BIOSYSTEMS ENGINEERING 229 (2023) 167–178

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

NIR attribute selection for the development of vineyard water status predictive models

Miguel Marañón [a], Juan Fernández-Novales [b,c], Javier Tardaguila [b,c], Salvador Gutiérrez [d], Maria P. Diago [b,c,*]

[a] University of La Rioja, Department of Mathematics and Computer Science, Madre de Dios 53, 26007, Logroño, La Rioja, Spain

[b] University of La Rioja, Department of Agriculture and Food Science, Madre de Dios 53, 26007, Logroño, La Rioja, Spain

[c] Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, Gobierno de La Rioja), Finca La Grajera, Ctra. Burgos Km. 6, 26007 Logroño, Spain

[d] Department of Computer Science and Artificial Intelligence, University of Granada, 18071, Granada, Spain

ARTICLE INFO

Article history:
Received 16 November 2022
Received in revised form
26 February 2023
Accepted 2 April 2023
Published online 14 April 2023

Keywords:
Grapevine
Stem water potential
Variable Importance in Projection scores
Manual wavelength selection
Interval Partial Least Squares

Near-Infrared spectroscopy (NIR) returns full spectra in the region between 750 and 2500 nm. Although a full spectrum provides extremely informative data, sometimes this enormous amount of detail is redundant and does not bring any additional information. In this work, different attribute selection methods for the development of vineyard water status predictive models are presented. Spectra from grapevine leaves were collected on-the-go (from a moving vehicle) along nine dates during the 2015 season in a commercial vineyard using a NIR spectrometer (1200–2100 nm). Contemporarily, the stem water potential (Ψ_{stem}) was also measured in the monitored vines. A manual selection, based on Variable Importance in Projection scores (VIP scores) to choose the spectrum intervals including the most important wavelengths (interval selection), the locally most important wavelengths in the spectrum (peak selection), as well as the Interval Partial Least Squares (IPLS) were tested as attribute selection methods. The results obtained for the estimation of Ψ_{stem} using the whole spectrum ($R^2_P = 0.84$, RMSEP = 0.167 MPa) were comparable to those yielded by the three attribute selection methods: the interval selection method ($R^2_P = 0.80$, RMSEP = 0.186 MPa), the peak selection method ($R^2_P = 0.77$, RMSEP = 0.201 MPa) and the IPLS ($R^2_P \sim 0.62$–0.79, RMSEP ~ 0.186–0.252 MPa). The highest simplification was provided by two IPLS models with three wavelengths and bandwidths of 20 and 4 nm that yielded $R^2_P \sim 0.78$ and RMSEP~ 0.190 MPa. These results corroborate the suitability of a highly reduced selection of NIR wavelengths for the prediction of grapevine water status, and its utility to develop simpler multispectral devices for vineyard water status estimation.

* *Corresponding author*. University of La Rioja, Department of Agriculture and Food Science, Madre de Dios 53, 26007, Logroño, La Rioja, Spain.

E-mail address: maria-paz.diago@unirioja.es (M.P. Diago).

https://doi.org/10.1016/j.biosystemseng.2023.04.001

1. Introduction

Water scarcity is being aggravated in recent years by climate change. In viticulture, the development of new irrigation scheduling strategies that optimise water resources' usage are aimed to be driven by plant stress sensing, as decision making regarding watering is based on plant response, considering the grapevine as a biosensor, which integrates soil and atmospheric water demands (Jones, 2004). Among the methods based on measuring the plant water status, plant water potential, either leaf (Ψ1eaf), stem water potential (Ψstem), or pre-dawn water potential (ΨPD) is one of the most widely accepted indicators due to its reliability and effectiveness (Choné et al., 2001; Rienth & Scholasch, 2019). This method is destructive, discontinuous, time-intensive and provides information of a reduced number of samples; hence, its ability to disclose the variability in plant water status within a vineyard plot is limited. To overcome this, in the last decades, non-destructive technologies, such as thermography (Araújo-Paredes et al., 2022; Gutiérrez et al., 2021; Zhou et al., 2022) or near infrared spectroscopy (Diago et al., 2018; Fernández-Novales et al., 2018; Pampuri et al., 2021) have been adapted to estimate plant water status in grapevines.

The NIR region is the part of the electromagnetic spectrum between 750 and 2500 nm and it is based on the absorption of energy from molecules or chemical constituents related with the overtones and combinations of these fundamental vibrations due to the stretching and bending of N–H, O–H and C–H groups (Nicolai et al., 2007). When incident light reaches the surface of the sample it interacts with the sample at a molecular level according to the sample's chemical composition and internal organic matter, and it is basically reflected, transmitted or absorbed (Pasquini, 2003). The water molecule, which is a predominant component of leaves, can partially or fully absorb the light at given wavelengths of 760, 971 or 1450 nm (O–H overtones) and at a combination band of 1940 nm (Nicolai et al., 2007) within the NIR spectral range.

In the last years, a growing interest has risen toward the development of portable Vis/NIR systems for the monitoring of water status directly in the vineyard (De Bei et al., 2011; Giovenzana et al., 2014, 2018; Guidetti et al., 2010; Tardaguila et al., 2017) or in greenhouse-located potted plants (Pampuri et al., 2021; Rapaport et al., 2015). In terms of Ψ1eaf and Ψstem, the outcomes, expressed as determination coefficients of cross validation (R2CV) or prediction (R2P) ranged between 0.59 and 0.87 in these works. Within the context of precision viticulture, several studies (Diago et al., 2018; Fernández-Novales et al., 2018, 2019; Gutiérrez et al., 2019) have recently reported the capabilities of NIR spectrometers and hyperspectral sensors installed in a mobile terrestrial platform to on-the-go monitor and map grape composition parameters and plant water status along the ripening period with reliable and encouraging results. Moreover, on plant water status specifically, the works of Pôças et al., 2015, 2017; Pôças, Tosin, Gonçalves, & Cunha, 2020 and Tosin et al. (2020, Tosin et al., 2021, 2022) are remarkable in the use of hyperspectral data and reflectance vegetation indices to model and predict grapevine ΨPD in Portugal vineyards, and the work of Wei et al. (2021) to predict Ψstem in vineyards of New Zealand using different data processing pipelines. As a matter of fact, Tosin et al. (2022) reviewed the existing literature covering the prediction of grapevine water potential (regardless the specific type) from Vis/NIR spectroscopy using either hand-held sensors or spectrometers covering different spectral ranges, and used either statically or from moving platforms, both aerial (e.g. UAVs) and terrestrial ones. While all these reviewed works corroborate that NIR spectroscopy has a great potential to predict grapevine water status appropriately, its practical implementation in commercial operations requires a higher degree of automation and robustness.

Further steps towards the automation of this spectral approach involve the miniaturisation and cost reduction of spectral devices as well as the simplification of spectral processing. In this regard, chemometric techniques are often used to select a few relevant variables that represent the most important information contained in the full spectral region. This selection eliminates variables containing mostly redundant information and spectral noise, and reduces the cost of the potential miniaturised devices built only with the selected wavelengths (Xiabobo et al., 2010). Different spectral selection strategies, such as manual approaches based on VIP scores and interval selection methods based on IPLS (Nørgaard et al., 2000) have been proved in this work to recognise the most significant wavelengths for grapevine water status estimation and their subsequent comparison with the PLS models yielded on the full spectrum. All of these selection methods have been previously used in a wide variety of research related to agriculture and food. The VIP scores based strategy has been considered in the determination of winter oilseed rape yield (Lantao et al., 2017), chilling injury in green bell peppers (Babellahi et al., 2020), leaf nitrogen concentration in winter wheat (Li et al., 2020) and internal quality indices in oriental melon (Cho et al., 2022). Regarding the IPLS method, it has been employed in wavelength selection to the assessment of soluble solids content in mandarin (Sun et al., 2009), grape juice composition (Wu et al., 2010), flavonoids in ginkgo leaves (Zou, Huang, et al., 2012), pigments in leaves of cucumber (Shi et al., 2011; Zou, Chen, et al., 2012), wheat moisture (Dong et al., 2013), blueberries' composition (Leiva-Valenzuela et al., 2014), soil specific surface area (Knadel et al., 2018), leaf primordia of potato tubers (Rady et al., 2018), neutral detergent fibre content in corn stover (Pan et al., 2020), lately food colorants (Al-Degs et al., 2021) and egg origin (Liu et al., 2022). Giovenzana et al. (2018) and Zhang et al. (2012) have addressed the potential of the variable selection methods to determine the leaf water status in ornamental plants and grapevines using Vis/NIR spectroscopy with promising results for the design of a simplified handheld device that supports 66 small-scale growers and optimises irrigation scheduling. More recently, Pampuri et al. (2021) built a PLS-model with Vis/NIR (350–2500 nm) spectra taken statically on leaves of Pinot Blanc (in potted vines) and ΨPD reaching R2P = 0.70 and RMSEP = 0.056 MPa, and a reduced model with only three relevant bands (530, 700 and 1400, with a 20 nm bandwidth), achieving R2P = 0.60 and similar RMSEP values. These authors carried out a SWOT analysis of the possibilities of Vis/NIR spectroscopy-derived models to assess ΨPD using a reduced set of relevant bands. As the main opportunities, these authors identified the monitoring of crop water status in a semicontinuous way, and the improvement of irrigation management. As one of the major weaknesses, the difficulty to positioning infield stand-alone sensors was highlighted. To overcome this last difficulty, the goal of the present study is two-fold. First to test the capability of different attribute selection methods to identify the crucial wavelengths able to predict the grapevine water status; second to compare the predictive capacity of the different variable selection methods among them and with respect to the model built using the full spectrum, employing a proximal NIR sensor from a motorised platform in a commercial vineyard.

2. Materials and methods

2.1. Experimental layout

The field experiment was performed in a commercial vineyard placed in Tudelilla, La Rioja, Spain (Lat. 42°18′18.26″ Long. −2°7′14.15″, Alt. 515 m) along eight weeks from 23rd July 2015 to 18th September 2015. Grapevines of (*Vitis vinifera* L.) Tempranillo (Clone 776) grafted on rootstock R-110 were planted in 2002. The vines were trained to a vertically shoot-positioned trellis system (VSP) on a double-cordon Royat. Vine spacing was 2.60 m between rows and 1.20 m in the row in a north-south orientation. The total rainfall from April to October 2015 was 245.8 mm at the study site, and the average temperature was 17.8 °C during this period.

Three different water treatments were deployed in a randomised block design (Hinkelmann & Kempthorne, 2007) with four blocks so as to induce a substantial variability of grapevine water status within the plot. The treatments were.

- T0 (low water stress, full irrigation). Two water pipelines irrigating 6 L h^{-1} were installed. The total amount of delivered water during the studied period was 406.5 mm H2O m^{-2}.
- T1 (moderate water stress and irrigation). A single water pipeline, which irrigated half the amount in T0, was installed and provided a total of 221.7 mm H2O m-2 during the studied period.
- T2 (high water stress, no irrigation). The plants were not irrigated in any moment during the experiment.

For each treatment, a completely randomised block design with four replications (one per block) were set up, coming to a total of 12 treatment replicates in three different vine rows (see Fig. 1 in Diago et al. (2018)). The experiment was conducted with 25 plants in each of three adjacent rows. However, only the middle row was used for measurement, and out of the 25 plants in the middle row, only the 15 middle plants were considered. These 15 plants were divided into sub replicate units consisting of groups of five vines (Fig. 1 in Diago et al. (2018)). The first and last five vines in each row were excluded, as well as the adjacent rows to avoid any edge effects. The plants subjected to water regimes T0 and T1 were watered for a total of 2 h per day, with four equal watering sessions of 30 min each, evenly spaced throughout the day.

2.2. Spectral measurements

Spectra were collected on-the-go under field conditions using a Polytec PSS 2120 spectrometer (Polytec GmbH, Waldbronn, Germany) working in the 1100–2100 nm spectral range, at a 2 nm resolution (501 datapoints per spectrum) and with an acquisition rate of 24 Hz. The spectrometer was a NIR optical device, based on a polychromator as reflection light source selector, and Indium Gallium Arsenide (InGaAs) diode array detectors. This device is made up of a sensor head (based on an integrated 20 W tungsten halogen lamp) for capturing light connected to a processing unit by optical fibre. The system was assembled in the front part of an all-terrain-vehicle (ATV) (Trail Boss 330, Polaris Industries, Minnesota, USA), pointing to the left and able to make spectral acquisitions controlled by a physical trigger while the ATV was in motion. In order to cover the mid-upper part of the vineyard's canopy, the sensor head was installed at a height of 0.95 m from the ground, so that it was located above the fruiting zone. Measurements were performed contactless, keeping a distance from 0.25 to 0.35 m between the canopy and the sensor head. The diameter of the measurement window was 19 mm, and it remained constant since it was not influenced by the distance to the canopy. All the measurements were conducted from the ATV moving at an average speed of 5 km per hour from the east side of the canopy. The collected spectra also contained information from non-grapevine objects such as gaps, wood, and metal. Therefore, a filtering step was necessary to select only those spectra that corresponded to grapevine canopy. To achieve this, a well-captured leaf signature was taken prior to each measurement day and used as a reference for comparison during the filtering process. The task of filtering was performed using the “Spectra Comparison & Filtering” tool, which is part of the SL Utilities software (version 3.1, Polytec GmbH, Waldbronn, Germany). The spectral measurements were conducted at solar noon (between 14:00 and 15:00 GMT+1) on nine different dates (23/07, 28/07, 06/08, 12/08, 19/08, 26/08, 07/09, 11/09 and 18/09) during season 2015.

2.3. Measurement of stem water potential

The gold reference indicator of the plant water status was the midday stem water potential (Ψ_{stem}). Measurements were taken at solar noon (14:00–15:00 local time), just a short time after spectral acquisition. For each treatment replicate, the first and last five plants were discarded, so as to avoid edge effects. The other 15 middle plants were considered of interest, and it was in these vines where spectral measurements, as well as Ψ_{stem} readings, were taken. From each group of five plants, one of them was marked randomly and then one leaf from the mid-upper part of its canopy, on the west side, was selected for the determination of Ψ_{stem}. Measurements of Ψ_{stem} were conducted with a Schölander pressure bomb (Model 600, PMS Instruments Co., Albany, USA) on leaves that had been previously subjected to dark adaptation during 1 h, covered with aluminium foil. Each day, 36 data for stem water potential were recorded, making a total of 324 data over the whole experiment.

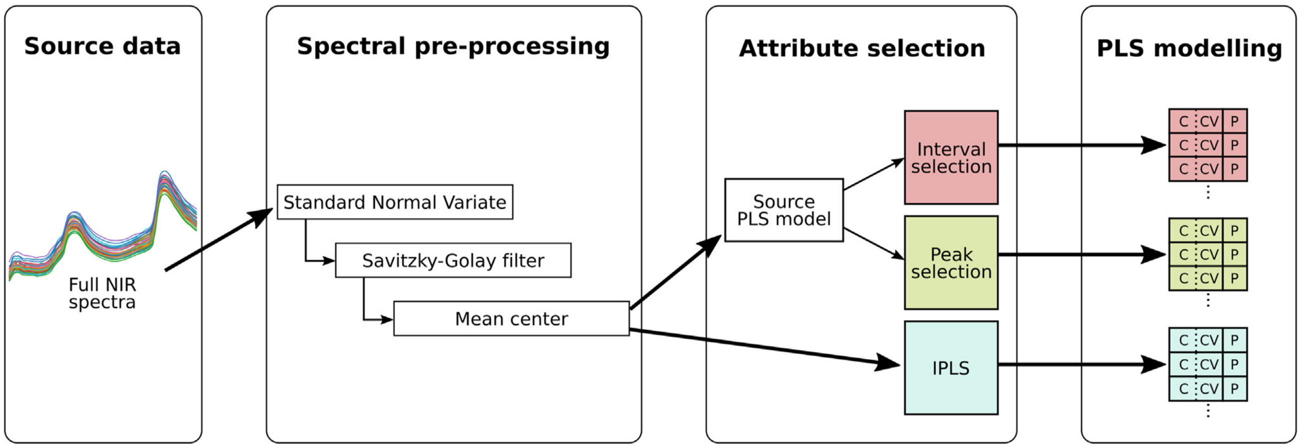

Fig. 1 – Diagram of the main steps from spectra acquisition to modelling.

2.4. *Data analysis*

Data processing and statistical analysis were performed with MATLAB version 7.6.0 (R2008a) and PLS_Toolbox version 5.5.1 (Eigenvector Research Inc., Manson, WA). Spectral pre-processing is useful for counterbalancing baseline offset and bias. Many combinations of spectral pre-processing filters were tested to obtain the models with the best calibration statistics. These filters involved the use of Standard Normal Variate (SNV) to correct scattering data (Barnes et al., 1898; Dhanoa et al., 1995) and the application of the Savitzky–Golay smoothing and derivative processes, which were used to emphasise small bands and to resolve overlapping peaks (Savitzky & Golay, 1964). Different values for the window size (7 and 15 points) and the degree of the derivative (first and second) were set. Moreover, the spectra were properly normalised by mean centring after pre-processing. With the aim of building robust models capable of predicting totally unknown samples, the original dataset was split up into two independent datasets: a calibration one (comprising 80% of all data) and an external validation set (comprising the remaining 20% of original data). The calibration dataset was used to train and to perform an internal cross-validation of the model, while the external validation set was only utilised for prediction purposes, using the calibration models. The ten-fold Venetian blind method was selected for the internal cross-validation. In an n-fold Venetian blind cross-validation, each fold i is built taking samples from the dataset of an n-multiple position until the end of the dataset (i.e., samples $i, i+n, i+2n, i+3n$). Once the folds are built, a traditional n-fold cross-validation is carried out, in which n models are trained with $n-1$ folds and tested with the remaining fold, rotating the latter until all of them have been used. The average performance of the n models is finally computed. For each model, the optimal number of latent variables (LVs) was selected as that yielding the minimum root mean square error of cross-validation (RMSECV). To evaluate the quality of the best models that were obtained, the coefficient of determination and the root mean square error of calibration (R^2_C, RMSEC), cross-validation (R^2_{CV}, RMSECV) and prediction (R^2_P, RMSEP) were calculated.

The expected number of samples was reduced due to diverse reasons, such as anomalous spectra, inconsistencies in the reference method and spectral outliers. So, after this data depuration, the final data set from the nine dates of the experiment comprised 288 samples. Since our aim was to select those features (wavelengths) in the data spectra that were most useful and relevant to the estimation of Ψ_{stem}, three attribute selection methods were considered: two manual selection methods—interval selection and peak selection—and IPLS. The manual selection methods were based on the calculation of the VIP scores, whereas the ones selected by IPLS were automatically selected by the algorithm. In addition, a full spectrum model based on partial least squares (PLS) regression was created to compare it against the models built using these attribute selection methods. PLS is an accurate, robust and reliable chemometric method which has been widely used to analyse spectral information as it is able to cope with huge amounts of data (Wold et al., 2001).

The referred-to attribute selection methods are described below.

- *Interval selection:* this method is about estimating the significance of each wavelength within the PLS model, according to the VIP scores function (where VIP stands for Variable Importance in Projection) and then choosing intervals whose wavelengths are relevant above a certain threshold. The VIP was calculated for each wavelength, as recommended by Wold et al. (2001) and Lin et al. (2013) to verify the wavelengths with greater impact on the predictions. The larger the VIP value, the more important the predictor is for estimating the response variable. According to Ericksson et al. (2001), variables with VIP greater than 1.0 are highly influential, so this was the threshold value above which a wavelength was considered important.
- *Peak selection:* this method is similar to the interval selection method, but in this case those wavelengths representing a local maximum for the VIP scores function is chosen, together with the contiguous ones. That is to say, only the intervals around the most relevant wavelengths are taken, with a determined bandwidth.

Five different bandwidths were considered: 4 nm, 8 nm, 12 nm, 16 nm and 20 nm.

- *IPLS:* The Interval PLS (IPLS) method (Leardi & Nørgaard, 2004; Nørgaard et al., 2000) is used to develop local PLS models on equidistant subintervals of the full spectrum region, selecting a subset of variables which will give a good prediction by doing a sequential, exhaustive search for the best variable or combination of variables. For the analysis of the gathered data, the MATLAB IPLS Variable Selection Interface was used in forward mode with eight latent variables at most and searching between one and four intervals with a bandwidth of 4, 8, 12, 16, 20 and 80 nm, respectively.

For both the interval selection and the peak selection methods, the choice of the most important wavelengths for evaluating plant water status was carried out after applying a combination of the SNV and the Savitzky–Golay first derivative with a 15-point window spectral pre-processing methods. After the selection, different additional pre-processing methods were applied only to the chosen wavelengths in the former spectrum. At the end, the pre-processing method that gave rise to the model with the greatest R^2_{CV} while keeping a small value for RMSECV was chosen. All of these PLS models were created with eight latent variables to avoid overfitting the data. With regard to IPLS selection, three kinds of interval searching algorithms were run, each one corresponding to only three different pre-processing steps: Savitzky–Golay first derivative with a 15-point window, no pre-processing, and SNV. Once the most relevant wavelengths in the spectrum were chosen, PLS models were built without any pre-processing from such wavelengths, again with a maximum of eight latent variables. The whole process is shown in diagram of Fig. 1.

3. Results

3.1. Vineyard water status

The irrigation treatments that were imposed produced a wide variability in plant water status within the vineyard (Table 1), comprising plants under no water stress at all (e.g. Ψ_{stem} = −0.55 MPa) to plant subjected to severe water stress (e.g. Ψ_{stem} = −2.25 MPa). As a result, significant differences ($p < 0.001$) in Ψ_{stem} among treatments across the measuring dates were induced (Fig. 2).

Table 1 – Descriptive statistics of the stem water potential (Ψ_{stem}) across the nine dates of the whole experiment ($n = 324$), expressed in megapascals (MPa).

Irrigation treatment	Minimum	Maximum	Mean	SD
T0-Full irrigation	−1.60	−0.55	−0.88	0.203
T1-Moderate irrigation	−1.70	−0.65	−1.17	0.238
T2-No irrigation	−2.25	−1.10	−1.68	0.286

SD. Standard deviation.

3.2. Spectral measurements

The absorbance spectra of the grapevine canopies (Fig. 3a) in the wavelength range of study (1100–2100 nm) and its transformation with SNV and first derivative (Fig. 3b) clearly revealed two absorption peaks: one at 1450 nm, which was related to the first overtone of the symmetric and asymmetric O–H bond stretching and/or combination bands, and another around 1940 nm, which can be assigned to the combination of the O–H stretching and bending bands (Nicolai et al., 2007). Stretching, bending and combinations are vibrational reactions of the organic groups to the electromagnetic excitation induced by NIR spectroscopy. Since leaves are mostly constituted by water, the prevalence of the hydroxyl (O–H) group absorbance in their NIR spectra is well justified.

3.3. Regression models for grapevine water status

3.3.1. Full spectrum

When no selection methods were carried out (full spectrum, 501 datapoints), a PLS model with eight latent variables coming from a pre-processing that combined SNV and Savitzky–Golay first derivative with a 15-point window (Fig. 3b) was built from the data gathered along the nine dates. This pre-treatment turned out to be the best one for the whole spectrum, and that is the reason why it was applied as well for the interval and peak selection methods. As shown in Table 2, the model built using the information contained in the full spectrum yielded an R^2_P of 0.84 and a RMSEP of 0.167 MPa.

3.3.2. Interval selection method

When the interval selection method was used, and the VIP scores function was considered (Fig. 4a), three wavelength intervals were chosen: 1402–1508 nm, 1676–1750 nm and 1870–1926 nm (Fig. 4b). All the selected intervals involved wavelengths around the O–H string first overtone (1450 nm)

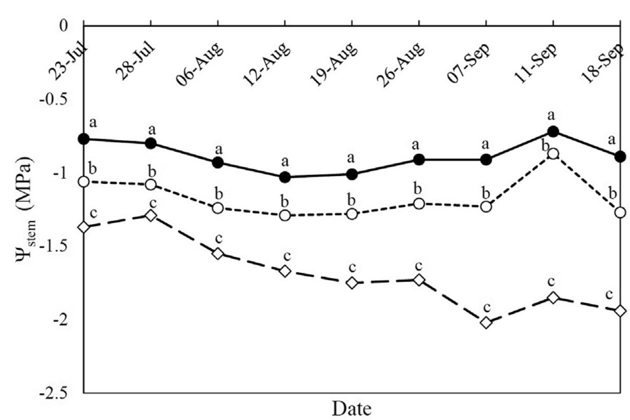

Fig. 2 – Evolution of the average stem water potential (Ψ_{stem}) stated in megapascals (MPa) for each irrigation treatment (T0: Full irrigation, T1: Moderate irrigation, T2: No irrigation) across the ripening season. For each date, the averaged data ($n = 12$) for each irrigation treatment was calculated. Significant differences among the three irrigation treatments at $p < 0.001$ were observed at all dates.

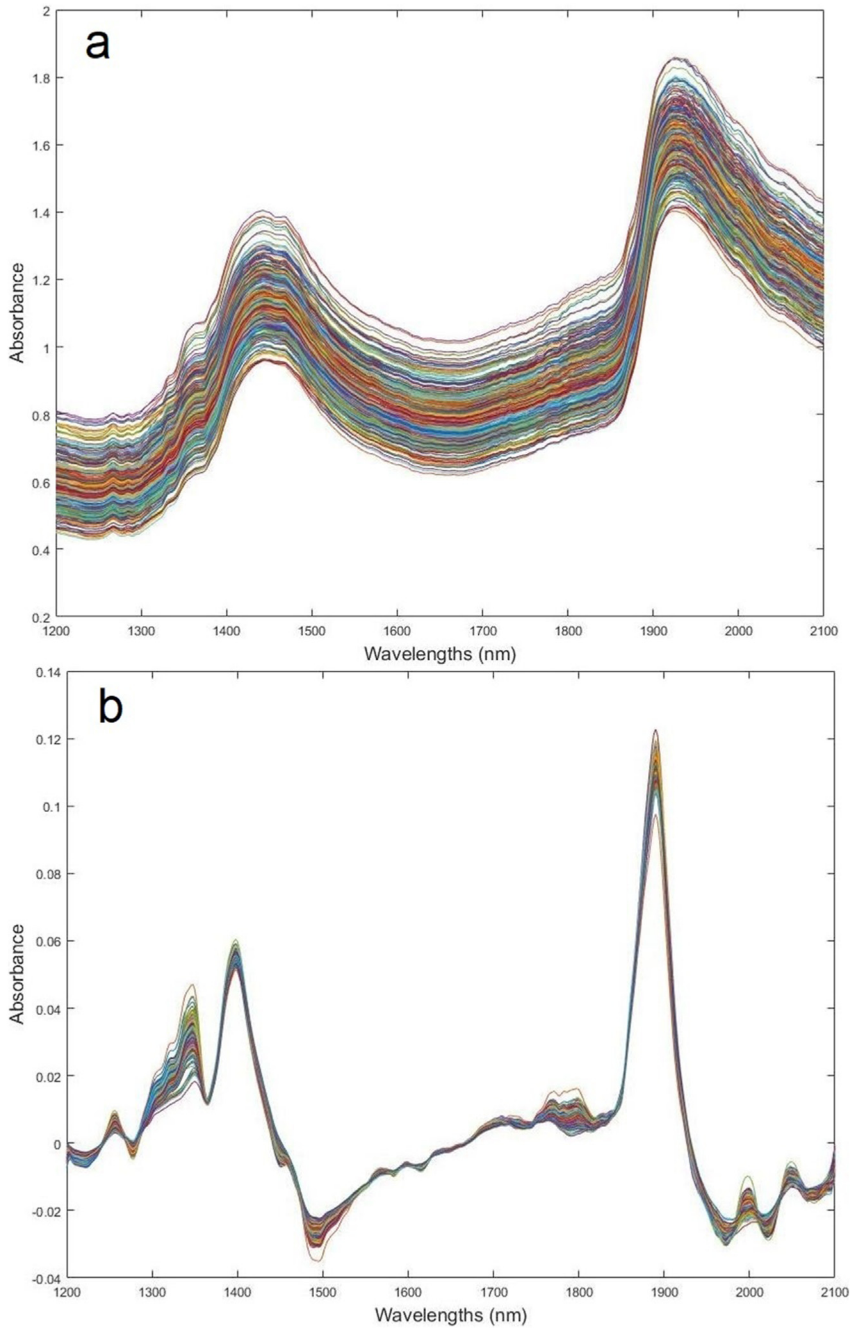

Fig. 3 – (a) Absorbance raw spectra acquired on-the-go (at 5 km h^{-1}) in the vineyard, (b) Absorbance preprocessed spectra with SNV and first derivative.

and the water combination band near the wavelength value of 1940 nm, except the interval between 1676 nm and 1750 nm, which may be related to carbon-hydrogen bonds. With these filters (Table 3), the number of datapoints of the PLS model was reduced from 501 to 121. After applying SNV pre-processing, a value for R^2_P of 0.80 was obtained while the RMSEP was 0.186 MPa.

3.3.3. *Peak selection method*

Regarding the peak selection method, after applying SNV and a Savitzky–Golay first derivative with a 15-point window to the full spectrum and considering the VIP scores function (Fig. 4a), four peaks were chosen: 1428 nm, 1704 nm, 1892 nm and 1912 nm (Fig. 4c). Again, all the selected peaks were near the O–H string overtone and water combination bands, with the exception of the peak at 1704 nm (carbon-hydrogen bonds). As it can be seen in Table 3, after trying different mathematical pretreatments and choosing the optimal one, a bandwidth of 8 nm produced a PLS model using only 20 data points that yielded a R^2_P of 0.77 and an RMSEP of 0.201 MPa.

3.3.4. *IPLS method*

Table 4 summarises the best regression models when the IPLS selection method was applied. For each number of selected

Table 2 – Calibration, validation and prediction results of the best PLS model to predict Ψ_{stem} considering the full spectrum (n = 288). The model was created with eight latent variables after applying a pre-processing consisting of SNV and 1st derivative with a 15-point window. RMSEC, RMSECV and RMSEP are shown in megapascals (MPa).

DP	Calibration		Validation		Prediction	
	R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_P	RMSEP
501	0.86	0.154	0.83	0.171	0.84	0.167

DP: Number of datapoints; RMSEC: Root mean square error of calibration; RMSECV: Root mean square error of cross-validation; RMSEP: Root mean square error of prediction.

intervals (one to four) and bandwidth, only the best PLS models among those created after having filtered the wavelengths according to the three above-mentioned interval searching algorithms were taken. Figure 5 shows the wavelength selections that yielded the best PLS models.

The best result for this method was obtained when three intervals with a bandwidth of 20 nm were selected (Fig. 5c). Unlike the other methods, only one of the intervals comprised wavelengths close to the O–H string overtone or water combination bands; despite this, a PLS model with 33 datapoints, eight latent variables, an R^2_P of 0.78 and an RMSEP of 0.190 MPa could be fitted (Table 4).

With the aim of reaching a trade-off between removing as many redundant wavelengths as possible from the whole spectrum but still producing robust models using the IPLS method, a model with no more than ten datapoints could be built when three intervals were chosen with a bandwidth of 4 nm. This nine-datapoint PLS model had six latent variables, an R^2_P of 0.77 and an RMSEP of 0.199 MPa (Table 4).

4. Discussion

Variable selection in NIR spectroscopy enables to reduce redundant information and/or to remove noisy wavelengths, which turns into less time needed for the system to collect and process the data. This makes sense only if the values of R^2_{CV}, R^2_P, RMSECV and RMSEP, which are key performance metrics, remain stable when compared to the ones yielded by the model developed using the entire spectrum. In all the considered cases in this work, it can be seen that a substantial reduction of the number of datapoints did not lead either to a big reduction of the values of R^2_{CV} and R^2_P nor to a considerable enhancement of the values of RMSECV and RMSEP, respectively, thus resulting in accurate predictions from reduced number of bands. For instance, when moving from 501 datapoints to 121 with the interval selection method (a reduction of 75.85% in the number of datapoints), the values of R^2_{CV} and R^2_P diminished from 0.83 to 0.84 to 0.78 and 0.80, respectively, while the values of RMSECV and RMSEP increased from 0.171 MPa to 0.167 MPa–0.195 MPa and 0.186 MPa, respectively. This trend was held even when the number of datapoints was reduced from 501 to 20 by the peak selection method (that is a decrease of 96.01% from the initial amount of datapoints). In this case, R^2_P was diminished from 0.84 to 0.77, while RMSEP reached 0.201 MPa (an increase of 20.36%), not far from the interval selection method's errors.

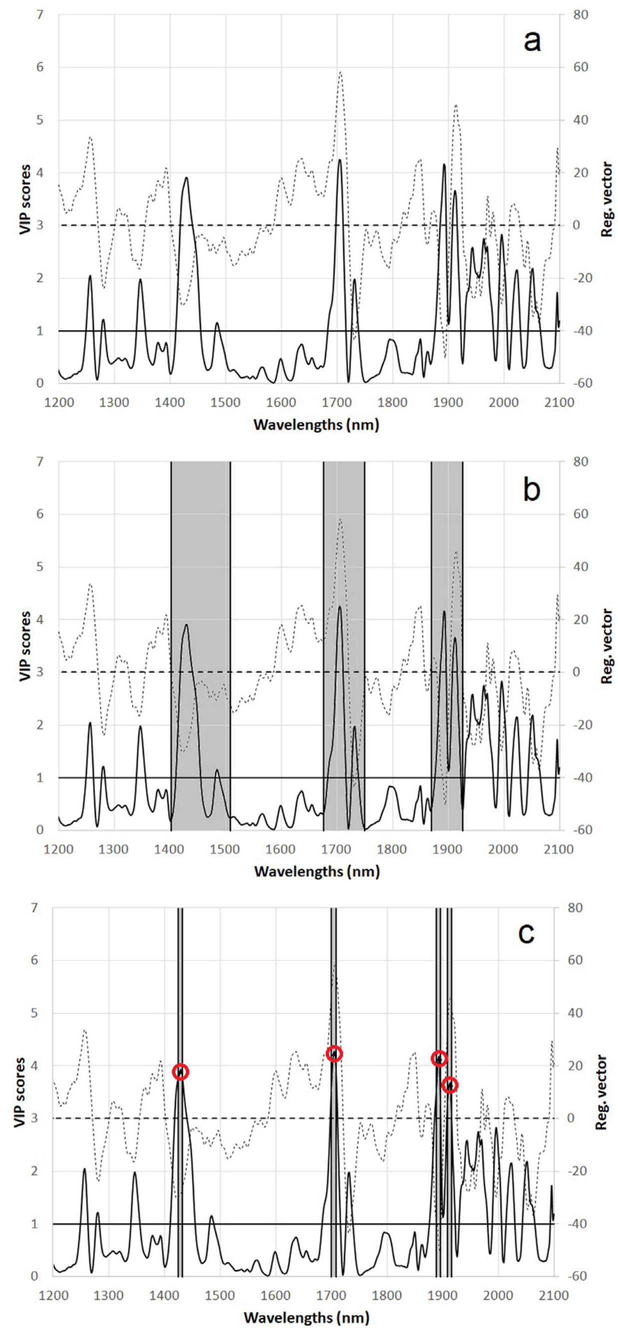

Fig. 4 – Manual selection methods based on VIP scores after applying SNV and Savitzky–Golay first derivative to the whole spectrum (a). The selected intervals are filled in grey. (b) Interval selection method. (c) Peak selection method. In the peak selection method's subplot, the chosen peaks are represented by triangles and the vertical solid lines represent intervals with a bandwidth of 8 nm. In all subplots, the VIP scores function is drawn by a solid line and the regression vector function is drawn by a dashed line.

Table 3 – Wavelength selection according to the manual selection methods based on VIP scores (interval selection and peak selection), along with calibration, validation and prediction results with the optimal pre-processing treatment for the selected intervals to predict Ψ_{stem}. All the models were created with eight latent variables. RMSEC, RMSECV and RMSEP are shown in megapascals (MPa).

BW (nm)	DP	Pre-processing	Calibration		Validation		Prediction	
			R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_P	RMSEP
Interval selection method								
–	121	SNV	0.82	0.175	0.78	0.195	0.80	0.186
Peak selection method								
8	20	1D (15)	0.77	0.201	0.74	0.213	0.77	0.201

BD: Bandwidth; DP: Number of datapoints; RMSEC: Root mean square error of calibration; RMSECV: Root mean square error of cross-validation; RMSEP: Root mean square error of prediction; 1D(15): Savitzky–Golay 1st derivative with a 15-point window; SNV: Standard Normal Deviation.

Although the peak selection method yielded slightly inferior models than the interval selection method, it is worth pointing out that the former has the advantage of selecting intervals with a fixed bandwidth. From a practical standpoint, this may result in more inexpensive filters in case a multispectral solution, using the relevant wavelengths to be assembled and used is aimed, instead of a conventional and more expensive full-spectrum spectrometer (Gutiérrez et al., 2019). Just as the peak selection method, IPLS also generated a selection of intervals with a fixed bandwidth, giving similar models in terms of performance. It should be highlighted that the model generated from three intervals with a bandwidth of 20 nm, by which a reduction of 91.22% in the total of datapoints was obtained, yielded an R^2_P six percentage points smaller (a decrease of 7.14%), an RMSEP of 0.190 MPa (an increase of 14.3%). Even two other IPLS models with yet less datapoints but still comparable to the full-spectra's was achieved: these models are those constituted by three intervals with a bandwidth of 4 or 8 nm (a reduction of more than 95% in the total of datapoints), which yielded a R^2_P of 0.77 and RMSEP between 0.197 and 0.199 MPa. It is worth mentioning that two of the central wavelengths (1420 nm and 1498 nm) of the 3-interval model with 4 nm bandwidth pivoted around the 1450 nm (first overtone of the symmetrical O–H bond).

Nowadays, there are few published works dealing with NIR wavelength selection used for water status predictive models, let alone those focused on vineyards. Of these, Giovenzana et al. (2018) used two portable contact sensors in the field and reported Multiple Linear Regression (MLR) models with an R^2_{CV} of 0.69 and an R^2_P of 0.63 after selecting 14 effective datapoints from the NIR region. These datapoints were selected by applying Regression Coefficient Analysis (RCA) derived from PLS analysis and choosing the highest absolute regression values, and were mainly related to the two main absorption peaks at 1200 and 1450 nm. MLR models with only five datapoints were created. In that work, some wavelengths capable of summarising the largest part of the useful information contained in the spectra were found, and the overall prediction results were comparable to those obtained in the present study.

Other approaches are reported in the literature regarding NIR attribute selection applied for the determination of water

Table 4 – Wavelength selection according to the IPLS[a] selection method, along with calibration, validation and prediction results for the selected wavelengths to predict Ψ_{stem} and the number of latent variables of each model. One to four intervals were picked to create the models. RMSEC, RMSECV and RMSEP are stated in megapascals (MPa).

Central wavelengths (nm)	BW (nm)	DP	Calibration		Validation		Prediction	
			R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_P	RMSEP
1714	80	41	0.77	0.197	0.72	0.219	0.65	0.242
1708	56	29	0.74	0.210	0.70	0.228	0.62	0.252
1550, 1714	80	82	0.84	0.165	0.80	0.184	0.75	0.205
1630, 1720	16	18	0.74	0.214	0.71	0.225	0.69	0.227
1304, 1550, 1714	80	123	0.82	0.174	0.80	0.184	0.79	0.189
1418, 1638, 1792	20	33	0.81	0.179	0.79	0.191	0.78	0.190
1344, 1604, 1704	8	15	0.78	0.197	0.74	0.211	0.77	0.197
1420, 1498, 1690	4	9	0.72	0.221	0.70	0.228	0.77	0.199
1222, 1304, 1550, 1714	80	164	0.83	0.172	0.80	0.185	0.79	0.186
1418, 1638, 1704, 1792	20	44	0.83	0.169	0.81	0.182	0.78	0.191
1344, 1494, 1604, 1704	8	20	0.81	0.183	0.78	0.194	0.77	0.200
1582, 1726, 1738, 1882	4	12	0.79	0.190	0.77	0.197	0.68	0.231

BD: Bandwidth; DP: Number of datapoints; RMSEC: Root mean square error of calibration; RMSECV: Root mean square error of cross-validation; RMSEP: Root mean square error of prediction.

[a] The IPLS method was developed without using any pretreatment, since better models were obtained this way.

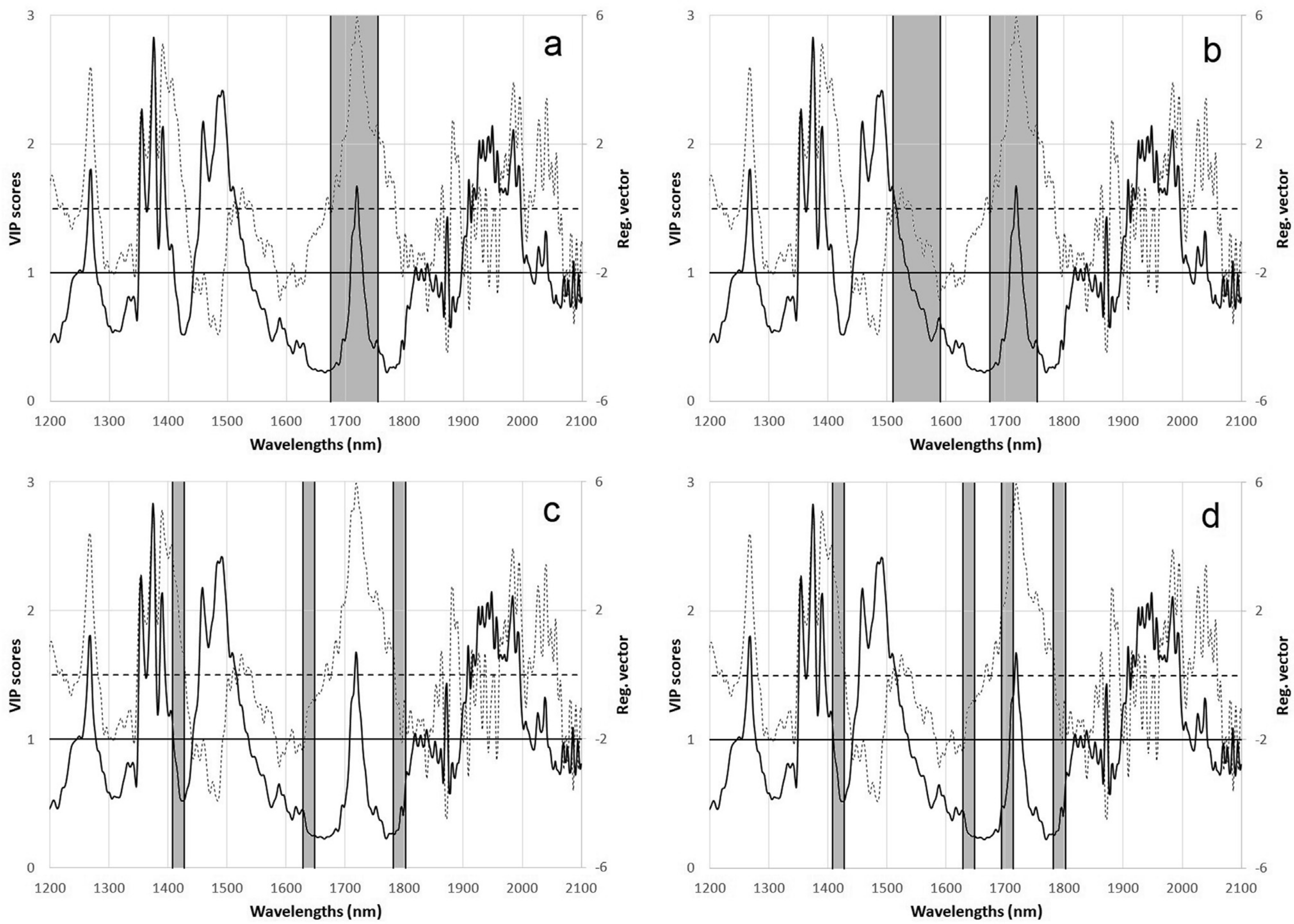

Fig. 5 – IPLS wavelength selection. The selected intervals are filled in grey. A suitable bandwidth to obtain the best models when one (a) or two (b) intervals were selected was 80 nm, and 20 nm when three (c) or four (d) intervals were chosen, respectively. In all subplots, the VIP scores function is drawn by a solid line and the regression vector function is drawn by a dashed line. Both functions were obtained from the full spectrum without any pre-processing.

content in plants other than vines. Song et al. (2011) selected the most sensitive wavelengths for the discrimination of the imperceptible spectral variations of paddy rice, identifying three wavelengths (1158, 1378 and 1965 nm) as the most useful bands to diagnose the stress condition. The PLS model that was created in their work was quite encouraging, as it had a coefficient of determination R^2_{CV} of 0.72. Principal Component Analysis (PCA) and band–band correlation methods were used to select the most significant wavelengths. Besides, Zhang et al. (2012) investigated the feasibility of detecting the water content in *Epipremnum Aureum* leaves using the diffuse reflectance spectra limited in the Vis/NIR region (400–1100 nm), proposing a hybrid wavelength selection that combines Backward Interval PLS (BIPLS) and Successive Projection Algorithm (SPA) to extract the efficient feature wavelength. This selection method returned an outstanding PLS model with 25 datapoints, an R^2_{CV} of 0.85, an RMSECV of 1.5%, an R^2_P of 0.96 and an RMSEP of 0.73%. Therefore, it proved to extract the most efficient information regarding water content, and improved the model precision and stability.

However, unlike the present work, all of these studies involved spectral data acquisition in a laboratory (under controlled conditions) from randomly chosen leaves that were then pulled up and destroyed or statically on potted vines. One of the major weaknesses towards the operational application of NIR spectroscopy to assess vineyard water status spatial variability, as identified in Pampuri's SWOT analysis (Pampuri et al., 2021) was the difficulty of positioning stand-alone sensors in the field. Towards this end, it has to be highlighted that the spectral data in the present work were acquired contactless, on-the-go, from a moving vehicle, and directly from the canopy in the field, therefore mitigating this pitfall. As opposed to indoor spectroscopy, on-the-go in-field spectroscopy has to overcome not only the changing natural illumination (mostly mitigated by the light of the lamp installed in the moving vehicle), but also vibrations derived from the unevenness of the terrain, and fluctuations in the distance to the target, inherent to the moving acquisition, particularly at speeds such as 5 km h^{-1}, which can be considered "operational speed" as most machinery in vineyards operates between 3 and 5 km h^{-1}. All these experimental factors may cause artefacts in the spectral signal acquired from the moving vehicle, and could potentially lead to reduced performance of the derived models, when compared to those obtained from spectral data acquired under controlled static conditions. As demonstrated in the present work, the performance of the derived models was comparable to that of those obtained in more controlled studies (Wei et al., 2021).

Likewise, the results obtained in this study suggest that PLS models with a reduced number of wavelengths (three or four, from the IPLS method) were almost as good in estimating the grapevine water status as those that used the whole spectrum, having the plus of lacking redundant wavelengths and being simpler and more economical to implement. This brings the opportunity of building simpler, more economical, miniaturised multispectral sensors involving only three to four wavelengths of interest to be used for on-the-go plant water status monitoring in a non-destructive way, directly in the vineyard. Additionally, the vast amount of data that could be retrieved by this sensor would enable the assessment of vineyard water status variability, hence ensuring the representativeness needed to design optimised irrigation strategies. Quantitatively speaking, the values of RMSEP yielded by these spectral methods using discrete reduced information are higher than those of the Schölander pressure bomb method to assess plant water potential when this technique is performed by well-trained operators (Rodríguez-Domínguez et al., 2022). However, the process of measuring with the Schölander pressure bomb in the field involves several experimental steps which are not always conducted in a similar way by operators, and that may induce increased uncertainty in final measurement (Rodríguez-Domínguez et al., 2022). These potential sources of error may include: the need for equilibration of previously transpiring leaves; leaf storage before measurement; the equilibration of Ψleaf for leaves on bagged branches, the use of ‘pulse’ pressurisation versus gradual pressurisation and different skill among operators. As shown in Rodríguez-Dominguez et al. (2022), the uncertainties associated to some of these factors may range between 0.1 and 0.2 MMPa, and become even larger when untrained operators carry out the measurements. As a result, only in plant water potential measurements with the Schölander pressure bomb are carried out following a best practice protocol, these sources or error are minimised. Moreover, from a qualitatively point of view, the accuracy achieved with the spectral method is enough to determine whether the water stress level is high, medium or low. Likewise, this classification is acceptable for industry operators to manage irrigation in a more efficient way.

4.1. *Future development and requirements*

With this experiment, it is intended to select certain wavelengths in the NIR region of the spectrum to develop a low-cost sensor for estimating water status of vineyards with an adequate precision, classifying it as high, medium or low.

In order to measure the referred water status in real time, it is necessary to carry out a rapid processing of the data, which can be obtained by initially taking the spectrum of the entire bandwidth in the NIR region and then develop an ad hoc software that takes into account only the wavelengths stated in this article when building the PLS models.

At this time, we believe that the filtering of the wavelengths should be done by a custom device with this type of software implemented in it, since it is currently very expensive to purchase specific filters in the NIR zone between 1600 nm and 2100 nm and, assuming that they are affordable, the filter bandwidth may not suit our needs.

5. Conclusions

The results obtained in this work allows for claiming that variable selection is a reliable method towards the simplification of the estimation of plant water status using NIR spectroscopy on-the-go in commercial vineyards. Despite the fact that a large number of bands of the full spectrum were removed when applying all the considered attribute selection methods, the robustness of the generated models was never compromised. Of the three NIR attribute selection methods tested in this work, the IPLS method proved to yield the best reduced models (with only three to four wavelengths) and narrow bandwidths between 4 and 8 nm, causing a diminishment in the number of points from 501 (whole spectrum) to 9 to 15 points.

This could be interpreted as a first step towards the design of low-cost, simpler multispectral instruments that would be able to rapidly and non-destructively determine grapevine water status by building accurate predictive models from a reduced amount of data collected from a moving terrestrial vehicle. The invention of these devices may boost the establishment of more profitable irrigation strategies in the future.

Funding

Grant PID2019-108330RA-I00 funded by MCIN/AEI/10.13039/501100011033 (Spain).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors want to thank Bodegas Dinastía Vivanco (La Rioja) for providing the vineyard and taking care of the irrigation procedures.

REFERENCES

Al-Degs, Y. S., El-Sheikh, A. H., Saleh, A. I., & Al-Reyahi, A. Y. (2021). Interval wavelength selection and simultaneous quantification of spectrally overlapping food colorants by multivariate calibration. *Food Measure, 15*, 2562–2575.

Araújo-Paredes, C., Portela, F., Mendes, S., & Valín, M. I. (2022). Using aerial thermal imagery to evaluate water status in *Vitis vinifera* cv. Loureiro. *Sensors, 22*, 8056.

Babellahi, F., Paliwal, J., Erkinbaev, C., Amodio, M. L., Chaudhry, M. M. A., & Colelli, G. (2020). Early detection of chilling injury in green bell peppers by hyperspectral imaging and chemometrics. *Postharvest Biology and Technology, 162*, 111100.

Barnes, R., Dhanoa, M., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy, 43*, 772–777.

Cho, B. H., Lee, K. B., Hong, Y., & Kim, K. C. (2022). Determination of internal quality indices in oriental melon using snapshot-type hyperspectral image and machine learning model. *Agronomy, 12*(9), 2236.

Chone, X., Van Leeuwen, C., Dubourdieu, D., & Gaudillère, J. P. (2001). Stem water potential is a sensitive indicator of grapevine water status. *Annals of Botany, 87*, 477–483.

De Bei, R., Cozzolino, D., Sullivan, W., Cynkar, W., Fuentes, S., Dambergs, R., Pech, J., & Tyerman, S. (2011). Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian Journal of Grape and Wine Research, 17*, 62–71.

Dhanoa, M., Lister, S., & Barnes, R. (1995). On the scales associated with near-infrared reflectance difference spectra. *Applied Spectroscopy, 49*, 765–772.

Diago, M. P., Fernández-Novales, J., Gutiérrez, S., Marañón, M., & Tardaguila, J. (2018). Development and validation of a new methodology to assess the vineyard water status by on-the-go near infrared spectroscopy. *Frontiers of Plant Science, 9*, 59.

Dong, X., & Sun, X. A. (2013). Case study of characteristic bands selection in near-infrared spectroscopy: nondestructive detection of ash and moisture in wheat flour. *Journal of Food Measurement and Characterization, 7*, 141–148.

Ericksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). *Multi-and megavariate data analysis principles and applications.* Umea: Umetrics Academ.

Fernández-Novales, J., Tardáguila, J., Gutiérrez, S., & Diago, M. P. (2019). On-The-Go VIS+SW-NIR spectroscopy as a reliable monitoring tool for grape composition within the vineyard. *Molecules, 24*, 2795.

Fernández-Novales, J., Tardaguila, J., Gutiérrez, S., Marañón, M., & Diago, M. P. (2018). In field quantification and discrimination of different vineyard water regimes by on-the-go NIR spectroscopy. *Biosystems Engineering, 165*, 47–58.

Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., & Guidetti, R. (2014). Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. *American Journal of Enology and Viticulture, 65*, 117–123.

Giovenzana, V., Beghi, R., Parisi, S., Brancadoro, L., & Guidetti, R. (2018). Potential effectiveness of visible and near infrared spectroscopy coupled with wavelength selection for real time grapevine leaf water status measurement. *Journal of the Science of Food and Agriculture, 98*, 1935–1943.

Guidetti, R., Beghi, R., & Bodria, L. (2010). Evaluation of grape quality parameters by a simple Vis/NIR system. *Transactions of the ASABE, 53*, 477–484.

Gutiérrez, S., Fernández-Novales, J., Diago, M. P., Íñiguez, R., & Tardaguila, J. (2021). Assessing and mapping vineyard water status using a ground mobile thermal imaging platform. *Irrigation Science, 39*, 457–468.

Gutiérrez, S., Tardaguila, J., Fernández-Novales, J., & Diago, M. (2019a). On-the-go hyperspectral imaging for the in-field estimation of grape berry soluble solids and anthocyanin concentration. *Australian Journal of Grape and Wine Research, 25*, 127–133.

Gutiérrez, S., Wendel, A., & Underwood, J. (2019b). Spectral filter design based on in-field hyperspectral imaging and machine learning for mango ripeness estimation. *Computers and Electronics in Agriculture, 164*, 104890.

Hinkelmann, K., & Kempthorne, O. (2007). Randomized block designs. In *Design and Analysis of experiments: Introduction to experimental design*. John Wiley & Sons, Inc.

Jones, H. G. (2004). Irrigation scheduling: Advantages and pitfalls of plant-based methods. *Journal of Experimental Botany, 55*, 2427–2436.

Knadel, M., Arthur, E., Weber, P., Moldrup, P., Greve, M. H., Chrysodonta, Z. P., & De Jonge, L. W. (2018). Soil specific surface area determination by visible near-infrared spectroscopy. *Soil Science Society of America Journal, 82*(5), 1046–1056.

Lantao, L. I., Tao, K. E. N., & Shanqin, W. A. N. G. (2017). Prediction models of winter oilseed rape yield based on hyperspeetral data at pod-filling stage [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery, 48*(3), 221–229.

Leardi, R., & Nørgaard, L. (2004). Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of Chemometrics, 18*, 486–497.

Leiva-Valenzuela, G. A., Lu, R., & Aguilera, J. M. (2014). Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths. *Innovative Food Science & Emerging Technologies, 24*, 2–13.

Li, L., Lin, D., Wang, J., Yang, L., & Wang, Y. (2020). Multivariate analysis models based on full spectra range and effective wavelengths using different transformation techniques for rapid estimation of leaf nitrogen concentration in winter wheat. *Frontiers of Plant Science, 11*, 755.

Lin, Z., Pei, Y., Chen, Z., Shi, X., & Qiao, Y. (2013). Improving the creditability and reproducibility of variables selected from near infrared spectra. In *Natural Computation (ICNC), 2013 Ninth International Conference on* (pp. 1370–1376). IEEE.

Liu, C. K., Wang, Q. H., Lin, W. G., & Yu, C. D. (2022). Origins classification of egg with different storage durations using FT-NIR: A characteristic wavelength selection approach based on information entropy. *Biosystems Engineering, 222*, 82–92.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., & Saeys, W. (2007). Theron, K.I.; Lammertyn, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*, 99–118.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy, 54*, 413–419.

Pampuri, A., Tugnolo, A., Bianchi, D., Giovenzana, V., Beghi, R., Fontes, N., Oliveira, H. M., Casson, A., Brancadoro, L., & Guidetti, R. (2021). Optical specifications for a proximal sensing approach to monitor the vine water status in a distributed and autonomous fashion. *Biosystems Engineering, 212*, 388–398.

Pan, X., Sun, L., Sun, G., Rong, P., Lu, Y., Li, J., & Song, Z. (2020). Prediction of neutral detergent fiber content in corn stover using near-infrared spectroscopy technique. *International Journal of Food Engineering, 16*(12).

Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society, 14*, 198–219.

Pôças, I., Gonçalves, J., Costa, P. M., Gonçalves, I., Pereira, L. S., & Cunha, M. (2017). Hyperspectral-based predictive modelling of grapevine water status in the Portuguese Douro wine region. *International Journal of Applied Earth Observation and Geoinformation, 58*, 177–190.

Pôças, I., Rodrigues, A., Gonçalves, S., Costa, P., Gonçalves, I., Pereira, L., & Cunha, M. (2015). Predicting grapevine water status based on hyperspectral reflectance vegetation indices. *Remote Sensing, 7*, 16460–16479.

Pôças, I., Tosin, R., Gonçalves, I., & Cunha, M. (2020). Toward a generalised predictive model of grapevine water status in Douro region from hyperspectral data. *Agricultural and Forest Meteorology, 280*.

Rady, A. M., Guyer, D. E., Kirk, W., & Donis-González, I. R. (2018). The potential application of sensor fusion and wavelength selection on prediction of leaf primordia of potato tubers. In *2018 ASABE annual international meeting* (p. 1). American Society of Agricultural and Biological Engineers.

Rapaport, T., Hochberg, U., Shoshany, M., Karnieli, A., & Rachmilevitch, S. (2015). Combining leaf physiology, hyperspectral imaging and partial least squares-regression (pls-r) for grapevine water status assessment. *ISPRS Journal of Photogrammetry and Remote Sensing, 109*, 88–97.

Rienth, M., & Scholasch, T. (2019). State of the art of tools and methods to assess vine water status. *Oeno One, 4*, 619–637.

Rodríguez-Dominguez, C. M., Forner, A., Martorell, S., Choat, B., Lopez, R., Peters, J. M. R., Pfautsch, S., Mayr, S., Carins-Murphy, M. R., McAdam, S. A. M., Richardson, F., Díaz-Espejo, A., Hernandez-Santana, V., Menezes-Silva, P. E., Torres-Ruiz, J. M., Batz, T. A., & Sack, L. (2022). Leaf water potential measurements using the pressure chamber: Synthetic testing of assumptions towards best practices for precision and accuracy. *Plant, Cell and Environment*, 1–25.

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*, 1627–1639.

Shi, J., Zou, X., Zhao, J., Mao, H., Wang, K., & Chen, Z. (2011). NIR spectra in non-invasive measurement of cucumber leaf chlorophylls content. *Nongye Jixie Xuebao, 42*.

Song, S., Gong, W., Zhu, B., & Huang, X. (2011). Wavelength selection and spectral discrimination for paddy rice, with laboratory measurements of hyperspectral leaf reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 672–682.

Sun, X., Zhang, H., Ouyang, A., & Liu, Y. (2009). Selection of NIR characteristic wavelength bands for soluble solids content in Nanfeng Mandarin fruit. *Nongye Jixie Xuebao= Transactions of the Chinese Society for Agricultural Machinery, 40*(7), 129–132.

Tardaguila, J., Fernández-Novales, J., Gutiérrez, S., & Diago, M. P. (2017). Non-destructive assessment of grapevine water status in the field using a portable NIR spectrophotometer. *Journal of the Science of Food and Agriculture, 97*, 3772–3780.

Tosin, R., Martins, R., Pôças, I., & Cunha, M. (2022). Canopy VIS-NIR spectroscopy and self-learning artificial intelligence for a generalised model of pre-dawn leaf water potential in *Vitis vinifera*. *Biosystems Engineering, 219*, 235–258.

Tosin, R., Pôças, I., Gonçalves, I., & Cunha, M. (2020). Estimation of grapevine predawn leaf water potential based on hyperspectral reflectance data in Douro wine region. *Vitis: Journal of Grapevine Research, 59*, 9–18.

Tosin, R., Pôças, I., Novo, H., Teixeira, J., Fontes, N., Graça, A., & Cunha, M. (2021). Assessing predawn leaf water potential based on hyperspectral data and pigment's concentration of Vitis vinifera L. In the Douro wine region. *Scientia Horticulturae, 278*.

Wei, H.-E., Grafton, M., Bretherton, M., Irwin, M., & Sandoval, E. (2021). Evaluation of point hyperspectral reflectance and multivariate regression models for grapevine water status estimation. *Remote Sensing, 13*(16).

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-Regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109–130.

Wu, D., He, Y., Nie, P., Cao, F., & Bao, Y. (2010). Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Analytica Chimica Acta, 659*, 229–237.

Xiaobo, Z., Jiewen, Z., Povey, M. J., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*, 14–32.

Zhang, Q., Li, Q., & Zhang, G. (2012). Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection. *Journal of Spectroscopy, 27*, 93–105.

Zhou, Z., Diverres, G., Chenchen, K., Sushma, T., Manoj, K., Qin, Z., & Keller, M. (2022). Ground-based thermal imaging for assessing crop water status in grapevines over a growing season. *Agronomy, 12*, 322.

Zou, X., Chen, Z., Shi, J., Huang, X., & Zhang, D. (2012). Rapid detection of cucumber leaves pigments based on near infrared hyper-spectral image technology. *Nongye Jixie Xuebao, 43*, 152–155.

Zou, X., Huang, X., Shi, J., Chen, Z., & Zhang, D. (2012). Selection of wavelength regions to determine flavonoids content in ginkgo leaves by FT-NIR spectroscopy. *Nongye Jixie Xuebao, 43*, 155–159.