Toward a generalized predictive model of grapevine water status in Douro region from hyperspectral data

Isabel Pôças a,b,c,d,⁎, Renan Tosina a,d, Igor Gonçalves e, Mario Cunha a,b,d

Abstract
The predawn leaf water potential ($\psi_{pd}$) is an eco-physiological indicator widely used for assessing vines water status and thus supporting irrigation management in several wine regions worldwide. However, the $\psi_{pd}$ is measured in a short time period before sunrise and the collection of a large sample of points is necessary to adequately represent a vineyard, which constitute operational constraints. In the present study, an alternative method based on hyperspectral data derived from a handheld spectroradiometer and machine learning algorithms was tested and validated for assessing grapevine water status. Two test sites in Douro wine region, integrating three grapevine cultivars, were studied for the years of 2014, 2015, and 2017. Four machine learning algorithms were tested for predicting the $\psi_{pd}$ as a continuous variable, namely Random Forest (RF), Bagging Trees (BT), Gaussian Process Regression (GPR), and Variational Heteroscedastic Gaussian Process Regression (VH-GPR). Three predicting variables, including two vegetation indices (NRI554,561 and WI900,970) and a time-dynamic variable based on the $\psi_{pd}$ ($\psi_{pd,t}$), were applied for modelling the response variable ($\psi_{pd}$). Additionally, the predicted values of $\psi_{pd}$ were aggregated into three classes representing different levels of water deficit (low, moderate, and high) and compared with the corresponding classes of $\psi_{pd}$ observed values. A root mean square error (RMSE) and a mean absolute error (MAE) lower or equal than 0.15 MPa and 0.12 MPa, respectively, were obtained with an external validation data set ($n = 71$ observations) for the various algorithms. When the modelling results were assessed through classes of values, a high overall accuracy was obtained for all the algorithms (82–83%), with prediction accuracy by class ranging between 79% and 100%. These results show a good performance of the predictive models, which considered a large variability of climatic, environmental, and agronomic conditions, and included various grape cultivars. By predicting both continuous values of $\psi_{pd}$ and classes of $\psi_{pd}$, the approach presented in this study allowed obtaining 2-levels of accurate information about vines water status, which can be used to feed management decisions of different types of stakeholders.

1. Introduction
The vineyard is traditionally rainfed in most parts of the world. However, many wine-producing regions undergo seasonal drought coincident with the grapevine growing season, and an increase in aridity is additionally foreseen related to climate change scenarios (Chaves et al., 2010; Flexas et al., 2010; Fraga et al., 2018). Consequently, in the last decades, deficit irrigation strategies have been adopted in critical grapevine phenological stages aiming to stabilize wine production and quality (Costa et al., 2016; Pisciotta et al., 2018; Serrano et al., 2010).

Nevertheless, water resources are increasingly under pressure due to the population growth and changes in lifestyle and diets, along with climate change and climate variability, with increasing occurrence of drought events (HLPE, 2015; Pereira, 2017). Therefore, the improved water management is now one of the most important challenges in agriculture, including in viticulture, and the enhancement of irrigation management is paramount within this context.
A key aspect for the improved irrigation management in viticultural systems relies on the accurate monitoring of the vine water status. The irrigation scheduling is often based on eco-physiological indicators of the vines response to water deficit, integrating both the plant and the environment influence (De Bei et al., 2011; Rodrigues et al., 2012; Williams et al., 2012). The predawn leaf water potential ($\psi_{pd}$) is one of these eco-physiological indicators, being considered adequate for assessing vines water status (Alves et al., 2012; Lopes et al., 2011; Williams and Araujo, 2002) and thus widely used to support irrigation management in several wine regions of the world. The $\psi_{pd}$ approximates the water potential of the soil before the sunrise and correspondingly the measurement period is restricted to a short daily time window, with operational constraints when the collection of a large sample of points is needed to adequately represent a vineyard. Additionally, some studies have shown a large variability of $\psi_{pd}$ measurements at the block and vineyard level, particularly in conditions of high water restriction, resulting in the need of an increased sampling density in order to support precise irrigation scheduling (e.g., Ojeda et al., 2005; Taylor et al., 2010). Therefore, it is extremely important setting alternative methods suitable for reliably and operationally assessing the plant water status of large vineyards in order to support precision viticulture production systems.

The spectral data has been increasingly used for retrieving information about vegetation, including greenness (e.g., leaf area index), dynamics (e.g., phenology), physiology (e.g., pigments content), and plant conditions (e.g., water status) (e.g., Mananze et al., 2018; Marshall et al., 2016; Üstün et al., 2009; Verrelst et al., 2015; Zarco-Tejada et al., 2013). The wavelengths of 970 nm, 1200 nm, 1450 nm, 1930 nm, and 2500 nm, from the near infrared (NIR) and mid infrared regions of the electromagnetic spectrum, are well recognized by their strong water absorption of radiation and thus represent natural regions for assessing crop water status (Roberto et al., 2012). The thermal infrared spectra has also been considered for assessing plants water status and water stress signs through the canopy temperatures (e.g., Bellvert et al., 2014; Buitrago et al., 2016; Neinavaz et al., 2017). However, the availability of remote sensing data in these spectral domains is still limited, particularly concerning hyperspectral data. Sensing the vegetation through hyperspectral data has proven advantages compared to broadband (multispectral) data, including for detecting plant stress, as discussed by Thenkabail et al. (2012). Therefore, several studies have focused on using hyperspectral data from the visible and NIR, which are more easily available and allow analyzing stress indicators that are considered proxy for the crop water status, e.g., the xanthophyll pigment cycle or the chlorophyll fluorescence (Hernández-Clemente et al., 2011; Middleton et al., 2012; Pôças et al., 2017; Zarco-Tejada et al., 2013).

Several approaches have been considered for estimating crop biophysical parameters using hyperspectral data, including physically-based and statistically-based methods (e.g., Clevers et al., 2010; Lázaro-Gredilla et al., 2014; Mananze et al., 2018). The physically-based approaches (inversion of radiative transfer models) establish a cause-effect relationship grounded on physical knowledge and are considered sound methodologies, although its application is rather challenging (Berger et al., 2018; Verrelst et al., 2015). Alternatively, the statistically-based approaches are less input demanding than physically-based models, while producing good results for retrieving several biophysical parameters, including plant water status (Mirzaie et al., 2014; Pôças et al., 2017, 2015; Rossini et al., 2013; Zarco-Tejada et al., 2013), and thus are more suitable for operational applications. Hence, the present study focuses on statistical approaches.

The statistically-based methodologies can be either parametric or non-parametric (Verrelst et al., 2015). The parametric models assume an explicit relationship between the spectral data (predictors) and the biophysical parameter (response variable), while non-parametric models assume a flexible approach to exploit the data and do not rely on a predefined relationship (Kuhn and Johnson, 2013). The parametric models are often based on spectral vegetation indices (VIs) that are mathematical combinations of a few selected spectral bands to describe the biophysical parameter (Jones and Vaughan, 2010). The VIs have been considered for approximating a large set of parameters related with plant water status, including plant water content, leaf water potential, and equivalent water thickness (Casas et al., 2014; González-Fernández et al., 2015b; Pôças et al., 2015; Rallo et al., 2014).

The machine learning algorithms, which represent non parametric methods, are often applied for the predictive modelling of crop biophysical parameters due to their potential to generate robust relationships between predictors and response variables with complex and non-linear patterns (e.g., crop water status). In machine learning based models, a training data set is used to learn the data patterns and train the model, which is further tested and assessed using a validation dataset. A large set of machine learning algorithms have been developed and applied for predicting biophysical parameters using hyperspectral data, e.g., random forest (e.g., Doktor et al., 2014; Pôças et al., 2017), bagging trees (e.g., Verrelst et al., 2015), partial least squares regression (e.g., González-Fernández et al., 2015a; Rapaport et al., 2015), Gaussian process regression (e.g., Lázaro-Gredilla et al., 2014; Verrelst et al., 2012a, 2015), artificial neural networks (Krishna et al., 2019; Neinavaz et al., 2017). The ultimate goal of the different types of machine learning algorithms is to develop a model that makes an accurate prediction. Several studies present predictive models specifically developed for estimating the crop water status based on spectral reflectance data. However, such models often rely on small sample sizes, corresponding to a few dates of field measurements (e.g., Romero et al., 2018), or a large number of predicting variables (e.g., González-Fernández et al., 2015a; Rallo et al., 2014), and thus their applicability is limited. Additionally, an important issue for the generalization of predictive models is associated with the validation process. Often, the validation of the predictive models is performed with small size independent data sets or through cross validation procedures (e.g., Pôças et al., 2017; Rallo et al., 2014), limiting its generalized application on an operational basis. Furthermore, it is important taking into account common inherent trends that may occur between response variables and predictors, which should be minimized to avoid spurious associations. For example, the hyperspectral data and the $\psi_{pd}$ of grapevines share a presumable common downward trend throughout the post-flowering period, which is due to phenology/vegetation growth and climatic conditions, respectively (Pôças et al., 2017). Therefore, to minimize the plausible common trend effects between time series (e.g., $\psi_{pd}$ and hyperspectral data), the predictor variables should be detrended or a time-varying predictor should be included in the model (Cunha and Richter, 2014).

The present study aims to develop and validate a simple generalized model for predicting vines water status based on hyperspectral data in Douro wine region. Specific goals include: (i) testing and validating multiyear modelling approaches on a diversified set of climatic, environmental, grape varieties and agronomic conditions; (ii) comparing the performance of various modelling approaches based on regression model considering different levels of water deficit; (ii) predicting two levels of vines water status data, i.e., numeric continuous values vs. classes of values, to assist different types of stakeholders.

2. Material and methods

2.1. Study area

The study area is located in Douro Wine Region, Northeast of Portugal (Fig. 1), where vineyards are mainly built over terraces and slopes with soils mostly derived from shale. The vineyards represent one of the most important features of Douro landscape, counting for 18.3% of the region total area. The Douro Wine Region covers around 250,000 ha, divided into three sub-regions: Baixo Corgo, Cima Corgo, and Douro Superior, distributed from the western up to eastern part of...
the region, and representing 32.2%, 22.0%, and 9.3% of land cover by vineyard, respectively (Fig. 1). This region, famous for its Port wine, is responsible for about one fourth (23%) of all wine produced in Portugal (IVDP, 2018), and its vineyard landscape is considered World Heritage by the UNESCO since 2001.

Overall, the region is characterized by a Mediterranean climate, with high temperatures and low precipitation values during summer period, resulting in frequent water deficit occurrence. The annual precipitation amount is 856 mm in Baixo Corgo, 658 mm in Cima Corgo and 539 mm in Douro Superior, with precipitation amounts during summer period corresponding to 10.3%, 8.8% and 7.4% of the annual precipitation, respectively (INMG, 1991). The average temperature in the summer is higher in Douro Superior and Cima Corgo (24.1 °C) compared to Baixo Corgo (22.4 °C) (INMG, 1991). A more detailed characterization of the region and sub-regions climate is presented by Pôças et al. (2017).

Two test sites, integrated in commercial vineyards, were considered for the study (Fig. 1): (i) Quinta dos Aciprestes (wine company Real Companhia Velha) located in Cima Corgo sub-region (Test site 1; Latitude 41.21° N; Longitude 7.43° W; 145 m a.s.l.), and (ii) Quinta do Ataíde (wine company Symington Family Estates), in Douro Superior (Test site 2; Latitude 41.25° N; Longitude 7.11° W; 161 m a.s.l.). Table 1 summarizes the vineyards overall characteristics in the test sites.

Two cultivars were studied in test site 1: (i) Touriga Nacional (TN; years 2014, 2015, and 2017), and (ii) Touriga Franca (TF; years 2015 and 2017). Two plots, with two replicate areas, were sampled for TN, each plot including three irrigation treatments: non-irrigated (TN_NI), irrigation treatment 1 (TN_IT1), and irrigation treatment 2 (TN_IT2). A single plot and a single irrigation treatment (TF_IT) were considered for TF.

In test site 2, four plots, covering three cultivars, were studied in the years 2015 and 2017: two plots of TN (TN1 and TN2), one plot of TF, and one plot of Tinta Barroca (TB). Two irrigation treatments were sampled for TN and TF: (i) Irrigated treatment: TN1_IT, TN2_IT, TF_IT, and (ii) non-irrigated treatment: TN1_NI, TN2_NI, TF_NI; and for TB only an irrigated treatment was sampled (TB_IT).

The irrigation amounts/dates per test site, plot, and irrigation treatment (Table 2) were managed by each wine company, following the \( \psi_{pd} \) regular measurements and aiming to adjust for quality criteria.

2.2. Ground measurements

The study was implemented in the years 2014, 2015, and 2017, with ground data being collected in the vineyards of both test sites. In the year 2014, no data were collected in the test site 2, while in the test site 1 only data of cultivar TN were collected.

The ground measurements were done between post-flowering and harvest, roughly between June and September. The climatic conditions during the ground measurements period presented large variability among years. The year 2017, in both test sites, was characterized by high temperatures and very low precipitation during the summer period, while 2014 was the coldest and wet year (Table 3). This climatic pattern among years can explain the irrigation amounts presented in Table 2 that increase from 2014 to 2017.

2.2.1. Predawn leaf water potential (\( \psi_{pd} \))

The \( \psi_{pd} \) measured with a pressure chamber (Scholander et al., 1965) (PMS600, Albany, OR, USA) was used as a reference for the plants water status. A minimum of six plants per plot were sampled in each test site. A total of 21 measurement dates was considered, five in 2014 (Test site 1) and eight both in 2015 and 2017 (four in test site 1 and four in test site 2), as shown in Table 3.

The \( \psi_{pd} \) data set presents a large intra and inter annual variability (Fig. 2(a)), covering, in each studied year, all the vineyard water deficit conditions (Fig. 2(b)) defined by Deloire et al. (2005): (i) None up to mild water deficit (0 MPa > \( \psi_{pd} \) > −0.2 MPa); (ii) mild to moderate (−0.2 MPa > \( \psi_{pd} \) > −0.4 MPa); (iii) moderate to high (−0.4 MPa > \( \psi_{pd} \) > −0.6 MPa); and (iv) high (−0.6 MPa > \( \psi_{pd} \)). The year 2014 presented lower water deficit conditions while the higher water deficit values were recorded in the year 2017 and intermediate values were observed in 2015 (Fig. 2(a)). Thus, the years 2014 and 2017 encompassed the largest variability of water deficit conditions.
Table 2
Irrigation period and irrigation amounts (L/Plant/day) per test site and irrigation treatment.

<table>
<thead>
<tr>
<th>Test site</th>
<th>Irrigation period</th>
<th>Irrigation events</th>
<th>Irrigation amount (L/Plant/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN,IT1</td>
<td>26/7-8/8</td>
<td>3/7-25/8</td>
<td>19/6-11/8</td>
</tr>
<tr>
<td>TN,IT2</td>
<td>26/7-8/8</td>
<td>3/7-25/8</td>
<td>19/6-11/8</td>
</tr>
<tr>
<td>TF,IT</td>
<td>26/7-8/8</td>
<td>3/7-25/8</td>
<td>19/6-11/8</td>
</tr>
<tr>
<td>TN1,IT</td>
<td>-</td>
<td>30/6-25/8</td>
<td>20/6-22/8</td>
</tr>
<tr>
<td>TN2,IT</td>
<td>-</td>
<td>30/6-25/8</td>
<td>20/6-22/8</td>
</tr>
<tr>
<td>TF,IT</td>
<td>-</td>
<td>30/6-25/8</td>
<td>24/6-14/8</td>
</tr>
<tr>
<td>TB,IT</td>
<td>-</td>
<td>30/6-25/8</td>
<td>20/6-22/8</td>
</tr>
</tbody>
</table>


Table 3
Ground measurement dates, number of observations in each grape cultivar and climatic conditions during the study period per test site and year.

<table>
<thead>
<tr>
<th>Test site</th>
<th>Measurement dates</th>
<th>Cultivar*</th>
<th>Average Temperature</th>
<th>Precipitation amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test site 1</td>
<td>June 16, June 25, July 5</td>
<td>TN (120), TF (40)</td>
<td>22.7°C</td>
<td>69.8 mm</td>
</tr>
<tr>
<td></td>
<td>June 10, July 16, July 20</td>
<td>TN (96), TF (40)</td>
<td>23.8°C</td>
<td>115.6 mm</td>
</tr>
<tr>
<td></td>
<td>July 26, August 6, August 3</td>
<td>TN (196), TF (36)</td>
<td>24.6°C</td>
<td>19.6 mm</td>
</tr>
<tr>
<td></td>
<td>August 19, September 3, August 31</td>
<td>TN (208), TF (48) and TB (24)</td>
<td>25.0</td>
<td></td>
</tr>
<tr>
<td>Test site 2</td>
<td>June 26, July 4</td>
<td>TN (96), TF (40)</td>
<td>23.8°C</td>
<td>115.6 mm</td>
</tr>
<tr>
<td></td>
<td>July 15, July 21</td>
<td>TN (96), TF (40)</td>
<td>24.6°C</td>
<td>19.6 mm</td>
</tr>
<tr>
<td></td>
<td>August 7, August 4</td>
<td>TN (196), TF (36)</td>
<td>25.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>September 2, September 1</td>
<td>TN (208), TF (48) and TB (24)</td>
<td>25.0</td>
<td></td>
</tr>
</tbody>
</table>

* TN, TF and TB represent the cultivars Touriga Nacional, Touriga Franca and Tinta Barroca, respectively, with the number of plants sampled in parenthesis. The temperature and precipitation refer to the means or sum for measurements period.

within the overall data set.

An one-way analysis of variance (ANOVA) with p-value associated to the Fischer test was performed to compare the mean of $\psi_{pd}$ between years regarding the test sites location, the irrigation treatment, and the cultivars.

2.2.2. Hyperspectral data
The hyperspectral data and $\psi_{pd}$ measurements were both measured in the same plants and dates (Table 3). The hyperspectral data were collected using a portable spectroradiometer (Handheld 2, ASD Instruments, Boulder, CO, USA) maintained approximately 30 cm above canopy and directed vertically downward. The spectroradiometer records reflectance signatures between 325 nm and 1075 nm of the electromagnetic spectrum (thus ranging from visible to NIR), with a wavelength interval of 1 nm. However, reflectance data below 400 nm and above 1010 nm were discarded due to noise occurrence in the spectroradiometer spectral limits.

Measurements were done between 11 h and 13 h (local time) to minimize changes in solar zenith angle, in cloud free days. Prior to canopy spectral data acquisition, a dark current correction was performed and the reflectance of a spectralon (white reference panel) was measured for directly obtaining a reflectance output. Ten repetitions per plant were collected and later averaged to minimize the effect of noise.

2.3. Data processing
The hyperspectral and $\psi_{pd}$ data were averaged per plot and treatment aiming for minimizing the noise effects. Following this procedure, 218 observations of hyperspectral profiles and $\psi_{pd}$ data were obtained. An analysis of outliers was performed resulting in the identification of 28 outliers, which were removed from the study. Therefore, a total of 189 observations were considered, covering different conditions regarding year, location, cultivar, and irrigation. The 189 observations accounted for 54 observations of the year 2014, 71 of 2015, and 64 of 2017.

Hyperspectral data were further processed into spectral VIs. Two VIs tested in a previous study in Douro vineyards (Pôças et al., 2017) were
computed: (i) \( NRI_{554,561} \), which was optimized for estimating \( \psi_{pd} \) (Pôças et al., 2017); and (ii) \( WI_{900,970} \), which was designed for assessing plant water content by Pehuelas et al. (1997). The \( NRI_{554,561} \) is a normalized index (Eq. (1)), while the \( WI_{900,970} \) is a simple ratio of bands (Eq. (2)).

\[
NRI_{i,j} = (b_i - b_j) / (b_i + b_j) \\
WI_{i,j} = b_i / b_j
\]

where \( b_i \) and \( b_j \) correspond to the reflectance in the band wavelengths \( i \) and \( j \), respectively.

A time-dynamic variable based on the \( \psi_{pd} \) was also computed by integrating, in each \( \psi_{pd} \) date to be predicted, the information of previous \( \psi_{pd} \) measurements (\( \psi_{pd,i-1} \)), aimed at assimilating all the available information regarding the crop water status dynamics in the post-flowering – harvest period. Additionally, this \( \psi_{pd,i-1} \) aimed at minimizing the spurious association between \( \psi_{pd} \) and hyperspectral data resulting from a common (downward) time trend inherent to each one of these two types of data.

The \( \psi_{pd,0} \) was defined for each measurement point and measurement date as the \( \psi_{pd} \) value corresponding to the previous measurement. For the first measurement date of each year, the \( \psi_{pd,0} \) was set to 70% of the \( \psi_{pd} \) recorded in that same measurement date, i.e., assuming that in a hypothetical measurement at an earlier date the \( \psi_{pd} \) value would be 30% lower. This percentage was set based on expert-knowledge.

2.4. Modelling approaches

For adequately training the predictive models, the training data set should be representative of the entire sample population, thus representing a broad range of the response variable (\( \psi_{pd} \)) conditions (Kuhn and Johnson, 2013). As shown in Fig. 2a, the year 2017 recorded high \( \psi_{pd} \) values, which were not observed in the previous years (2014 and 2015) and thus could not be learnt by the models if the data of 2014 and 2015 were used for training the models. Therefore, the data from the years 2014 and 2017 were used as training data set (118 observations) aiming to adequately learn the data patterns and train the machine learning algorithms with the larger possible variability of the \( \psi_{pd} \) conditions.

Various machine learning techniques were tested, in the regression mode, for modelling the \( \psi_{pd} \) using a training data set integrating the pairs of concurrent measurements of the \( \psi_{pd} \) and the corresponding values of the predicting variable. Four state-of-the-art machine learning algorithms were tested for predicting the \( \psi_{pd} \) as a continuous variable: (i) Random Forest, RF (Breiman, 2001); (ii) Bagging Trees, BT (Breiman, 1996); (iii) Gaussian Process Regression, GPR (Rasmussen and Williams, 2006); and (iv) Variational Heteroscedastic Gaussian Process Regression, VH-GPR (Lázaro-Gredilla et al., 2014). These machine learning algorithms have been successfully applied in previous studies related with the retrieval of vegetation biophysical parameters (e.g., Lázaro-Gredilla et al., 2014; Pôças et al., 2017; Verrelst et al., 2012a, 2015).

The RF and BT algorithms are tree-based models, while GPR and VH-GPR algorithms are Bayesian statistical inference models (Kuhn and Johnson, 2013; Verrelst et al., 2015).

In the tree-based models, the data are progressively split into smaller groups, more homogenous regarding the response variable, and learning decision rules inferred from the training data are used to predict the values of the response variable (Kuhn and Johnson, 2013). In the BT, multiple versions of the predictive model are generated (ensemble technique) by making bootstrap of the training set and each model is then used for predicting a new sample (Breiman, 1996). The multiple versions of the predictive model are then averaged to obtain an aggregated model prediction. Contrary to the BT, where all the original predictors are considered at every split of every tree, in the RF, the algorithm randomly selects predictors at each split, consequently reducing the trees correlation and potentially impacting on models performance (Breiman, 2001; Kuhn and Johnson, 2013).

In the Bayesian statistical inference models, a process of assigning and refining probability statements about unknown quantities is applied, which incorporates and updates prior knowledge and accounts for all sources of uncertainty (Link and Barker, 2010). The GPR algorithm provides a probabilistic approach for learning generic regression problems using flexible kernels functions (Verrelst et al., 2012a). An assumption of homoscedasticity is considered in GPR algorithm, i.e. assumes a constant noise power (error) in the relationship between the predicting variables and the response variable; such assumption is often not verified in biophysical retrieval studies because the noise can affect differently the acquisition process depending on the range of the measured variable (Lázaro-Gredilla et al., 2014). The VH-GPR overcomes this issue by allowing the noise power to vary throughout the input space, i.e. allowing a nonstandard variational approximation, which is more adjusted to signal-dependent noise scenarios (Lázaro-Gredilla et al., 2014).

In the application of the various machine learning regression algorithms, the \( NRI_{554,561} \) the \( WI_{900,970} \) and the \( \psi_{pd,0} \) were used as predicting variables for modelling the response variable (\( \psi_{pd} \)). The selection of the predicting variables \( NRI_{554,561} \) and \( WI_{900,970} \) was knowledge-assisted, following previous results in the same vineyards (Pôças et al., 2017), and the \( \psi_{pd,0} \) was newly-added, allowing the model (s) to learn, in each moment, from prior grapevine water status data.

The software ARTMO (Automated Radiative Transfer Models Operator) (Verrelst et al., 2012b), through the machine learning regression algorithm toolbox (Rivera et al., 2014), was used in the implementation of the four algorithms.

2.5. Models performance assessment

In the machine learning approaches, the relationship between the predicting variables and the response variable is learnt by fitting a flexible model from the data (training data) and adjusted to minimize the prediction error of an independent data set (validation data). Thus, the training data set (corresponding to the data from 2014 and 2017) was randomly split into calibration (70%; 83 observations) and validation (30%; 35 observations). Additionally, a validation procedure through an external data set (71 observations), corresponding to the data set built with data collected in the year 2015, was considered.

Several goodness-of-fit indicators were used to evaluate the prediction error of the regression-mode models (RF, BT, GPR, and VH-GPR), including the root mean squared error (RMSE), the mean absolute error (MAE), and the normalized root mean square error (NRMSE) as suggested by Kuhn and Johnson (2013).

Additionally, two statistical indicators were used to assess the weight and contribution of each predicting variable to the best performing machine learning algorithms: (i) Sigma value, and (ii) normalized root mean square error (NRMSE). The sigma value is the length-scale per input feature (predicting variable) in the algorithm and represents the relevance of each predicting variable, as discussed by Verrelst et al. (2016); lower sigma values represent higher relevance of the predicting variable. The NRMSE, which is the ratio between the RMSE and the range of observed values (Richter et al., 2012), was analyzed by successively including the various predicting variables in the algorithms.

When applying the \( \psi_{pd} \) for supporting irrigation scheduling, often data are aggregated into classes of vines water deficit (Deloire et al., 2005; Ojeda et al., 2001), which allows an easier data use by stakeholders, while minimizing the inherent variability of \( \psi_{pd} \). Therefore, the predicted values of \( \psi_{pd} \) obtained by each machine learning algorithm were further aggregated into classes of water deficit and compared with the corresponding classes of \( \psi_{pd} \) observed values. The definition of \( \psi_{pd} \) classes was based on the analysis of the \( \psi_{pd} \) distribution (Fig. 2b), resulting three class labels: (i) Class low water deficit:
0 MPa > $\psi_{pd}$ > −0.25 MPa; (ii) class moderate water deficit: −0.25 MPa > $\psi_{pd}$ > −0.50 MPa; and (iii) class high water deficit: −0.50 MPa > $\psi_{pd}$. This class definition was further supported by the fact that, in Mediterranean regions, the irrigation (under deficit irrigation strategies) most often starts when plants are under $\psi_{pd}$ values below −0.5 MPa to stabilize the wine quality conditions (Lopes et al., 2011; Van Leeuwen et al., 2009).

The 71 observations of $\psi_{pd}$ of the validation dataset were distributed into the three classes of water deficit as: (i) 5 observations in class low, (ii) 24 observations in class moderate, (iii) and 42 observations in class high. For evaluating the results of the comparison between observed and predicted classes following the data aggregation, a confusion matrix was built and the percentages of the overall model accuracy as well as the positive prediction value by class were obtained. The overall model accuracy reflects the agreement between observed and predicted classes and corresponded to the ratio of the number of cases correctly predicted (represented in the diagonal position of the confusion matrix) and the total number of cases, expressed as a percentage (Kuhn and Johnson, 2013). The overall accuracy makes no distinction about the type of errors being made. The positive prediction value by class was also computed considering the ratio of the number of cases correctly assigned in a class (true positives) and the total cases of that class, expressed as a percentage, thus taking in consideration the prevalence of the event (Kuhn and Johnson, 2013).

The validation procedures were implemented in ARTMO (Verrelst et al., 2012b), through the machine learning regression algorithm toolbox (Rivera et al., 2014).

### 3. Results

#### 3.1. Comparison between study years

The average results of $\psi_{pd}$ for each year, test site, irrigation treatment, and cultivar show the large variability of climate, agronomic, and environmental conditions (Table 4). Nevertheless, no statistically significant differences between the years 2015 and 2017 were observed for the test site 1, while the opposite was observed for the test site 2. Regarding the irrigation conditions, there were statistically significant differences between years for the non-irrigated treatment but not for the irrigated treatment. For the cultivars, statistically significant differences between years were observed only for TN. Overall, the mean $\psi_{pd}$ of the three years (2014, 2015, and 2017) was statistically different between test sites and between irrigation treatments, but not between cultivars. Nevertheless is noteworthy that the number of observations for cultivars TF and TB is much lower when compared to TN. The mean values of the $\psi_{pd}$ are significantly different between 2014 and the years 2015 and 2017, and the absolute value increases from 2014 to 2017.

#### 3.2. Models performance

The predicting variables used in the application of the machine learning algorithms were NRI_{554,561}, WI_{900,970}, and $\psi_{pd}$. The performance of the machine learning regression algorithms based on these predictors was assessed through indicators of the predicting errors for the validation data sets (Table 5).

The results show RMSE, MAE and NRMSE lower or equal than 0.151 MPa, 0.120 MPa and 19.9%, respectively, for the application of all the algorithms in the external validation dataset (Table 5). These results are close to those obtained with the validation dataset (lower or equal to 0.130 MPa, 0.100 MPa and 13.3%, respectively; Table 5), indicating a good robustness of the models.

Fig. 3 shows the performance of the machine learning algorithms based on the comparison between observed and predicted values of $\psi_{pd}$ for the validation and external validation datasets.

Overall, the results of Table 5 and Fig. 3 indicate the GPR and VH-GPR as the best performing algorithms to predict the $\psi_{pd}$. The weight of each predicting variable in the models obtained through GPR and VH-GPR, expressed by the sigma values, is shown in Table 6. The $\psi_{pd,0}$ was the predicting variables with more relative importance in both models (lower sigma values), followed by NRI_{554,561} and WI_{900,970}, respectively. Nevertheless, the values of NRMSE improved when each predictive variable was successively added to the models as shown in Table 6, which highlights both the relative importance of each predicting variable and the interaction between these variables.

The predicted values of $\psi_{pd}$ were further grouped into three classes (low, moderate, and high water deficit) and the comparison with the corresponding classes of observed $\psi_{pd}$ is shown in a confusion matrix (Table 7). The overall accuracy (percentage), obtained by the ratio of the cases correctly assigned (in the matrix diagonal) and the total number of cases, was 82% for GPR and RF and 83% for VH-GPR and BT. For all the algorithms, the prediction accuracy in the class low water deficit was 100% and for classes moderate and high water deficit was always higher or equal than 81% and 79%, respectively.

Through this approach it is possible to predict 2-levels of information about $\psi_{pd}$ (values and classes of values) that can assist different types of stakeholders.

#### 4. Discussion

The predicting variables NRI_{554,561}, and WI_{900,970} were also used for assessing grapevine water status in Douro wine region in the years 2014 and 2015 (Pôças et al., 2017). However, in this previous study, the variability regarding water deficit conditions was smaller, as 2014 and 2015 presented a shorter range of water deficit, without $\psi_{pd}$ values exceeding −1 MPa, when compared to the year 2017, as shown in Fig. 2(a). The NRI_{554,561} integrates spectral information of the green domain (520–570 nm), similarly to the Photochemical Reflectance Index (PRI; Gamon et al., 1992) that combines wavelengths of 531 nm and 570 nm and is often used for assessing crop water status (Hernández-Clemente et al., 2011; Pôças et al., 2015; Zarco-Tejada et al., 2013). The green spectral domain has been used as a proxy of crop stress conditions, including water deficit conditions, due to its ability for detecting subtle changes in the xanthophyll cycle pigment activity resulting from stress conditions (Gamon et al., 1992; Middleton et al., 2012). The WI_{900,970} integrates wavelengths of the
near infrared domain, specifically including the wavelength 970 nm corresponding to a water absorption peak (Peñuelas et al., 1993, 1997; Roberto et al., 2012). A novelty was introduced in the model by integrating the $\psi_{pd,0}$ as a time-dynamic predictor, which accounts for the natural dynamics of $\psi_{pd}$ along the grapevine cycle and aims to strengthen the learning process by the machine learning algorithms, while minimizing the presumable common trend effects between the $\psi_{pd}$ and the predictor variables. The day of the year (DOY) was previously used as a time-varying predictor (Pôças et al., 2017), partially accounting for the intra- and inter-annual variability relative to vegetation dynamics. However, a more versatile time-predictor, as $\psi_{pd,0}$, is needed to account for the complex relationships associated to a large variability of climatic, environmental, and agronomic conditions.

The results of the goodness-of-fit indicators of the various algorithms for the validation dataset (Table 5) were similar or better than the results of previous studies. The RMSE results were similar to those obtained for assessing water status in grapevine (Rapaport et al., 2015) and significantly better than those obtained for olive orchard (Rallo et al., 2014) using a partial least squares regression algorithm. Nevertheless, in these studies the spectral domains included range from the visible up to the shortwave infrared (SWIR), thus larger than the spectral range considered in the present study (Jones and Vaughan, 2010). The RMSE results were also similar to those obtained by Pôças et al. (2017) when assessing grapevines water status in Douro region in 2014 and 2015 when conditions of lower variability of $\psi_{pd}$ were observed (Fig. 2(a)), which suggests a robust performance of the models in the present study. Additionally, it is important to highlight that the results of RMSE, MAE and NRMSE were obtained for an external data set with 71 observations, instead of considering a cross validation procedure as in most of the previous studies (e.g., Pôças et al., 2017; Rallo et al., 2014).

The good performance of the four algorithms is likely related with the validation and external validation datasets using the four machine learning algorithms.
5. Conclusions

Predictive models applied in the regression mode were used for retrieving crop water status in vineyards of Douro wine region. A large set of climatic, environmental, and agronomic conditions were sampled to test the models robustness. Two years of data, representing a large variability of ψpd, were used for training the models, and a third year was used for an external validation of the models generated. Overall, the models showed good performance in predicting ψpd as continuous response variable, particularly those obtained through the GPR and VH GPR algorithms. When the results were further aggregated according to water deficit classes, the classes of ψpd values were accurately predicted including for the class of high water deficit, which represents the class of ψpd considered for irrigation management decisions. These two types of information can be used to support management decisions related with irrigation scheduling, management of water allocation, adjustment of irrigation sectors, and thus serve different types of stakeholders, including wine producers, water managers, and researchers.

Additionally, the results obtained constitute the basis for future applications of improved irrigation decisions based on hyperspectral data derived from aerial or satellite platforms. The upcoming global mapping imaging spectroradiometric missions and the increasing availability of hyperspectral cameras onboard unmanned aerial vehicles set good perspectives for using the predictive models proposed in this study for mapping the ψpd over large vineyards. Such applications would have great impact in the domain of precision irrigation in viticulture.

Acknowledgments

Isabel Pôças acknowledges FCT by the postdoctoral grant SFRH/BD/79767/2011 and the Post-Doctoral grant of the project ENGAGE-SKA POCI-01-0145-FEDER-022217, co-funded by FEDER through COMPETE (POCI-01-0145-FEDER-022217). The authors also thank the wine company Real Companhia Velha (and its Coordinator for Viticulture Rui Soares) and the wine company Symington Family Estates (and its R&D Viticulture Manager Fernando Alves) for the facilities provided for fieldwork, as well as the Associação para o Desenvolvimento da Viticultura Duriense (ADVID) for funding part of the field-work missions. This study was implemented under a cooperation protocol integrating Faculty of Sciences, University of Porto, ADVID, Real Companhia Velha, and Symington Family Estates.

References
