USING THE FRACTION OF TRANSPIRABLE SOIL WATER TO ESTIMATE GRAPEVINE LEAF WATER POTENTIAL: COMPARING THE CLASSICAL STATISTICAL REGRESSION APPROACH TO MACHINE LEARNING ALGORITHMS

Authors: Ricardo EGIPTO[1*], J. Miguel COSTA[2], José SILVESTRE[1], Carlos M. LOPES[2]

[1]*INIAV IP – Polo de Inovação de Dois Portos, 2565-191 Dois Portos, Portugal*
[2]*LEAF* - Linking Landscape, Environment, Agriculture and Food*, Instituto Superior de Agronomia, Univ. Lisboa, Portugal*

*Corresponding author: *ricardo.egipto@iniav.pt*

Abstract:

Context and purpose of the study – Weather uncertainty is forcing Mediterranean winegrowers to adopt new irrigation strategies to cope with water scarcity while ensuring a sustainable yield and improved berry and wine quality standards. Therefore, more accurate and high-resolution monitoring of soil water content and vine water status is a major concern. Leaf water potential measured at pre-dawn (Ψ_{PD}) is considered to be in equilibrium with soil water potential and is highly correlated with soil water content at the soil depth where roots extract water.
The aim of this study is to evaluate a dataset of eco-physiological data collected in a 3-year vineyard irrigation trial to assess the explanatory power of the fraction of transpirable soil water (FTSW) to predict Ψ_{PD} by comparing the classical statistical regression approach with a machine learning algorithm (MLA).

Material and methods – Deficit irrigation trials were conducted from 2013 to 2015 in a commercial vineyard in the Alentejo (southern Portugal). Trial plot was planted with *Vitis vinifera* (L.) cv. Aragonez (ARA)(syn. Tempranillo), grafted onto 1103 Paulsen rootstock and spaced 1.5 m within and 3.0 m between N-S oriented rows. The experimental layout was a randomized complete block design with two treatments: sustained deficit irrigation (SDI – control; ~30% Etc) and regulated deficit irrigation (RDI; ~15% Etc) and 4 replicates per treatment. The Ψ_{PD} and soil water content were measured the day before and the day after each irrigation event by using a capacitance probe down to a soil depth of 1 m and a Scholander pressure chamber. Models predicting Ψ_{PD} from FTSW were trained on 600 data cases and validated on an independent dataset (10% of all available data) using MATLAB R2022b (Mathworks, USA) and STATISTICA 13 (Tibco, USA).

Results – Our results show that 87.6% of the observed Ψ_{PD} variability is explained by the FTSW using a linear regression model (LRM) with a linear-logarithmic transformation of the independent variables. The accuracy of the prediction model, as measured by root mean squared error (RMSE), in the independent validation dataset, was 0.08 MPa. These results were compared to the estimation accuracy of a set of MLAs. Two support vector machine (SVM) algorithms with a quadratic and a medium Gaussian kernel function, and three Gaussian process regression (GPR) algorithms with an exponential, a squared exponential and a rational quadratic kernel functions were tested. All trained MLAs showed an accuracy in explaining the variability of the Ψ_{PD} (86-87%) similar to the LRM. An increase in the model explained variability of the independent dataset from 89 to 91% was observed in all MLAs, with an accuracy of 0.087 to 0.096 MPa as measured by the RMSE.
Both statistical methods indicate that Ψ_{PD} can be estimated with good accuracy using FTSW as an explanatory variable. Regarding the comparative performance of the two types of statistical models no differences were found in the ability of the tested models to estimate Ψ_{PD}.

Keywords: Deficit irrigation, soil water content, machine learning algorithms.

1. Introduction

The optimization of irrigation management is a major issue to promote viticulture sustainability. New irrigation strategies that cope with warmer climates and water scarcity while ensuring a sustainable yield and adequate berry and wine quality standards are on demand (Costa et al., 2020). A precise and high-resolution monitoring of soil water content and vine water status is needed to efficiently manage such irrigation strategies. Leaf water potential measured at pre-dawn (Ψ_{PD}) is considered to be in equilibrium with soil water potential and highly correlated with soil water content at soil depth where roots extract water (Ameglio et al, 1999; Williams and Araujo, 2002; Lebon et al., 2003; Pellegrino, 2003). However, measuring Ψ_{PD}, is a destructive and time-consuming method, has low spatial and temporal resolution, and involves measurements during the night, at pre-dawn. In order to overcome such a series of disadvantages, faster, continuous and accurate methods are on demand. An alternative approach is to use data from soil moisture monitoring devices as a proxy for Ψ_{PD}. Modelling Ψ_{PD} from soil water content (SWC) is not new and has been proposed to assess plant water status and support irrigation management of vineyards (Williams and Araujo, 2002; Pellegrino, 2003; Pellegrino et al., 2004). These studies used classic regression analysis to determine the relationships between SWC and Ψ_{PD}. However, few studies used a machine learning approach for this task. The present study aims to evaluate a dataset of eco-physiological data collected in a 3-year vineyard irrigation trial to assess the explanatory power of the fraction of transpirable soil water (FTSW) to predict Ψ_{PD} by comparing the classical statistical regression approach with a machine learning algorithm (MLA).

2. Material and methods

Plant material and growing conditions

Plant material – the experiments was conducted from 2013 to 2015 in a commercial vineyard (Herdade do Esporão), located at Reguengos de Monsaraz, Alentejo winegrowing region (Southern Portugal, lat. 38◦ 23′ 55.00′′ N; long.7◦ 32′ 46.00′′ W). The plot was planted with *Vitis vinifera* (L.) cv. Aragonez (ARA; syn. Tempranillo), grafted onto 1103 Paulsen rootstock and spaced 1.5 m within and 3.0 m between N-S oriented rows. The experimental layout was a randomized complete block design with two irrigation treatments: sustained deficit irrigation (SDI - control; ~30% ETc) and regulated deficit irrigation (RDI; ~15% ETc) and 4 replicates per treatment. The elemental plot comprised three adjacent rows (two buffer rows and a central one for data collection). Measurements were done in two blocks per irrigation treatment.

Plant and soil water content measurements - Predawn leaf water potential (Ψ_{PD}) and soil water content were measured the day before and the day after each irrigation event by using a Scholander pressure chamber and a portable capacitance probe down to a soil depth of 1 m, respectively. The fraction of transpirable soil water (FTSW) was calculated for each smart point according to Pellegrino et al. (2003, 2004). The Ψ_{PD} of two exposed and fully unfolded healthy leaves, collected from the vines near the access tube of the capacitance probe, were measured before dawn. A total of 8 leaves spread over 4 sampling points were measured on each sampling date.

Statistical analysis – The original data was split into two datasets, ensuring that both datasets have a similar data distribution. A dataset with 600 data cases, was used for training and internal validation of the models, and another dataset with 10 % of all original data, was used as an independent validation dataset. A linear regression model with a linear-logarithmic transformation of the independent variable (FTSW) was fitted from conventional least squares method. The same dataset was used to train and validate a set of supervised MLAs. Two support vector machine (SVM) algorithms with a quadratic and a medium Gaussian kernel function, and three Gaussian process regression (GPR) algorithms with an exponential, a squared exponential and a rational quadratic kernel function, were fitted to predict Ψ_{PD} from FTSW with MATLAB R2022b (Mathworks, USA). The supervised MLAs were fitted during the training phase using a five-fold cross-validation process. All models have been subjected to a validation process using an external validation dataset. The measures of the differences between the model predicted and observed values were analysed by the root mean square error (RMSE) and the mean absolute error (MAE) and used as measures of model accuracy.

3. Results and discussion

3.1. Linear regression model

The trained LRM with a linear-logarithmic transformation of the independent variable (FTSW) explained 87.6% of the observed Ψ_{PD} variability (Figure 1). The model showed a tendency to overestimate Ψ_{PD} values lower than -0.4 MPa (Figure 2A). Nevertheless, the model presented a good accuracy, as measured in the independent validation dataset by RMSE values (0.082 MPa), and the MAE (0.067 MPa) (Table 1).

3.2. Supervised machine learning algorithm

The trained MLAs were able to explain 86-87% of the Ψ_{PD} variability in the training process (data not shown). In the validation process, the MLAs showed good accuracy with the RMSE and MAE of the Ψ_{PD}, predicted with the independent validation dataset, in the range of 0.087 - 0.096 MPa and 0.066 - 0.069 MPa, respectively (Table 1). As observed in the LRM, the MLAs showed a tendency for higher estimation errors when the Ψ_{PD} was below -0.4 MPa (Figures 2B-F). Between the MLAs, the Gaussian process regression (GPR) algorithms performed slightly better than the support vector machines (SVM) (Table 1).

3.3. Comparing the accuracy of both methods for estimating Ψ_{PD}

To compare the accuracy of LRM and the MLAs to estimate Ψ_{PD} from FTSW, a validation using an independent dataset was performed. The scatter plots of observed *vs* estimated Ψ_{PD}, using the independent validation dataset show an overall good agreement between observed and model predictions (Figure 2). In turn, the correlation between observed and estimated values showed a very high and significant agreement (see Table 1). indicating that the model is suitable for accurate Ψ_{PD} prediction.

4. Conclusions

Aiming to find alternative methods to pre-dawn leaf water potential measurements to assess soil and plant water status in vineyards, we tested the explanatory power of FTSW for Ψ_{PD} estimation. In this paper, we compared a classical linear model approach with non-parametric MLAs approaches, using different non-linear covariance functions (kernels). Results showed that all models estimated the Ψ_{PD} from FTSW with high prediction accuracy, with a maximum prediction error of around 0.06 MPa in the validation process. The high goodness of fit of all models indicate almost no differences between the classic linear regression model with logarithmic transformation of the independent variable (FTSW) and the machine learning algorithms. Although the MLA models show only slightly higher accuracy than LRM, the advantage of MLA to automatically learn data patterns from new data inputs and to optimize the Ψ_{PD} prediction models should not be disregarded. Because of its simplicity and computational power, MLA using Ψ_{PD} estimated from FTSW can be an important tool to support deficit irrigation management strategies in modern viticulture. Since root distribution and extraction patterns may not always be homogeneous across the soil profile and in different soil textures, also due to the large influence of atmospheric demand on the thresholds of FTSW, it should be considered that the model needs to be validated before use in other climates, soils as well as grapevine variety/rootstock combinations.

5. Acknowledgements

This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement FP7-311775, Project INNOVINE.

6. Literature cited

Ameglio, T., Archer, P., Cohen, M., Valancogne, C., Daudet, F.-A., Dayau, S., Cruiziat, P., 1999. Significance and limits in the use of predawn leaf water potential for tree irrigation. Plant and Soil 207, 155–167.

Costa, J.M., Vaz, M., Escalona, J.M., Egipto, R., Lopes, C.M., Medrano, H., Chaves, M.M., 2020. Water as a critical issue for viticulture in southern Europe: sustainability vs competitiveness. IVES, International Viticulture and Enology Society, Technical Reviews Vine & Wine, pp. 1-2

Lebon E., Dumas V., Pieri P., R. Shlutz H., 2003. Modelling the seasonal dynamics of the soil water balance of vineyards. Functional Plant Biology 30, 699-710.

Pellegrino A., 2003. Élaboration d'un outil de diagnostic du stress hydrique utilisable sur la vigne en parcelle agricole par couplage d'un modèle de bilan hydrique et d'indicateurs de fonctionnement de la plante. PhD de l'Ecole Nationale Supérieure Agronomique de Montpellier, France.

Pellegrino A., Lebon E., Voltz M., Wery J., 2004. Relationships between plant and soil water status in vine (*Vitis vinifera L*.). Plant and Soil Sciences 266, 129–142.

Williams, L.E., Araujo, F.J., 2002. Correlations among predawn leaf, midday leaf, and midday stem water potential and their correlations with other measures of soil and plant water status in *Vitis vinifera*. J. Am. Soc. Hortic. Sci. 127, 448–454.

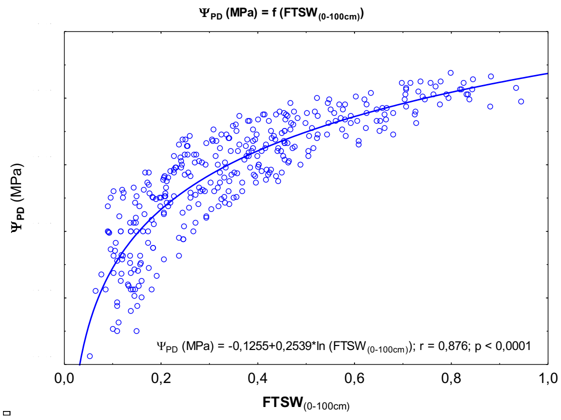

Figure 1. Linear regression model (LRM) with a linear-logarithmic transformation of the independent variable (FTSW) versus the Ψ_{PD}, fitted for Aragonez. The training dataset (n=600) was used for modelling purposes.

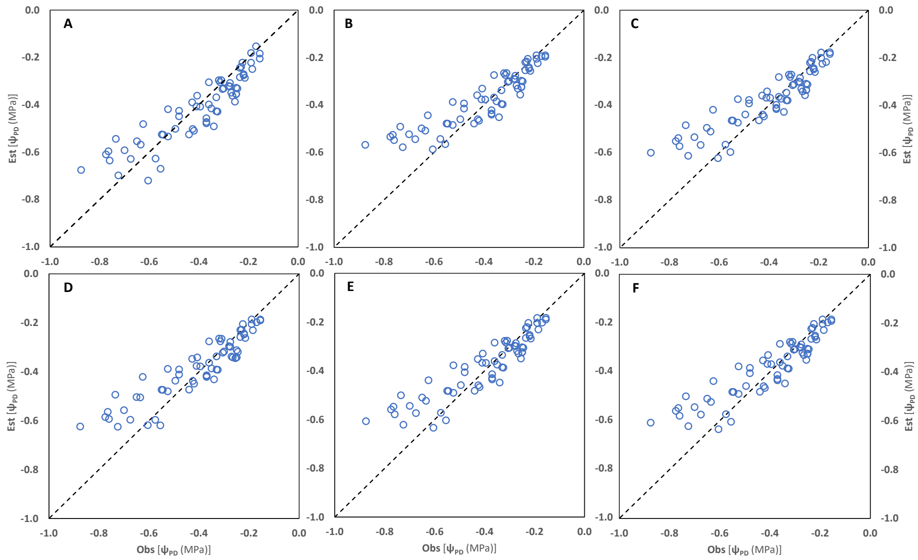

Figure 2: Observed versus predicted Ψ_{PD} using validation dataset. **(A)** Linear Regression Model (LRM) with linear-logarithm transformation of FTSW, **(B)** Support Vector Machine (SVM) algorithm with a quadratic kernel function, **(C)** SVM algorithm with a medium Gaussian kernel, **(D)** Gaussian Process Regression (GPR) with an exponential kernel, **(E)** GPR with a squared exponential kernel and **(F)** GPR with a rational quadratic kernel. The black dashed line represents the 1:1 line.

Table 1. Statistical indicators of goodness of fit of the predicted models for Ψ_{PD} estimation. Linear regression model (LRM) with a linear-logarithmic transformation of the independent variable (FTSW) and machine learning algorithms (MLA) accuracy to predict Ψ_{PD} from an independent validation dataset. The Pearson correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE) are presented as accuracy measures. All the correlation coefficients presented were statistically highly significant, p < 0.001.

Model	Transformation	r	RMSE (MPa)	MAE (MPa)
Linear Regression	Logarithmic	0.902	0.082	0.067
Support Vector Machine	Quadratic	0.899	0.096	0.069
	Medium Gaussian	0.906	0.093	0.068
Gaussian Process Regression	Exponential	0.911	0.087	0.066
	Squared Exponential	0.909	0.089	0.066
	Rational Quadratic	0.909	0.089	0.066