Physiological variable predictions using VIS-NIR spectroscopy for water stress detection on grapevine: interest in combining climate data using multiblock method

Maxime Ryckewaert[a,d], Daphné Héran[a], Thierry Simonneau[b], Florent Abdelghafour[a], Romain Boulord[b], Nicolas Saurin[c], Daniel Moura[a], Silvia Mas-Garcia[a,d], Ryad Bendoula[a]

[a] *ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France*
[b] *LEPSE, Univ Montpellier, INRAE, Institut Agro Montpellier SupAgro, Montpellier, France*
[c] *Pech-Rouge, Univ Montpellier, INRAE, Montpellier, France*
[d] *ChemHouse Research Group, Montpellier, France*

Abstract

In recent years, climate fluctuations have been increasingly extreme, affecting agricultural production. The development of digital agriculture driven by new intelligent sensors is one of the privileged paths to improve farm management. Assessing transpiration E and stomatal conductance g_s in real time with optical instruments is a real challenge to detect water stress. In this study, the objective is to evaluate VIS-NIR spectroscopy to predict transpiration E and stomatal conductance g_s of grapevine plants (Vitis vinifiera L.). For this purpose, a water stress gradient was obtained using vine pots of three varieties (Syrah, Merlot, Riesling) tested under two water conditions where precise monitoring of physiological variables has been carried out. Hyperspectral images were acquired to form a spectral database and a weather station provided radiation (Rg), relative humidity (RH), temperature (Ta) and wind speed (Ws). First, Partial Least Squares (PLS) models were established to relate spectral data to physiological variables. Then, Sequential Orthogonalized-Partial Least Squares (SO-PLS) was used to predict these physiological variables with two blocks: spectral and climate data. PLS models are obtained for g_s (R^2= 0.656, bias=8.76 mmol.m^{-2}.s^{-1}, RMSE=64.7 mmol.m^{-2}.s^{-1}) and E (R^2= 0.625, bias=-0.02 mmol.m^{-2}.s^{-1}, RMSE=0.67 mmol.m^{-2}.s^{-1}). For E, improved results (R^2= 0.699, bias=0.055 mmol.m^{-2}.s^{-1}, RMSE=0.614 mmol.m^{-2}.s^{-1}) are obtained by using climate data with SO-PLS. Generic PLS models achieved good predictive quality despite different coloured berry varieties. Quality of these prediction models could be improved by defining varietal models on a larger data set. Merging spectral data with climate data improves prediction quality of transpiration variable providing insights by adding further information with the aim of improving predictive qualities.

Keywords: Spectroscopy, Water Stress, Physiological variables, Digital Agriculture, Multivariate Data Analysis, Fusion data

---

1. Introduction

In recent years, climate fluctuations have been increasingly extreme, affecting agricultural production (Lobell et al., 2011). In this context, one of the biggest challenges for agriculture is to adapt agricultural practices to these constraints.

Moreover, resulting abiotic stresses may modify crop sensitivity to diseases (Mittler, 2006). In the event of a severe and persistent stress episode, the consequence is an inexorable reduction in yield. Therefore, it becomes critical to provide new methods to manage crops in order to avoid episodes of severe abiotic stresses (Olesen et al., 2011).

The development of digital agriculture driven by new intelligent sensors is one of the privileged paths to improve farm management (Fountas et al., 2020; Zhai et al., 2020). The increasing adoption of sensors for agriculture provides valuable data driving farm actions such as input use (pesticides, irrigation, fertilisers) or harvest planning.

In crop production, soil drying is especially at stake on plant status. In fact, water stress induces several physiological markers on plants (Simonneau et al., 2017). At the leaf level, transpiration and stomatal conductance are two physiological parameters closely related to the plant water status. These physiological variables are highly influenced by multiple environmental factors, such as soil water availability, air humidity, radiation and temperature (Damour et al., 2010).

Assessing such physiological parameters in real time with optical instruments is a real challenge (Schultz and Stoll, 2010). Thermography has long been studied as an optical instrument that gives an indication of transpiration rate (Romano et al., 2011; Zhou et al., 2021). Indeed, a decrease in transpiration, *e.g.* following a soil drying, leads to vegetation warming, which is a consequence of stomatal closure.

Alternatively, visible and near-infrared (VIS-NIR) spectroscopy has been widely used to evaluate vegetation status. In this spectral region (VIS-NIR), information is related to both pigments and cell structure parameters (Xu et al., 2019; Ryckewaert et al., 2021) which are parameters altered by water stress. Moreover, maintaining a high spectral resolution in the VIS-NIR region improves the description of vegetation responses to water stress (Ryckewaert et al., 2021). This technology indirectly reveals different properties that can describe water stress such as a decrease in chlorophyll content (Steidle Neto et al., 2017) and water content (Zhang et al., 2012) in leaves. The use of VIS-NIR spectroscopy provides the capability to study spectral bands most related to the prediction of physiological variables such as stomatal conductance (Rapaport et al., 2015; Dao et al., 2021).

Methodological efforts using chemometrics are expected to further exploit spectral information from the vegetation (Mulla, 2013). A reference regression method called Partial Least Squares (PLS) (Wold et al., 2001) provides models to predict response variables from spectral data. Recently, data fusion methods, also called multiblock methods, have been developed to combine several data sources. Adding an additional block such as climate data to spectral data can potentially improve the performances of prediction models. The most commonly used multiblock method is Sequential Orthogonalized-Partial Least Squares (SO-PLS) regression (Naes et al., 2011) which sequentially exploits information from different blocks *i.e.* types of data. With this method, the association of spectral data with climate data has the potential to improve physiological variable prediction.

The objective of this paper is to study the potential of VIS-NIR spectroscopy combined with climate data to predict physiological markers of water stress such as transpiration and stomatal conductance of grapevine plants (Vitis vinifiera L.).

For this purpose, an experimental campaign with precise monitoring of physiological variables has been carried out. First of all, physiological variables are predicted by using spectral data exclusively. Then, spectral and climate data combination is used for predictions using the data fusion method called SO-PLS method.

2. Materials and methods

2.1. Experimental design

The experiment took place in summer 2020 at "Institut Agro" in Montpellier, France (fig. 1). The objective of this experiment was to obtain a water stress gradient on three grape varieties, Syrah, Riesling and Merlot, of different ages (Syrah: two years old; Riesling: three years old; Merlot: four years old) treated with two irrigation conditions also called modalities. Each plant was grafted and pruned to bear only one main axis, then potted in a 9L pot filled with 3800 $\pm$ 1 g of a dry mixture of 70 % peat and 30 % clay (custom-made potting soil, Klasmann-Deilmann), and staked vertically. The plants were arranged in rows oriented almost north-south with a density of 2.8 plants/linear meter along the row. The main branch of each plant was pruned at 22 leaves and all secondary branches were retained.

The first irrigation condition, called well-watered (WW), consisted of irrigating the pots several times a day to maintain the weight of each pot at a level that did not constrain plant growth (this level was determined in previous experiments). The second irrigation condition was defined in order to induce a water deficit (WD). This second condition consisted of stopping irrigation during one week. Each modality contained 6 plants per variety, representing a total of 36 pots. The experiment was repeated in its entirety twice with new pots at one-week interval with suspension of irrigation on

06/22/2020 for the first experiment, and on 07/06/2020 for the second.

Figure 1: Vine pots used for experimentation

2.2. *Physiological measurements*

The transpiration of each plant was determined from the weight evolution of each pot placed on a strain gauge load cell (Micro Load Cell model CZL635, range 20 kg, mean error ± 70 g), recorded every 30 seconds (data logger CR1000 Campbell Scientific, Leicestershire, UK). Transpiration rate was given by a linear regression of weight versus time over a 4h time-frame. Transpiration per leaf surface (E), corresponding to water loss through leaves only, was then adjusted to the total leaf area of each plant, estimated from planimeter measurements (LI-3100C LI-COR Biosciences Inc., Nebraska, USA), and vein size measurements, which were converted to area from previously established charts.

Stomatal conductance was measured with a porometer (Model AP4, Delta-T Devices, Burwell, Cambridge, UK) on young mature leaves with good sun exposure (east side of the row in the morning and west side in the afternoon).

This measurement was carried out on one leaf per plant chosen at random three times a day (9am, 11.30am and 3.30pm, UTC+2). The porometer was calibrated before each series of measurements (three times a day).

2.3. Climate data

Four climate parameters were measured: Air temperature (Ta), relative humidity (RH), global radiation (Rg) and wind speed (W_s).

T and RH were measured with a capacitive thermohygrometer (HMP35A Vaisala; Oy, Helsinki, Finland) placed in a naturally aspirated radiation shield at 2.5m height. Rg was measured with a PPFD sensor (LI-190SB; LI-COR, Lincoln, NE, USA). W_s was measured with a 3-cup anemometer (A100L2, Vector Instruments, Denbighshire, UK). Data were collected every 30 seconds, averaged over 1800 seconds and stored in a datalogger (CR10X; Campbell Scientific Ltd, Shepshed, Leicestershire, UK).

2.4. Hyperspectral acquisitions

Hyperspectral images were acquired using a hyperspectral camera (Specim, Specim IQ, Finland) covering the spectral range from 400 nm to 1000 nm with 204 spectral bands. Spectral regions were cut off after 800 nm due to the high level of noise in this experiment. The distance between the camera and vine pots was set to approximately 1 meter. Images were acquired each day for each modality at three different times (8am, 2pm and 4pm, UTC+2) producing a set of 160 hyperspectral images. Camera orientation during image capture was defined to minimise direct sunlight. Thus, west side was privileged at 8am and 2pm and east side at 4pm (see fig. 2).

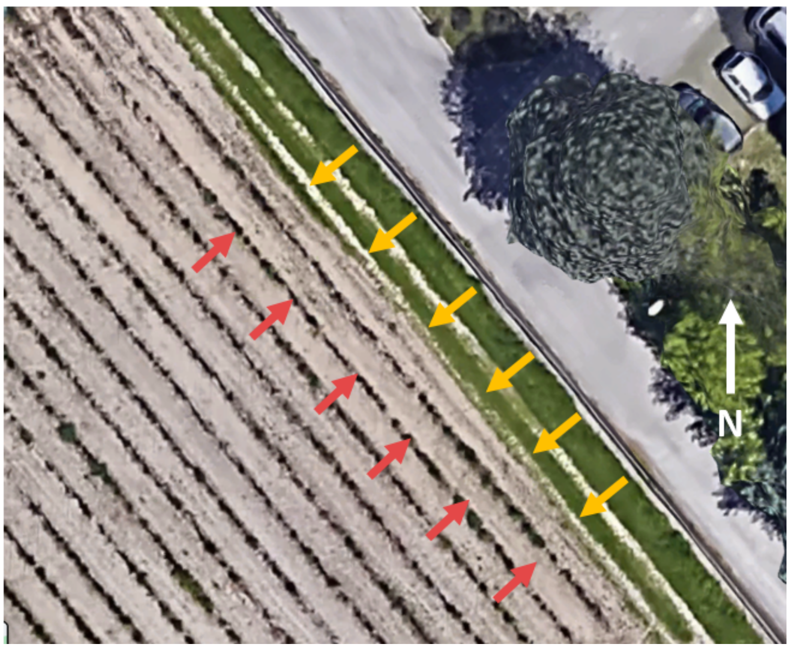


Figure 2: Camera orientation during hyperspectral image capture: towards west (orange arrows) at 8am and 2pm and east (red arrows) at 4pm. Source: https://www.geoportail.gouv.fr/

Figure 3: Typical scene where hyperspectral images were acquired. One image corresponded to one of the modalities of the experimental design with 6 pots of vines. The reference is placed in the scene.

A white reference (SRS99, Spectralon ®) was used to measure natural incident light ($I_0(\lambda)$) to standardise all measured images from non-uniformities of all instrumentation components (light source, lens, detector). This reference was systematically placed in the scene for each image acquisition (see fig. 3). From these measurements, reflectance ($R_s(\lambda)$) was calculated for each image:

$$R_s(\lambda) = \frac{I_s(\lambda) - I_b(\lambda)}{I_0(\lambda) - I_b(\lambda)} \tag{1}$$

where $I_s(\lambda)$ is the reflected light intensity, $I_b(\lambda)$ the dark current image recorded by the camera.

2.5. Data analysis

All computations, data processing and multivariate data analysis were performed with MATLAB software v.R2015b (The Mathworks Inc., Natick,MA, USA).

2.5.1. Image preprocessing workflow

The presented workflow was established to generate a spectral database. This workflow was defined in three main steps:

The first step was to manually extract an area corresponding to foliage to obtain a reference spectrum $\mathbf{s}_{ref}$. The second step was to identify, for all images, vegetation pixels that present similar spectra to the reference vegetation spectrum. For this purpose, the Spectral Angle Mapper (SAM) (Kruse et al., 1993) was used as an indicator that describes spectral similarity between two spectra. Expressed in degrees, this indicator calculates the angle formed between $\mathbf{s}_{ref}$ and all spectra of an image in the vector space defined by the wavelengths. For a given pixel i, the SAM between $\mathbf{s}_{ref}$ and $\mathbf{s}_i$ is written as follows:

$$\mathrm{SAM}(\mathbf{s}_{ref}, \mathbf{s}_i) = \arccos\left(\frac{\langle \mathbf{s}_{ref}, \mathbf{s}_i \rangle}{\|\mathbf{s}_{ref}\| \, \|\mathbf{s}_i\|}\right) \tag{2}$$

With $\|.\|$ being euclidean norm. Spectra are similar when angle value is close to 0°. Conversely, the larger the SAM value, the higher the difference between the two spectra. This indicator value has the advantage of being independent of signal intensity.

After identifying vegetation-related pixel in one image, the third step was to create a subset of 500 vegetation pixels without any outlier based on their spectra. For this purpose, a principal component analysis (PCA) was applied on all vegetation spectra of one image. Then, Q-residuals and T^2 criteria were computed in order to identify potential outlier spectra. 500 pixels were randomly selected excluding outliers.

Finally, the 500 collected spectra were averaged by modality (*i.e.* per image) forming a total of 160 spectra.

2.5.2. PLS predictions based on spectral data

In chemometrics, PLS regression (Wold et al., 2001) is the widely used method to predict a reference variable $\mathbf{y}$ from spectral data $\mathbf{X}$. $\mathbf{X}$ dimension is n x p where n is the total number of observations and p the number of variables or wavelengths. $\mathbf{y}$ dimension is n x 1. The final equation of PLS regression can be written as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}_X \tag{3}$$

Where $\mathbf{b}$ is the vector containing regression coefficients and $\mathbf{r}_X$ is a vector containing residuals of the model.

To do so, a model is established between intermediate variables, called latent variables computed respectively from $\mathbf{X}$ and $\mathbf{y}$. The adjustment of these latent variables is performed according to different iterative algorithms. Essentially, $\mathbf{X}$ is decomposed into scores represented by a matrix $\mathbf{T}$ of dimension n x k, and into loadings $\mathbf{P}$ of dimension p x k, where k represents the number of latent variables retained for the model. Similarly, $\mathbf{y}$ is decomposed into a matrix of scores $\mathbf{U}$ and loadings $\mathbf{q}$ of dimension n x k and of 1 x k. This intermediate variables can be defined by these equations:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E}_x \tag{4}$$

$$\mathbf{y} = \mathbf{U}\mathbf{q}^t + \mathbf{e}_y \tag{5}$$

Where $\mathbf{T}$ and $\mathbf{U}$ are the scores of $\mathbf{X}$ and $\mathbf{y}$ respectively. $\mathbf{P}$ and $\mathbf{q}$ represent loadings for $\mathbf{X}$ and $\mathbf{y}$ respectively. $\mathbf{E}_x$ and $\mathbf{e}_y$ represent residuals in the decomposition of $\mathbf{X}$ and $\mathbf{y}$.

2.5.3. SO-PLS predictions with both spectral and climate data

SO-PLS (Naes et al., 2011) regression is a multi-block method where prediction model is built sequentially from each data block. First, the SO-PLS algorithm started as PLS method with the first block containing spectral data, as previously described (eq. 3).

Then, an orthogonalisation procedure was performed to remove information (already exploited from the first regression) on the second block containing climate data, defined by the matrix $\mathbf{Z}$. This orthogonalisation, providing $\mathbf{Z}_\perp$, can be written as follows:

$$\mathbf{Z}_\perp = \mathbf{Z} - \mathbf{T}(\mathbf{T}^t\mathbf{T})^{-1}\mathbf{T}^t\mathbf{Z} \tag{6}$$

Where $\mathbf{T}$ represented scores of $\mathbf{X}$ described eq. 4. Then, a second PLS model is established between the residual matrix, corresponding to the matrix $\mathbf{r}_X$ (eq. 3) and the matrix $\mathbf{Z}_\perp$. This regression is established by following the same procedure as previously for the regression between $\mathbf{X}$ and $\mathbf{y}$ (eq. 3 4 and 5). At the end of this procedure, a vector $\mathbf{c}$ containing the regression coefficients is obtained. The final equation of the SO-PLS multiblock method can be written as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_{\perp}\mathbf{c} + \mathbf{r}_{X,Z} \tag{7}$$

With $\mathbf{r}_{X,Z}$, the residual matrix of the SO-PLS model.

2.6. Evaluation strategies of prediction models

The spectral data set was divided into two independent data sets: a calibration set of 106 images and a test set. The test set was formed with the 54 remaining hyperspectral images. This test set was constructed to reflect all modalities of the experimental design.

A cross-validation step was performed to select the number of latent variables per block using a k-fold validation procedure (Camacho and Ferrer, 2012) performed with five blocks repeated twice. The maximum number of latent variables was set at 20 for the spectral data and 3 for the climate data.

The validation errors were then calculated and used to determine the optimal number of latent variables. The parameters chosen for the evaluation of the models are the root-mean-square error (RMSE), the bias and the coefficient of determination R^2. These parameters were calculated as follows:

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - y_m)^2} \tag{10}$$

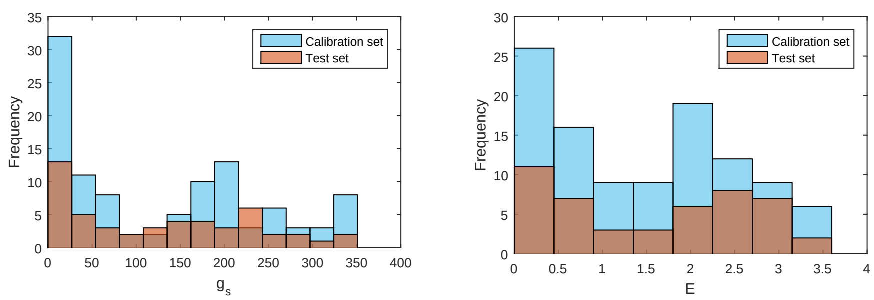
Where $\hat{y}_i$ denotes the predicted value, y_i the observed value, y_m the mean value and N the total number of observations.

3. Results and discussion

3.1. Data visualisation

3.1.1. Y variables: transpiration and stomatal conductance

Value distributions of transpiration (E) and stomatal conductance (g_s) are shown in figure 4b and figure 4a.



(a) g_s (b) E

Figure 4: Histograms of response variable values of (a) stomatal conductance g_s (mmol.m^{-2}.s^{-1}) and (b) transpiration E (mmol.m^{-2}.s^{-1}) .

g_s values range from 0 to 350 mmol.m^{-2}.s^{-1} (fig. 4a). Values above 250 mmol.m^{-2}.s^{-1} correspond to well-watered plants and optimal sun exposition. Then, a high frequency around 200 mmol.m^{-2}.s^{-1} appears. This range of values corresponds to plants in water deficit and/or to lower sun radiation at the beginning or end of the day. A majority of low values between 0 and 25 mmol.m^{-2}.s^{-1} are observed corresponding to closed stomata. This occurs when there is a complete cessation of the photosynthetic process. Plants are then considered as stressed and correspond to individuals whose irrigation has been stopped for a long time.

E values range from 0 to 4 mmol.m^{-2}.s^{-1} (fig. 4b). High values ($\geq$ 2.5 mmol.m^{-2}.s^{-1}) of E correspond to a transpiration level expected when no water stress is applied. Low values, lower than 1, mean that plant transpiration is reduced. With such values during the day, the metabolic activity of the plant is considered as suboptimal.

For both physiological variables, test sets have similar value distributions than those of the corresponding calibration set and cover the whole range of values.

3.1.2. Spectral data

The figure 5 shows average spectra obtained by modality (WD and WW) and by grape variety.

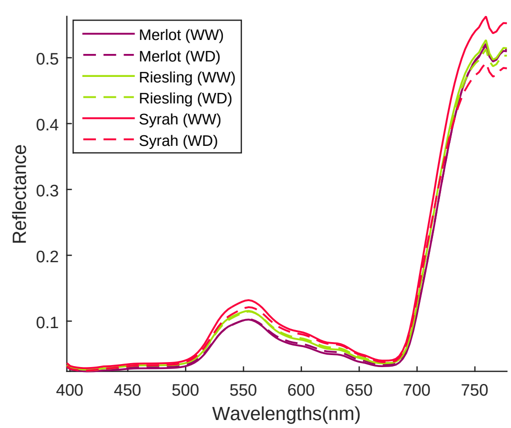


Figure 5: Spectral data, mean spectra per modality

These spectra corresponds to typical vegetation spectra (Xu et al., 2019; Ryckewaert et al., 2021) with specific characteristics at 450 nm, 550 nm, 650 nm related to pigments (carotenoids, chlorophyll and anthocyanins) and the red-edge, which corresponds to a slope between the visible and near-infrared range towards 720 nm.

Average spectra are very close to each other regardless the irrigation regime both for Merlot and Riesling but substantially differ from these of Syrah. Reflectance values are higher for Syrah over the whole spectrum when irrigated but lower in the NIR domain when non-irrigated.

The spectrum in irrigated condition differs from the spectrum in non-irrigated condition only for Syrah. The visualisation of average spectra is not sufficient to observe significant differences between the irrigation conditions for other varieties.

3.2. PLS with one block using spectral data

3.2.1. Cross-validation procedure

Table 1 shows criterion values obtained with the cross-validation step to predict variables E and g_s using the PLS method.

Table 1: PLS criteria obtained after the cross validation procedure.

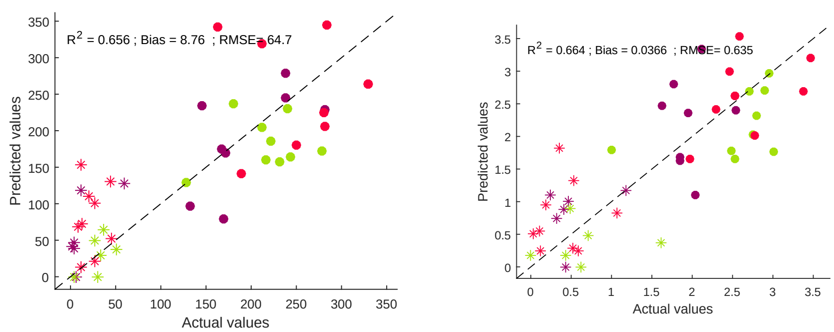
Variable	LV	R^2_{cv}	$bias_{cv}$ ($mmol.m^{-2}.s^{-1}$)	$RMSE_{cv}$ ($mmol.m^{-2}.s^{-1}$)
g_s	9	0.639	5.41	67.3
E	8	0.625	-0.02	0.67

With the PLS method, the cross-validation suggests 8 LV for E and 9 LV for g_s. For g_s, R^2_{cv}, $bias_{cv}$ and $RMSE_{cv}$ have values of 0.639, 5.41 $mmol.m^{-2}.s^{-1}$ and 67.3 $mmol.m^{-2}.s^{-1}$ respectively. For E, R^2_{cv}, $bias_{cv}$ and $RMSE_{cv}$ have values of 0.625, -0.02 $mmol.m^{-2}.s^{-1}$ and 0.67 $mmol.m^{-2}.s^{-1}$ respectively. $RMSE_{cv}$ values obtained for the prediction of these two variables are to be compared with the observed values (fig. 4a and 4b).

3.2.2. Model evaluation

.

E and g_s PLS models calibrated with calibration set are applied to the independent test set. Figures 6b and 6a show predicted values according to observed values.



(a) (b)

Figure 6: PLS-model evaluation on the test set of (a) stomatal conductance g_s (mmol.m^{-2}.s^{-1}) and (b) transpiration E (mmol.m^{-2}.s^{-1}). Symbol indicates irrigation condition: • well-watered and * without irrigation. Colour identifies varieties: red: Merlot; green: Riesling; violet: Syrah.

Figure 6a shows criterion values obtained for g_s prediction. The separation between the irrigated and non-irrigated modalities is clearly observed. For this variable, criteria (R^2= 0.656, bias=8.76 mmol.m^{-2}.s^{-1}, RMSE=64.7 mmol.m^{-2}.s^{-1}) are close to the calibration model with a larger bias (tab. 1). Some observed values below 50 mmol.m^{-2}.s^{-1} seem more difficult to predict. On the other hand, above 50 mmol.m^{-2}.s^{-1}, satisfying predictions are obtained.

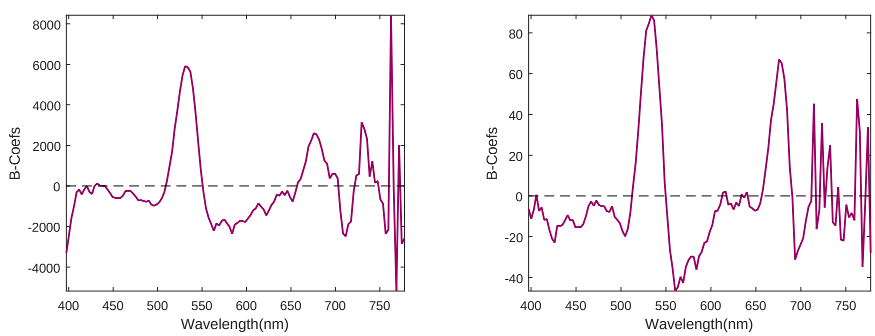
As previously observed, separation between the irrigated and non-irrigated modalities is clearly identified for E. Criteria values obtained for E prediction are 0.664 for R^2, -0.0366 $mmol.m^{-2}.s^{-1}$ for the bias and 0.635 $mmol.m^{-2}.s^{-1}$ for the RMSE. These values are close to values obtained during the cross-validation procedure (tab. 1).

Error values of these two variables are sufficient to identify occurrence of water stress on plants.

It is interesting to note that the error is of the same nature for irrigated plants, at the beginning of desiccation, or in more severe desiccation, *i.e.* in the complete range of variation of the studied variables.

3.2.3. Regression coefficients: contribution of the different wavelengths to models

Figure 7 shows regression coefficients (B-coefficients) of the two PLS models to predict g_s (fig. 7a) and E (fig. 7b).



(a) (b)

Figure 7: Regression coefficients of the PLS-models predicting (a) stomatal conductance g_s and (b) transpiration E.

Regression coefficients are given by PLS models according eq. 3. These coefficients provide the contributions of wavelengths or spectral regions considered in the PLS model.

For g_s, a very high peak is observed around 531 nm. At this wavelength, spectra are sensitive to anthocyanin content (Ryckewaert et al., 2021). Thus, difference observed in the stomatal conductance values is probably related to a difference in anthocyanin content in leaves. An acute metabolic use of anthocyanins or a disturbance in the xanthophyll cycle seems to occur when water status deteriorates. Besides, high values of these regression coefficients are also observed around 680 nm. This spectral region is known to be related to chlorophyll content (Ryckewaert et al., 2021). As a consequence, the combination of spectral information related to chlorophyll and anthocyanin contents seem to be important for stomatal conductance predictions.

Negative peaks are also visible. A negative peak is visible at 714 nm, corresponding to the middle of the red-edge slope (fig. 5). Another negative contribution can be found in the spectral region at 570 nm. This wavelength is often used as a reference to 533 nm to compute the Photochemical Reflectance Index. Beyond 750 nm, regression coefficients vary rapidly sometimes with abrupt reversals of sign, suggesting the complexity or the absence of a significant interpretation.

In the case of E (fig. 7b), two positive peaks are observed around 533 nm and 675 nm. These peaks are located at the same wavelengths previously observed (fig. 7a). However, slight differences are noticeable for the variable E such as a change in ratio between peak values at 533 nm and 675 nm or a sign change occurring at a lower wavelength (before 700 nm for E coefficients and after 700 nm for g_s coefficients). Negative peaks are also visible as described for g_s around 570 nm. Other negative coefficients are visible in the carotenoid region, around 426 nm. This result could be related to the decrease in concentration of some carotenoids with increasing water stress as reported in other works (Salazar-Parra et al., 2015). In the same figure (fig. 7b), noise seems to appear in the near-infrared region between 700 and 800 nm.

The anthocyanin and chlorophyll contents are the two pigments most closely related to the values obtained for transpiration E and stomatal conductance g_s.

When experiencing sudden water deficit, stomatal conductance changes very rapidly. A complex relationship between anthocyanin and chlorophyll occurs as a response to this water stress and is linked to stomatal conductance to regulate the photosynthetic process. These results are consistent with the well-known effect of water stress on stomatal closure (and the resulting decrease in transpiration), but also on chlorophyll degradation and xanthophyll cycling (Doupis et al., 2020). The stability of the model for predicting g_s or E from hyperspectral data therefore depends *a priori* on the stability of the relationships between g_s or E and pigment concentrations. The behaviour of the model for the three varieties, suggests a similar evolution of their physiological characteristics, that influence the reflectance spectrum in response to the water regime.

3.3. SO-PLS using a second block with climate data

3.3.1. Cross-validation procedure

Table 2 shows results of optimal parameters obtained with cross-validation step for the prediction of variables g_s and E using the SO-PLS multi-block method.

Table 2: Number of latent variables (LV) obtained after cross validation procedure for the first block (spectral data) and the second block (climate data).

Variable	LV 1^{st} block (spectral data)	LV 2^{nd} block (climate data)	R^2_p	$bias_p$ ($mmol.m^{-2}.s^{-1}$)	$RMSE_p$ ($mmol.m^{-2}.s^{-1}$)
g_s	9	0	0.639	5.41	67.3
E	8	3	0.684	0.02	0.613

Regarding g_s model, cross-validation suggests 9 latent variables for the first block and 0 for the second block corresponding respectively to spectral data and climate data. Information from the second block does not improve the g_s prediction model. SO-PLS parameterisation then corresponds to the PLS model calculated previously using only the spectral data (table 1).

This result is surprising because radiation level is known to influence stomatal opening (Jones, 2013). However, considering the environmental conditions of the measurements, only relatively high level of radiation (exposed leaves and sunny days) were encountered. Therefore, the stomatal conductance was not driven by this factor during experiments.

Regarding E model, the number of latent variables retained is 8 for the first block and 3 for the second block. In this case, both blocks are exploited to estimate the prediction model. The climate data seem to provide additional information to the spectral data for transpiration prediction. This result is consistent with the fact that transpiration depends on the one hand on g_s and on the other hand on the evaporative capacity of the air, which is itself determined by relative humidity, temperature, radiation and wind speed (Jones, 2013).

From a theoretical point of view, a relationship was primarily expected between VIS-NIR signature and leaf water status (with possible, additional influences of changes in leaf pigment composition and other constituents during soil drying). Considering that stomatal conductance fairly well correlates with leaf water status (Damour et al., 2010), our observation that stomatal conductance also correlated with VIS-NIR characteristics is in line with the theoretical expectation. Regarding transpiration, which is roughly equal to the stomatal conductance multiplied by the evaporative demand (mostly vapour pressure deficit), a stronger influence of climate on the relationship with VIS-NIR characteristics could also be expected.

3.3.2. Model evaluation

As mentioned before, SO-PLS model of g_s corresponds to PLS model previously studied in section 3.2.2 (fig. 6a and 7a).

SO-PLS results for E prediction are shown in figure 8.

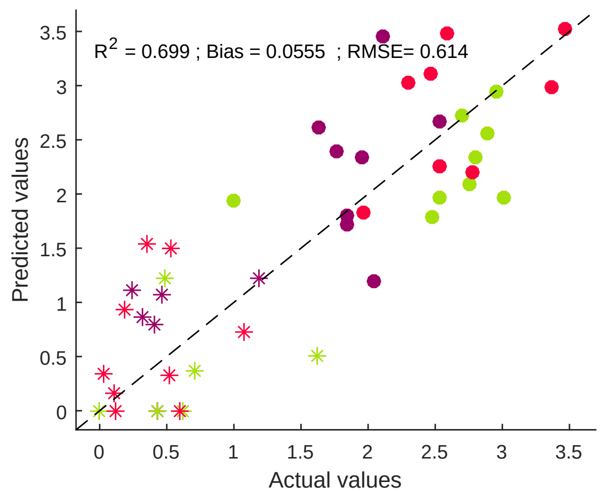
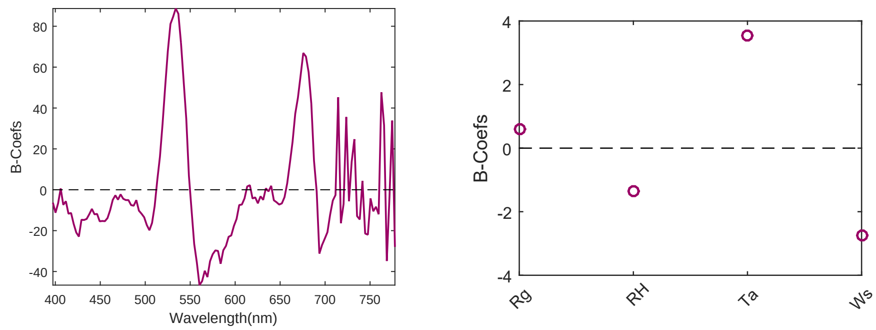


Figure 8: Model evaluation on a test set for the prediction of transpiration E (mmol.m^{-2}.s^{-1}). Symbol indicates irrigation condition: • well-watered and * without irrigation. Colour identifies varieties: red: Merlot; green: Riesling; violet: Syrah

R^2, bias and RMSE criteria obtained have values of 0.699, 0.0555 mmol.m^{-2}.s^{-1} and 0.614 mmol.m^{-2}.s^{-1}, respectively. These criteria are improved compared to the results obtained with PLS method (fig. 6b). Including the second block corresponding to climate data in the model improves the E prediction. In a perspective of use for agronomic diagnosis, it is worth noticing that the same model is used for all three grape varieties tested here, despite the slightly different spectral signature of Syrah.

3.3.3. Regression coefficients for each block

Figures 9a and 9b show regression coefficients for spectral and climate data, respectively.



(a) (b)

Figure 9: Regression coefficients of SO-PLS model to predict transpiration E (a) for the first block (spectral data) and (b) for the second block (climate data)

These vectors show variables involved in E prediction. For the first block corresponding to spectral data (fig. 9a), the same coefficients are obtained as previously (fig. 7b). Indeed, models are established with 8 latent variables in both cases.

Coefficients obtained from the second block show non-zero values for all four climate variables: radiation (Rg), relative humidity (RH), temperature (Ta) and wind speed (Ws). These parameters are then taken into account in the prediction model.

Both coefficients related to Rg and Ta have positive signs while those of RH and Ws have negative signs. This shows that for given spectral characteristics, an increase of Rg or Ta induces an increase of E. This result could be expected as the evaporative capacity of the air increases with Rg and Ta. It can be noted, however, that Rg has a lower impact compared to Ta. This was probably due to a lower range of variation of Rg at the time of our measurements (as commented above). Conversely, increasing values of RH and Ws will tend to decrease transpiration. This result was also expected for RH, because when RH increases, the evaporative capacity of the air decreases. The effect of wind is more complex, because, on the one hand, it decreases leaf temperature, which reduces transpiration. But, on the other hand, it increases aerodynamic conductance, which increases transpiration. It seems that the first effect dominates here. This effect should be tested under other radiation conditions (especially lower level of radiation on leaves) and other wind speed ranges.

Furthermore, absolute values of these coefficients show the impact of the associated variables in the prediction model. Thus, Ta and Ws have a greater impact than Rg and RH on transpiration.

4. Conclusion

This article proposed a study of prediction models established from spectral data, for two major variables related to water stress, namely stomatal conductance g_s and transpiration E. Despite different coloured berry varieties (one white and two red), generic PLS models achieved good predictive quality. Quality of these prediction models could be improved by defining varietal models on a larger data set. Combining predictions of these two variables is a promising solution to assess plant water stress.

In addition, merging spectral data with climate data improves prediction quality of the transpiration variable. Moreover, if additional information from other sensors is available, multi-block methods could improve predictive qualities of physiological variables. The proposed methodology enables to consider coupling spectral point acquisitions with connected objects in the field in order to improve the prediction of agronomic variables.

Acknowledgement

This work has benefited from a financial support from the InterregSudoe under the reference SOE3/P2/E0911.

References

David B. Lobell, Wolfram Schlenker, and Justin Costa-Roberts. Climate trends and global crop production since 1980. *Science*, 333(6042):616–620, 2011. URL http://science.sciencemag.org/content/333/6042/616.short.

Ron Mittler. Abiotic stress, the field environment and stress combination. *Trends in Plant Science*, 11(1):15–19, January 2006. ISSN 1360-1385. doi: 10.1016/j.tplants.2005.11.002. URL https://www.sciencedirect.com/science/article/pii/S1360138505002918.

J. E. Olesen, M. Trnka, K. C. Kersebaum, A. O. Skjelvåg, B. Seguin, P. Peltonen-Sainio, F. Rossi, J. Kozyra, and F. Micale. Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, 34(2):96–112, February 2011. ISSN 1161-0301. doi: 10.1016/j.eja.2010.11.003. URL https://www.sciencedirect.com/science/article/pii/S1161030110001061.

Spyros Fountas, Borja Espejo-Garcia, Aikaterini Kasimati, Nikolaos Mylonas, and Nicoleta Darra. The Future of Digital Agriculture: Technologies and Opportunities. *IT Professional*, 22(1):24–28, January 2020.

ISSN 1520-9202, 1941-045X. doi: 10.1109/MITP.2019.2963412. URL https://ieeexplore.ieee.org/document/8994135/.

Zhaoyu Zhai, José Fernán Martínez, Victoria Beltran, and Néstor Lucas Martínez. Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170:105256, March 2020. ISSN 0168-1699. doi: 10.1016/j.compag.2020.105256. URL https://www.sciencedirect.com/science/article/pii/S0168169919316497.

Thierry Simonneau, Eric Lebon, Aude Coupel-Ledru, Elisa Marguerit, Landry Rossdeutsch, and Nathalie Ollat. Adapting plant material to face water stress in vineyards: which physiological targets for an optimal control of plant water status? *OENO One*, 51(2):167, 2017. doi: 10.20870/oeno-one.2016.0.0.1870. URL https://hal.archives-ouvertes.fr/hal-01602544. Publisher: Institut des Sciences de la Vigne et du Vin (Université de Bordeaux).

Gaëlle Damour, Thierry Simonneau, Hervé Cochard, and Laurent Urban. An overview of models of stomatal conductance at the leaf level. *Plant, Cell & Environment*, 33(9):1419–1438, 2010. ISSN 1365-3040. doi: 10.1111/j.1365-3040.2010.02181.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2010.02181.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3040.2010.02181.x.

H. R. Schultz and M. Stoll. Some critical issues in environmental physiology of grapevines: future challenges and current limitations. *Australian Journal of Grape and Wine Research*, 16(s1):4–24, 2010. ISSN 1755-0238. doi:

10.1111/j.1755-0238.2009.00074.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-0238.2009.00074.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1755-0238.2009.00074.x.

Giuseppe Romano, Shamaila Zia, Wolfram Spreer, Ciro Sanchez, Jill Cairns, Jose Luis Araus, and Joachim Müller. Use of thermography for high throughput phenotyping of tropical maize adaptation in water stress. *Computers and Electronics in Agriculture*, 79(1):67–74, October 2011. ISSN 01681699. doi: 10.1016/j.compag.2011.08.011. URL http://linkinghub.elsevier.com/retrieve/pii/S0168169911001888.

Zheng Zhou, Yaqoob Majeed, Geraldine Diverres Naranjo, and Elena M. T. Gambacorta. Assessment for crop water stress with infrared thermal imagery in precision agriculture: A review and future prospects for deep learning applications. *Computers and Electronics in Agriculture*, 182:106019, March 2021. ISSN 0168-1699. doi: 10.1016/j.compag.2021.106019. URL https://www.sciencedirect.com/science/article/pii/S0168169921000375.

Jun-Li Xu, Alexia Gobrecht, Daphné Héran, Nathalie Gorretta, Marie Coque, Aoife A. Gowen, Ryad Bendoula, and Da-Wen Sun. A polarized hyperspectral imaging system for in vivo detection: Multiple applications in sunflower leaf analysis. *Computers and Electronics in Agriculture*, 158:258–270, March 2019. ISSN 0168-1699. doi: 10.1016/j.compag.2019.02.008. URL http://www.sciencedirect.com/science/article/pii/S0168169918314455.

Maxime Ryckewaert, Nathalie Gorretta, Fabienne Henriot, Alexia Gobrecht, Daphné Héran, Daniel Moura, Ryad Bendoula, and Jean-Michel Roger. Potential of high-spectral resolution for field phenotyping in plant breeding: Application to maize under water stress. *Computers and Electronics in Agriculture*, 189:106385, October 2021. ISSN 0168-1699. doi: 10.1016/j.compag.2021.106385. URL https://www.sciencedirect.com/science/article/pii/S0168169921004026.

Antonio José Steidle Neto, Daniela C. Lopes, Francisco A. C. Pinto, and Sérgio Zolnier. Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. *Biosystems Engineering*, 155:124–133, March 2017. ISSN 1537-5110. doi: 10.1016/j.biosystemseng.2016.12.008. URL https://www.sciencedirect.com/science/article/pii/S1537511015304505.

Qianxuan Zhang, Qingbo Li, and Guangjun Zhang. Rapid Determination of Leaf Water Content Using VIS/NIR Spectroscopy Analysis with Wavelength Selection. *Spectroscopy: An International Journal*, 27(2): 93–105, May 2012. ISSN 2314-4920. doi: 10.1155/2012/276795. URL https://www.hindawi.com/journals/jspec/2012/276795/. Publisher: Hindawi.

Tal Rapaport, Uri Hochberg, Maxim Shoshany, Arnon Karnieli, and Shimon Rachmilevitch. 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, November 2015. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2015.

09.003. URL https://www.sciencedirect.com/science/article/pii/S092427161500204X.

Phuong D. Dao, Yuhong He, and Cameron Proctor. Plant drought impact detection using ultra-high spatial resolution hyperspectral images and machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 102:102364, October 2021. ISSN 0303-2434. doi: 10.1016/j.jag.2021.102364. URL https://www.sciencedirect.com/science/article/pii/S0303243421000714.

David J. Mulla. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114 (4):358–371, April 2013. ISSN 1537-5110. doi: 10.1016/j.biosystemseng.2012.08.009. URL https://www.sciencedirect.com/science/article/pii/S1537511012001419.

Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130, 2001.

T. Naes, O. Tomic, B.-H. Mevik, and H. Martens. Path modelling by sequential PLS regression. *Journal of Chemometrics*, 25(1):28–40, January 2011. ISSN 08869383. doi: 10.1002/cem.1357. URL http://doi.wiley.com/10.1002/cem.1357.

F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993. URL http://www.sciencedirect.com/science/article/pii/003442579390013N.

José Camacho and Alberto Ferrer. Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: theoretical aspects. *Journal of Chemometrics*, 26(7):361–373, 2012. ISSN 1099-128X. doi: https://doi.org/10.1002/cem.2440. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.2440. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.2440.

Carolina Salazar-Parra, Iker Aranjuelo, Inmaculada Pascual, Gorka Erice, Álvaro Sanz-Sáez, Jone Aguirreolea, Manuel Sánchez-Díaz, Juan José Irigoyen, José Luis Araus, and Fermín Morales. Carbon balance, partitioning and photosynthetic acclimation in fruit-bearing grapevine (Vitis vinifera L. cv. Tempranillo) grown under simulated climate change (elevated CO2, elevated temperature and moderate drought) scenarios in temperature gradient greenhouses. *Journal of Plant Physiology*, 174:97–109, February 2015. ISSN 0176-1617. doi: 10.1016/j.jplph.2014.10.009. URL https://www.sciencedirect.com/science/article/pii/S0176161714002922.

George Doupis, Konstantinos S. Chartzoulakis, Demetris Taskos, and Angelos Patakas. The effects of drought and supplemental UV-B radiation on physiological and biochemical traits of the grapevine cultivar "Soultanina". *OENO One*, 54(4):687–698, October 2020. ISSN 2494-

1271. doi: 10.20870/oeno-one.2020.54.4.3581. URL https://oeno-one.eu/article/view/3581. Number: 4.

Hamlyn G. Jones. *Plants and Microclimate: A Quantitative Approach to Environmental Plant Physiology*. Cambridge University Press, Cambridge, 3 edition, 2013. ISBN 978-0-521-27959-8. doi: 10.1017/CBO9780511845727. URL https://www.cambridge.org/core/books/plants-and-microclimate/A7420295742E44198D8984F21D921058.