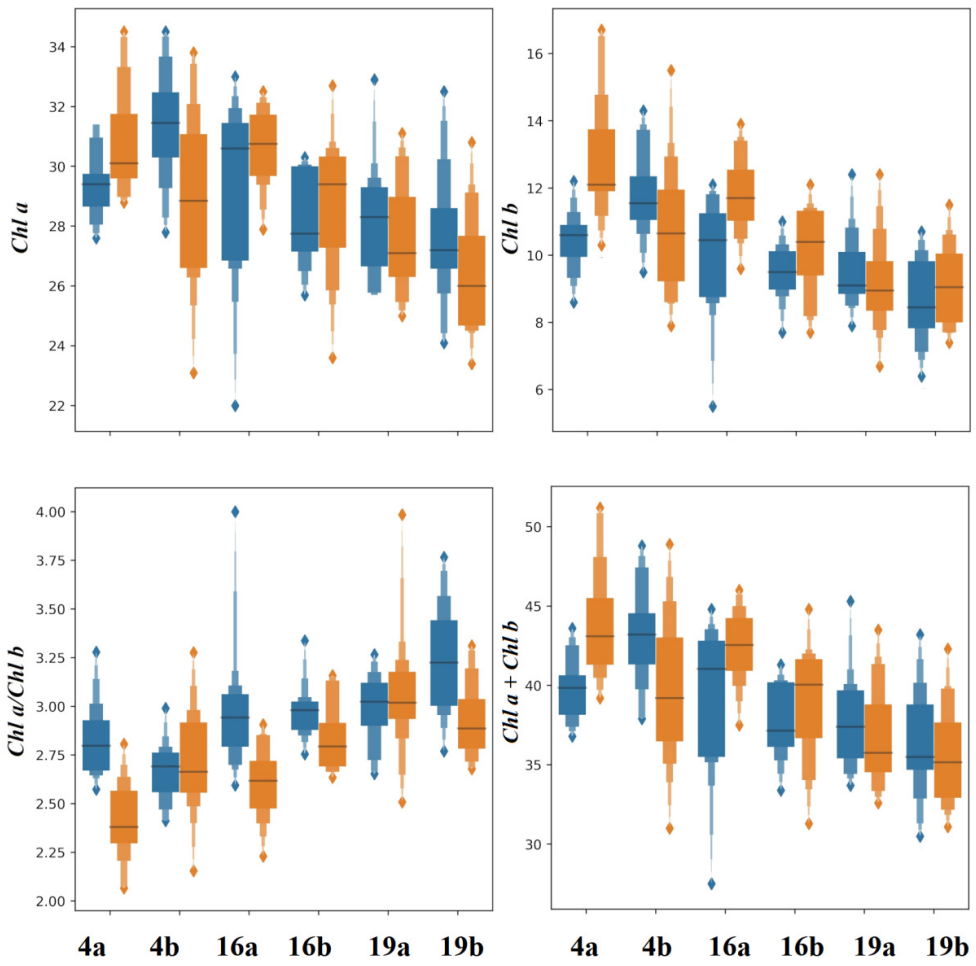

Figure 2 - Chlorophyll measurements for the six studied plots at first acquisition date (blue) and second (brown).

had a systematic decrease from the first observing date to the second, in agreement with what was reported by BIELCZYNSKI et al. (2017), considering also that solar illumination in the first date (December 16) was more intense than in second data (February 27); however, total levels *Chl a* + *Chl b* did not follow this trend, even if, in general, chlorophyll levels tend to rise toward more advance phases of the phenological cycle. Parcels 16a (first acquisition), 16b (first), 19a (first and second) and 19b (first) presented medians of about 3.0, an indication of plants under high luminosity conditions what is confirmed by the open landscape at the site. Parcels 19a and 19b presented vines with smaller, open canopy, being noted that 19a is located at the highest elevation, with exposed, rocky soils and little vegetation between rows. The highest *Chl a* and *Chl b* concentrations were found at parcels 4a (second), 4b (first) and 16a (first).

Correlational spectral analysis

The correlograms for the reflectance spectra and the HVIs at both acquisition dates are presented at figure 3. Graphs a) and b) suggest a decrease of the correlation between the more distant wavelengths, and higher correlations between neighboring wavelengths, at the NIR spectral range. Graph b) presents positive and negative correlations between indices. Changes in leaf characteristics along the cycle lead to a decrease in water content, with a reduction in photosynthetic activity and changes in leaf colors, effects due to senescence which affects the spectrum (BOYER et al., 1988).

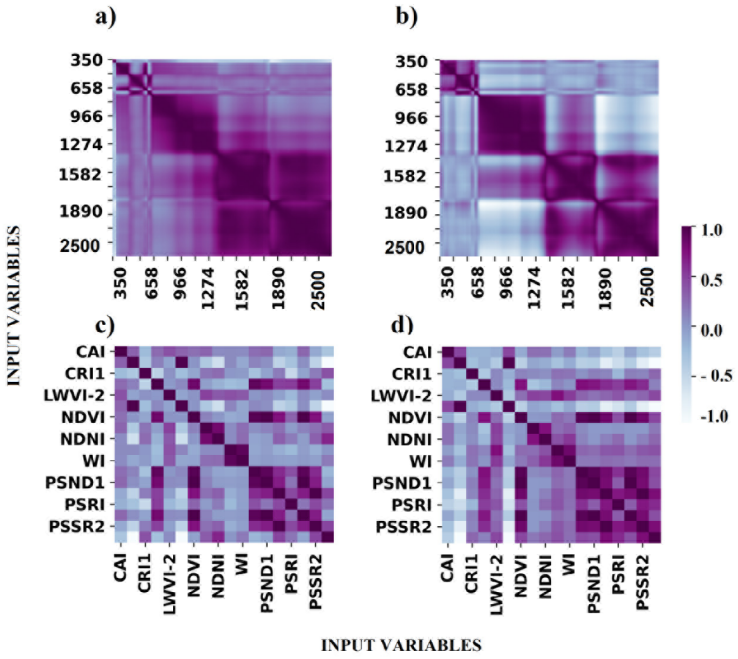

Figure 3 - Spearman Correlation Coefficient Rank (r), applied to wavelengths (350nm - 2500nm) of hyperspectral leaf reflectance and to HVIs, for acquisition dates: Dec. 16, 2017 (a and c) and Feb. 27, 2018 (b and d).

Prediction of chlorophyll parameters

The metrics which qualify the chlorophyll predictions from the PLSR model, for the three input sets and two acquisition dates, with a total of twenty-four results, are presented in table 1. As an average, each sub-model had used from two to eight Principal Components (PC) as input data for the prediction. The estimate model for *Chl b* on the second date was the data set reduced by PCA with the largest number of wavelengths (892) in the analysis; conversely, the models for *Chl a* and *Chl a/Chl b* had the smaller number of wavelengths used on the PCs, being nine and eight, respectively.

As the model's difficulty to predict chlorophyll levels increased, so increased the number of input variables; with the cycle's progress toward senescence, leaves have smaller photosynthetic activity, and leaf chlorophyll content gets smaller leading to color change to yellow-orange shades (BOYER et al., 1988; JENSEN, 2006). It was noted that those models using to chlorophyll parameters with larger number of input PCs presented a better performance as expressed by prediction metrics.

The estimate of *Chl b* from the Reflectance data set was the one which obtained the largest R^2 (0.622) at the second acquisition date. Predictions using only the HVI data set presented lower performances, with $R^2 < 0.468$. However, using the whole input data set, Reflectance + HVI, the PLSR model performed well for *Chl a*, *Chl b* and *Chl a+ Chl b*, with R^2 values above 0.60 for the second acquisition date. Metrics expressed by R^2 with cross validation (R^2(CV)) were smaller than 0.462, suggesting insufficient entries to model validation. Finally, *Chl a/Chl b* obtained the smallest RMSE at both dates, between 0.183 and 0.250, and similar values for RMSE (CV).

Going now to the Random Forest Regressor (RFR) model, R^2 values larger than 0.874 were obtained, regardless of the input set (Table 2). Using the Index input dataset, parameter *Chl a/Chl b)* got RMSE values between 0.085 and 0.097 for the two measuring dates, these being the smaller errors among all measured parameters. These results indicated that the RFR model presented higher predictive accuracy that the PLSR model. In general, the predictive metrics with cross validation presented values which were smaller than the training ones, with R^2(CV) around 0.113 and 0.391 and RMSE(CV) from 0.228 to 3.579. It is to be noted that this poorer performance for PLSR was

Table 1 - Metrics used for the predictions conducted with the PLSR model for the three input databases and on the two acquisition dates. The parameters used in the analyses are principal components number (PC), Wavelengths numbers (WN), coefficient of determination (R^2), coefficient of determination with cross validation (R^2 (CV)), root mean square error (RMSE), root mean square error with cross validation (RMSE (CV)).

	Parameters	------PC------		------WN------		-------R^2-------		----R^2(CV)----		----RMSE----		-RMSE(CV)-	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Reflectance	*Chl a*	3	5	9	141	0.294	0.551	0.214	0.415	1.995	1.748	2.104	1.994
	Chl b	8	8	332	892	0.585	0.622	0.359	0.432	1.025	1.254	1.274	1.537
	Chl (a+b)	3	6	391	360	0.380	0.579	0.296	0.462	3.073	2.952	3.274	3.336
	Chl (a/b)	7	4	8	157	0.576	0.528	0.471	0.391	0.183	0.214	0.205	0.243
HVI	*Chl a*	2	5	6	7	0.286	0.468	0.228	0.336	2.006	1.902	2.086	2.125
	Chl b	3	4	5	5	0.404	0.411	0.317	0.274	1.228	1.565	1.315	1.738
	Chl (a+b)	3	5	4	6	0.345	0.446	0.278	0.311	3.157	3.385	3.316	3.775
	Chl (a/b)	3	4	6	9	0.365	0.352	0.204	0.115	0.225	0.250	0.251	0.293
Reflect +HVI	*Chl a*	2	8	273	120	0.293	0.639	0.239	0.383	1.996	1.567	2.070	2.047
	Chl b	2	6	103	203	0.409	0.630	0.354	0.399	1.224	1.241	1.248	1.581
	Chl (a+b)	2	8	9	121	0.333	0.659	0.292	0.380	3.188	2.656	3.283	3.581
	Chl (a/b)	3	4	254	271	0.415	0.500	0.312	0.369	0.215	0.220	0.234	0.247

not reported in other studies, as, for example, by LI et al. (2019), who arrived to R^2 values as high as 0.86, and by LIU et al. (2019), who reported even higher accuracies from PLSR application.

For the PLSR coefficients the impact of each wavelength in the estimates of chlorophyll parameters are presented in figure 4. Wavelengths selected by the model were concentrated at red and red-edge spectral regions, between 670nm to 760nm; however, information at blue and green regions (from 550nm to 520nm) was also important. Parameters *Chl* a and *Chl a/Chl b)* presented the lesser number of wavelengths in the spectral modeling, with the red and red-edge spectral regions presenting high sensitivity to chlorophyll content; in these spectral regions, reflectance depends linearly on leaf chlorophyll content (STEELE et al., 2008). For other parameters, the model used data from several spectral regions, from the visible to SWIR. A significant increase of wavelengths inserted at

Table 2 - Predictions performance metrics for the Random Forest Regressor model for the three input databases on the two acquisition dates. The parameters used in the analyses are coefficient of determination (R^2); coefficient of determination with cross validation (R^2 CV); root mean square error (RMSE); root mean square error with cross validation (RMSE (CV)).

	-------Parameters-------	----------R^2----------		------R^2(CV)------		----------RMSE----------		----RMSE(CV)----	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd
Reflectance	*Chl a*	0.874	0.893	0.113	0.242	0.844	0.853	2.230	2.271
	Chl b	0.897	0.901	0.290	0.337	0.510	0.643	1.341	1.661
	Chl (a+b)	0.877	0.897	0.293	0.288	1.312	1.458	3.403	3.836
	Chl (a/ b)	0.899	0.956	0.258	0.293	0.090	1.250	0.243	1.100
HVI	*Chl a*	0.876	0.909	0.157	0.365	0.836	0.786	2.180	2.078
	Chl b	0.911	0.905	0.391	0.355	0.475	0.629	1.242	1.638
	Chl (a+b)	0.894	0.915	0.282	0.387	1.269	1.327	3.307	3.561
	Chl (a/b)	0.910	0.902	0.307	0.258	0.085	0.097	0.235	0.268
Reflect+HVI	*Chl a*	0.899	0.905	0.231	0.353	0.793	0.803	2.082	2.097
	Chl b	0.909	0.908	0.375	0.373	0.480	0.617	1.258	1.614
	Chl (a+b)	0.898	0.913	0.304	0.380	1.247	1.338	3.255	3.579
	Chl (a/b)	0.906	0.911	0.342	0.339	0.087	0.093	0.228	0.253

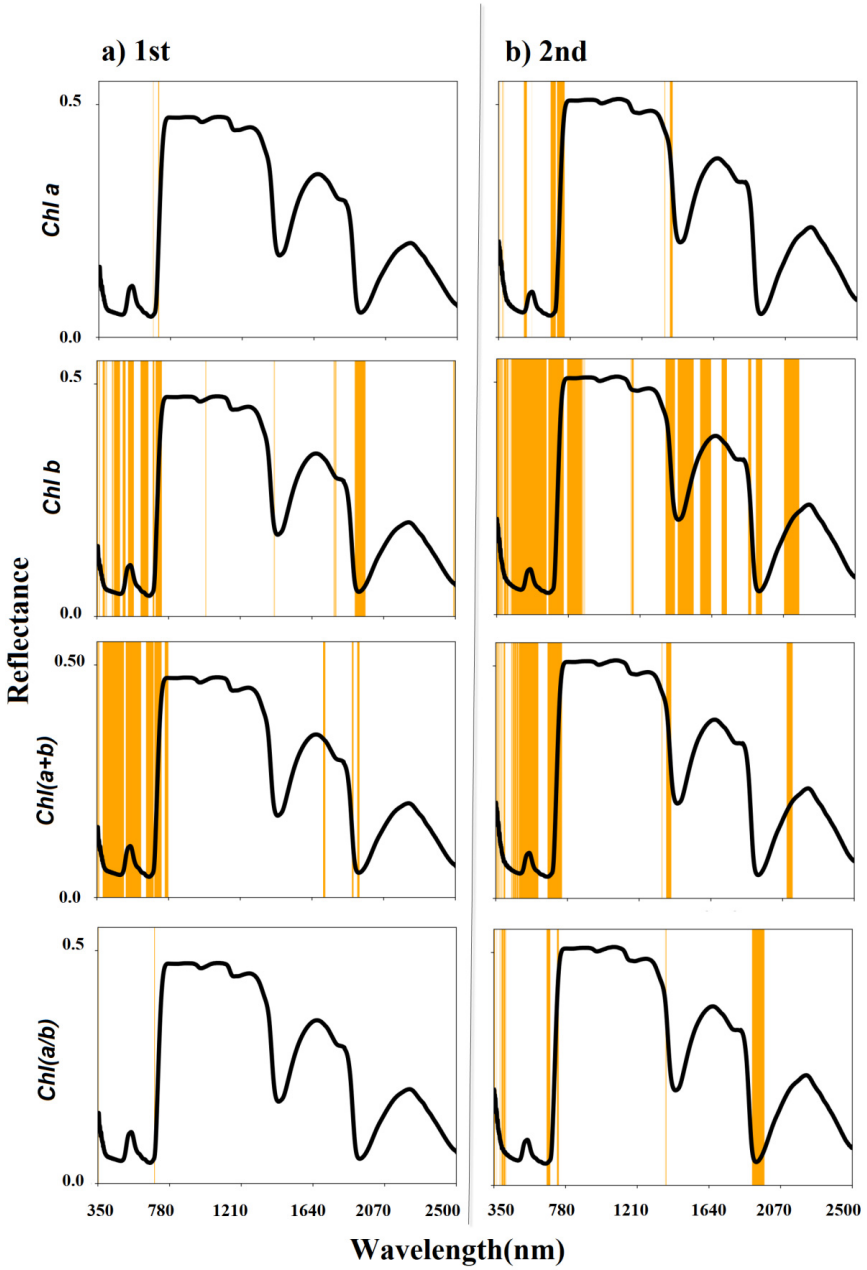

Figure 4 - Selected wavelengths (highlighted in yellow) with the best coefficients of the PSLR components, at the observed spectral domain (350nm-2500nm), in each prediction; column a corresponds to the first acquisition date, and column b to the second.

the Principal Components happened at the second acquisition date.

The importance of the Gini index to vegetation indices and wavelengths is expressed at figure 5. The index is used to select the variables more relevant to prediction in the RFR model, and presently, given the HVIs input data set, at first acquisition the most important HVIs were CAI, NDWI and PSR2. For data from the second acquisition date, indices CAI and PRI were the most important. For the Reflectance data set, the important wavelengths were concentrated in the blue to yellow regions (450nm to 580nm) for

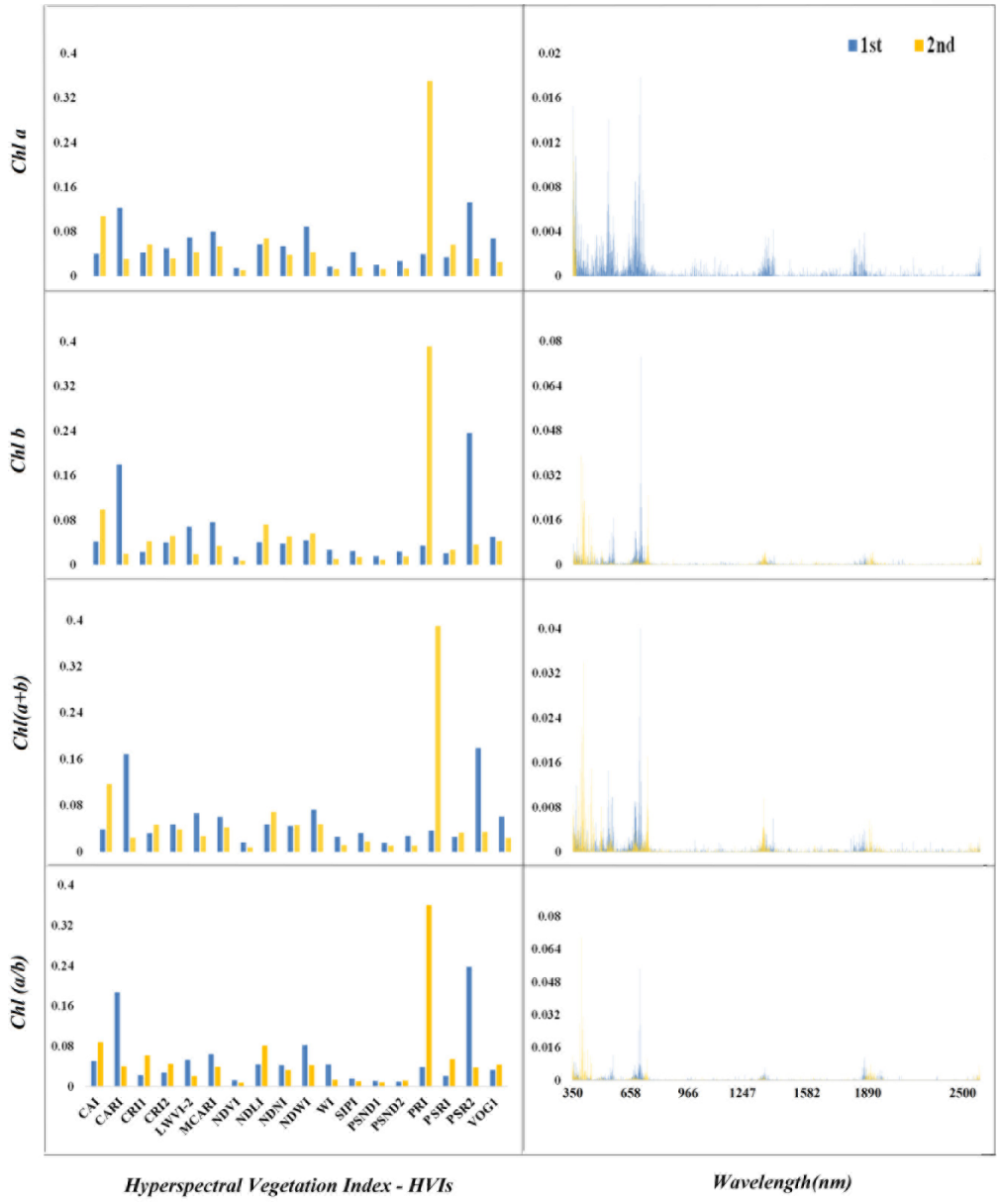

Figure 5 - Gini Importance to the HVIs(a) and wavelengths(b) in modeling of each chlorophyll parameter measured in the field.

first acquisition date, while for the second acquisition date, besides the blue region, information from near infrared, around 1350nm, was also important.

CONCLUSION

The results presented in this investigation suggested that chlorophyll content can be predicted from hyperspectral data. Collinearity between wavelengths presented some stability in specific spectral regions, later in the vegetative cycle, when the second data acquisition took place; at this stage, the observed indices tend to be more stable. Parameters *Chl a* and *Chl a/Chl b*) were in general the ones with less wavelengths used as input data sets in prediction models PLSR and RFR . The PRI index was the most

important vegetation index in all calibrations done using the RFR model, while PLSR model included a larger variable number at second acquisition. Compared to RFR, results from PLSR were relatively poor; the number of measurements was apparently adequate for application of RFR, but possibly results from PLSR would improve with more observations.

Forthcoming studies may focus on deeper investigations chlorophyll detection and content prediction, correlating spectroradiometric data with data from remote sensors at high resolutions, both in space and time. Besides, data on physiological parameters bring more information of environmental effects on leaf characteristics, and extending the analysis presently reported to other grape varieties will help to a better understanding on the spectral behavior due to genetic factors.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES**)** – Finance code 001, grants 88882.48939/2019-1 and 88887.488339/2020-00. The authors are grateful to Luiz Argenta Winery for the kind reception and full support during the whole investigation.

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 the final version.

REFERENCES

ABBASI, M. et al. Optimal spectral wavelengths for discriminating orchard species using multivariate statistical techniques. **Remote Sensing**, v.12, n.1, 2020. Available from: <https://doi.org/10.3390/rs12010063>. Accessed: Mar. 3, 2018. doi: 10.3390/rs12010063.

ARGENTA, G. et al. Leaf relative chlorophyll content as an indicator parameter to predict nitrogen fertilization in maize. **Ciência Rural**, v.34, n.5, 2004. Available from: <https//doi.org/10.1590/S0103-84782004000500009>. Accessed: Mar. 3, 2018. doi: 10.1590/S0103-84782004000500009.

BENERAGAMA, C. K.; GOTO, K. Chlorophyll *a*: *b* ratio increases under low-light in 'shade-tolerant' *Euglena gracilis*. **Tropical Agricultural Research**, v.22, n.1,p.12-25, 2011. Available from: <https://doi.org/10.4038/tar.v22i1.2666>. Accessed: Mar. 3, 2018. doi:10.4038/tar.v22i1.2666.

BERGSTRÄSSER, S. et al. HyperART: Non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging. **Plant Methods**, v.11, n.1, p.1–17, 2015. Available from: <https://doi.org/10.1186/s13007-015-0043-0>. Accessed: Mar. 3, 2018. doi: 10.1186/s13007-015-0043-0.

BIELCZYNSKI, L. W. et al. Leaf and plant age affects photosynthetic performance and photoprotective capacity. **Plant Physiology**, v.175, n.4, p.1634-1648, 2017. Available from: <https//doi.org/10.1104/pp.17.00904>. Accessed: Dec. 19, 2021. doi: 10.1104/pp.17.00904.

BLACKBURN, G. A. Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. **Remote Sensing of Environment**, v.66, n.3, p.273–285, 1998. Available from: <https://doi.org/10.1016/S00344257(98)00059-5>. Accessed: Dec. 19, 2021. doi: 10.1016/S0034-4257(98)00059-5.

BOYER, M. et al. Senescence and spectral reflectance in leaves of northern pin oak (*Quercus palustris* Muenchh.). **Remote Sensing of Environment**, v.25, n.1, p.71–87, 1988. Available from: <https://doi.org/10.1016/0034-4257(88)90042-9>. Accessed: Aug. 6, 2020. doi: 10.1016/0034-4257(88)90042-9.

BREIMAN, L. Random Forests. **Machine Learning**, v.45, p.5-32, 2001. Available from: <http://dx.doi.org/10.1023/A:1010933404324>. Accessed: Aug. 6, 2020. doi: 10.1023/A:1010933404324.

CHENG, J.-H.; SUN, D.-W. Partial Least Squares Regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle. **Food Engineering Reviews**, v. 9, n.1, p.36–49, 2017. Available from: <https://doi.org/10.1007/s12393-016-9147-1>. Accessed: Aug. 6, 2020. doi: 10.1007/s12393-016-9147-1.

EL-HENDAWY, S. et al. Performance of optimized hyperspectral reflectance indices and partial least squares regression for estimating the chlorophyll fluorescence and grain yield of wheat grown in simulated saline field conditions. **Plant Physiology and Biochemistry**, v.144, p.300–311, 2019. Available from: <https://doi.org/10.1016/j.plaphy.2019.10.006>. Accessed: Sept. 15, 2020. doi: 0.1016/j.plaphy.2019.10.006.

ETTABAA, K. B.; SALEM, M. B. Adaptive progressive band selection for dimensionality reduction in hyperspectral images. **Journal of the Indian Society of Remote Sensing**, v.46, n.2, p.157–167, 2017. Available from: <https://doi.org/10.1007/s12524-017-0691-9>. Accessed: Jun. 16, 2019. doi: 10.1007/s12524-017-0691-9.

FASSNACHT, F. E. et al. Non-destructive estimation of foliar carotenoid content of tree species using merged vegetation indices. **Journal of Plant Physiology**, v.176, p.210–217, 2015. Available from: <https://doi.org/10.1016/j.jplph.2014.11.003>. Accessed: Sept. 15, 2020. doi: 10.1016/j.jplph.2014.11.003.

GALVÃO, L. S et al. Discrimination of surface varieties in Southeast Brazil with EO-1 Hyperion data. **Remote Sensing of Environment**, v.94, n.4, p.523-534, 2005. Available from: <https://www.sciencedirect.com/science/article/pii/S0034425704003669>. Accessed: May, 18, 2021. doi: 10.1016/j.rse.2004.11.012

GAO, B.-C. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, v.58, n.3, p.257–266, 1996. Available

from: <https://doi.org/10.1016/j.agwat.2017.08.003>. Accessed: Dec. 19, 2021. doi: 10.1016/j.agwat.2017.08.003.

GITELSON, A. A. et al. Assessing carotenoid content in plant leaves with reflectance spectroscopy. **Photochemistry and Photobiology**, v.75, n.3, p.272, 2002. Available from: <https://doi.org/10.1562/0031-8655(2002)075<0272:accipl>2.0.co;2>. Accessed: Dec. 19, 2021. doi: 10.1562/0031-8655(2002)075<0272:accipl>2.0.co;2.

GITELSON, A. A. et al. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. **Photochemistry and Photobiology**, v.74, n.1, p.38-45, 2001. Available from: <https://doi.org/10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2>. Accessed: Dec. 19, 2021. doi: 10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2.

JENSEN, JR. **Remote Sensing of the Environment**: An Earth Resource Perspective. New York: Prentice Hall, 2006.

KAWAMURA, K. et al. Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. **Remote Sensing**, v.9, n.10, p.1081, 2017. Available from: <https://doi.org/10.3390/rs9101081>. Accessed: Jul. 25, 2021. doi: 10.3390/rs9101081.

LACAR, F. M. et al. Use of hyperspectral reflectance for discrimination between grape varieties. In: **IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium**. IEEE, 2002, p.2878–2880. Available from: <https://doi.org/10.1109/IGARSS.2001.978192>. Accessed: Jul. 05, 2018. doi: 10.1109/IGARSS.2001.978192.

LI, Y. et al. Spectroscopic determination of leaf chlorophyll content and color for genetic selection on *Sassafras tzumu*. **Plant Methods**, v.15, n.1, 2019. Available from: <https//doi.org/10.1186/s13007-019-0458-0>. Accessed: Jul. 16, 2021. doi: 10.1186/s13007-019-0458-0.

LIU, J. et al. Nondestructive detection of rape leaf chlorophyll based on Vis-NIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, 222, 117202, 2019. Available from: <https//doi.org/10.1016/j.saa.2019.117202>. Accessed: Jul. 16, 2021. doi: 10.1016/j.saa.2019.117202.

LOGGENBERG, K. et al. Modelling water stress in a Shiraz vineyard using hyperspectral imaging and machine learning. **Remote Sensing**, v.10, n.2, p.202, 2018. Available from: <https://doi.org/10.3390/rs10020202>. Accessed: Dec. 05, 2020. doi: 10.3390/rs10020202.

LORENZ, D. H et al. Growth stages of the grapevine: phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. *vinifer*a)—Codes and descriptions according to the extended BBCH scale. **Australian Journal of Grape and Wine Research**, v.1, n.2, p.100–103, 1995. Available from: <https://doi.org/10.1111/j.1755-0238.1995.tb00085.x>. Accessed: Dec. 17, 2018. doi: 10.1111/j.1755-0238.1995.tb00085.x.

MERZLYAK, M. N. et al. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. **Physiologia Plantarum**, v.106, n.1, p.135–141, 1999. Available from: <https://doi.org/10.1034/j.1399-3054.1999.106119.x>. Accessed: Feb. 17, 2017. 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**, v.80, n.January, p.26–37, 2019. Available from: <https://doi.org/10.1016/j.jag.2019.04.002>. Accessed: Feb. 10, 2021. doi: 10.1016/j.jag.2019.04.002.

MITRA, S. et al. Effect of vineyard soil variability on chlorophyll fluorescence, yield and quality of table grape as influenced by soil moisture, grown under double cropping system in protected condition. **PeerJ**, v.6, n.9, 2018. Available from: <https://doi.org/10.7717/peerj.5592>. Accessed: Feb. 10, 2021. doi: 10.7717/peerj.5592.

NAGLER, P. L et al. Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. **Remote Sensing of Environment**, v.87, n.2–3, p.310–325, 2003. Available from: <https://doi.org/10.1016/j.rse.2003.06.001>. Accessed: Feb. 10, 2019. doi: 10.1016/j.rse.2003.06.001.

NEMBRINI, S. et al. Supplementary material for "The revival of the Gini importance?". **Bioinformatics**, v.21, p.3711-3718, 2018. Available from: <https://doi.org/10.1093/bioinformatics/bty373>. Accessed: Feb. 10, 2019. doi: 10.1093/bioinformatics/bty373.

ORDÓÑEZ, C. et al. Determining optimum wavelengths for leaf water content estimation from reflectance: A distance correlation approach. **Chemometrics and Intelligent Laboratory Systems**, v.173, p.41–50, 2018. Available from: <https://doi.org/10.1016/j.chemolab.2017.12.001>. Accessed: Mar. 25, 2019. doi: 10.1016/j.chemolab.2017.12.001.

PALMER, D. S. et al. Random forest models to predict aqueous solubility. **Journal of Chemical Information and Modeling**, v.47, n.1, p.150-158, 2007. Available from: <https://pubs.acs.org/doi/10.1021/ci060164k>. Accessed: Mar. 25, 2019. doi: 10.1021/ci060164k.

PEÑUELAS, J. et al. Estimation of plant water concentration by the reflectance water index WI (R900/R970). **International Journal of Remote Sensing**, v.18, n.13, p.2869–2875, 1997. Available from: <https://doi.org/10.1080/014311697217396>. Accessed: Mar. 25, 2018. doi:10.1080/014311697217396.

PEÑUELAS, J. et al. Semi-empirical indices to assess carotenoids/chlorophyll *a* ratio from leaf spectral reflectance. **Photosynthetica**, v.2, p.221–230, 1995. Available from: <https://www.researchgate.net/publication/235645501>. Accessed: Mar. 25, 2018.

PEÑUELAS, J. et al. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. **New Phytologist**, v.131, n.3, p.291–296, 1995. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.1995.tb03064.x>. Accessed: Mar. 25, 2018. doi: 10.1111/j.1469-8137.1995.tb03064.x.

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**, v.42, n.15, p.5680–5697, 2021. Available from: <https://www.tandfonline.com/doi/full/10.1080/01431161.2021.1929542>. Accessed: Aug. 16, 2021. doi: 10.1080/01431161.2021.1929542.

POWER, A. et al. From the laboratory to the vineyard - Evolution of the measurement of grape composition using NIR spectroscopy towards high-throughput analysis. **High Throughput**, v.8, n.4, p.21, 2019. Available from: <https://doi.org/10.3390/ht8040021>. Accessed: Aug. 16, 2021. doi: 10.3390/ht8040021.

ROUSE, J. W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **Proceedings of the Third Earth Resources Technology Satellite- 1 Symposium**, p.301–317, 1974. Available from: <ntrs.nasa.gov/citations/19740022614>. Accessed: Aug. 16, 2021.

SCHLICHTING, A. F. et al. Efficiency of portable chlorophyll meters in assessing the nutritional status of wheat plants. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.19, n.12, p.1148–1151, 2015. Available from: <https://doi.org/10.1590/1807-1929/agriambi.v19n12p1148-1151>. Accessed: Mar. 16, 2019. doi: 10.1590/1807-1929/agriambi.v19n12p1148-1151.

SERRANO, L. et al. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. **Remote Sensing of Environment**, v.81, n.2–3, p.355–364, 2002. Available from: <https://doi.org/10.1016/S0034-4257(02)00011-1>. Accessed: Mar. 16, 2019. doi: 10.1016/S0034-4257(02)00011-1.

SESTAK, Z. Changes in the chlorophyll content as related to photosynthetic activity and age of leaves. **Photochemistry and Photobiology**, v.2, p.101-110, 1963. Available from: <onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-1097.1963.tb08207.x>. Accessed: Mar. 16, 2019.

STEELE, M. et al. Nondestructive estimation of leaf chlorophyll content in grapes. **American Journal of Enology and Viticulture**, v.59, n.3, p.299–305, 2008. Available from: <ajevonline.org/content/59/3/299>. Accessed: Mar. 16, 2019. doi: 10.5344/ajev.2008.59.3.299.

TANAKA, R.; TANAKA, A. Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. **Biochimica et Biophysica Acta – Bioenergetics**, v.1807, n.8, p.968-976, 2011. Available from: <pubmed.ncbi.nlm.nih.gov/21216224>. Accessed: Mar. 16, 2019. doi: 10.1016/j.bbabio.2011.01.002.

THUM, AB 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://www.tandfonline.com/doi/abs/10.1080/01431161.2020.1798547>. Accessed: Jan. 16, 2021. doi: 10.1080/01431161.2020.1798547.

VISCARRA ROSSEL, R. A et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, v.131, n.1–2, p.59–75, 2006. Available from: <https://doi.org/10.1016/j.geoderma.2005.03.007>. Accessed: Mar. 16, 2019. doi: 10.1016/j.geoderma.2005.03.007.

VOGELMANN, J. E. et al. Red edge spectral measurements from sugar maple leaves. **International Journal of Remote Sensing**, v.14, n.8, p.1563–1575, 2007. Available from: <http://dx.doi.org/10.1080/01431169308953986>. Accessed: Mar. 18, 2019. doi: 10.1080/01431169308953986.

YANG, X.-H. et al. A modified chlorophyll absorption continuum index for chlorophyll estimation. **Journal of Zhejiang University - Science A**, v.7, n.12, p.2002–2006, 2006. Available from: <https://doi.org/10.1631/jzus.2006.A2002>. Accessed: Mar. 18, 2019. doi: 10.1631/jzus.2006.A2002.

ZHANG, N. et al. Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging. **Food Chemistry**, v.237, p.811–817, 2017. Available from: <https://doi.org/10.1016/j.foodchem.2017.06.007>. Accessed: Jan. 6, 2019. doi: 10.1016/j.foodchem.2017.06.007.

ZHANG, C. et al. Relationship between hyperspectral measurements and mangrove leaf nitrogen concentrations. **Remote Sensing**, v.5, n.2, p.891–908, 2013. Available from: <https://doi.org/10.3390/rs5020891>. Accessed: Oct. 6, 2019. doi: 10.3390/rs5020891.

ZHAO, J. et al. Hyperspectral measurements of severity of stripe rust on individual wheat leaves. **European Journal of Plant Pathology**, v.139, n.2, p.401–411, 2014. <https://doi.org/10.1007/s10658-014-0397-6>. Accessed: Jan. 18, 2018. doi: 10.1007/s10658-014-0397-6.