

Identifying the most relevant issues to minimize an automated irrigation water supply in vineyards using Data Mining

A. Rabasa Dolado¹, J. M. Molina Martínez², H. Puerto Molina³,
J. J. Rodríguez-Sala¹, L. Santamaría Arana¹ & A. Ruiz Canales³

¹*Centro de Investigación Operativa,*

Universidad Miguel Hernández de Elche, Spain

²*Dpto. Ingeniería de Alimentos y del Equipamiento Agrícola,*

Universidad Politécnica de Cartagena, Spain

³*Dpto. Ingeniería, Universidad Miguel Hernández de Elche, Spain*

Abstract

In order to carry out optimal daily irrigation programming for vineyard production, different variables have to be considered. Some of them are related to soil (soil moisture variation), other variables are related to crop (variety, dendrometry, leaf temperature) and finally some climatic variables are considered (reference evapotranspiration ETo and rainfall). This daily irrigation programming is traditionally based on physical–mathematical models (soil water balance, radiation, drainage among others). This is a model-driven methodology. However, in this paper a data-driven methodology is proposed. With this methodology, the daily water consumption of a crop is estimated from daily values acquired through the variables described above. This approach predicts minimal water consumption. In addition, the significant variables in daily water consumption according to crop and phenological phases are determined. The experiment was developed in a vineyard farm located in Villena, Spain. Representative daily data of five vineyard varieties (Cabernet Blanc, Merlot, Petit Verdot, Sauvignon Blanc and Syrah) were collected from 2005 to 2009. These data were subsequently analysed by Data Mining techniques that are oriented to extraction and rule systems classification. This methodology will be used in the development of applied algorithms in software for decision tools in irrigation programming for vineyard production.

Keywords: Data Mining, irrigation, vineyard.



1 Introduction

Wine production is one of the main economic activities in Villena (Alicante), in the South of the Valencian Community (Spain). Nowadays, several approved wineries in this area are experimenting with foreign varieties of vines such as Cabernet Sauvignon or Merlot, both French varieties. These varieties have been adapted to the climatology, altitude and soil conditions of this area.

In order to adapt these new varieties to these specific conditions, several Agronomical practices have to be implemented. One aspect of these practices is irrigation scheduling. Due to the specific climatology (annual rainfall, average temperature and evapotranspiration of 420 mm, 15°C and 1,248 mm respectively), irrigation supply has to be applied in order to obtain an adequate yield.

Moreover, irrigation enables increased yields [1], though a moderate water deficit is often desirable to improve wine composition [2]. Deficit irrigation has been suggested as a strategy to improve fruit composition for premium quality wines reducing canopy vigour, increasing fruit exposure to light and reducing berry growth to avoid dilution effects [3].

Deficit irrigation consists in applying water rates to replace only part of the vine's potential evapotranspiration either during the whole season or only during some previously established phenological periods. Regulated deficit irrigation can be applied as a strategy to reduce the possible negative impact of irrigation on wine quality. In the past, Sal6n et al. [4] studied the response to irrigation of cv. Bobal. However, in the previously cited varieties (Cabernet Sauvignon or Merlot) the effects of different irrigation regimes on vine performance and wine quality have not been reported yet for this area.

Soil water availability is a critical factor for vine performance and wine composition. This availability is monitored by means of soil moisture gauges. The use of soil-based water status indicators have been adopted as an adequate strategy for water balance estimation.

The use of plant-based water status indicators has become very popular in recent years for studying plant-water relations and for planning more precise irrigation programmes, because it is recognized that the tree itself is the best indicator of its water status. One of these methodologies, monitoring trunk diameter variations (TDV), is widely used to detect plant water status [5, 6]. This methodology is commonly named as dendrometry.

The ultimate goal of these techniques is to provide vine growers with information about the more appropriate volumes of irrigation to apply in each phenological period, depending on the desired yield levels and wine styles.

An added structural problem of this area is the aridity of the climate and the scarcity of water resources, which is a historical and recurring characteristic in southern Europe, where during several periods in the year, water supply is not guaranteed. In order to optimize water supply and to rationalize the management of demand per vineyard an adequate irrigation scheduling could be established. This scheduling is based mainly on soil moisture gauge measurements,



dendrometry and climatology. Several other variables are also related to irrigation scheduling of vineyards.

In this paper, a data-driven methodology is proposed, where the daily water consumption of a crop is estimated from the daily values acquired for the variables described above. This approach predicts minimal water consumption and determines the significant variables of daily water consumption according to crop and phenological phases. This methodology is based on experimental data from a commercial vineyard.

The experiment was developed at a vineyard situated in Villena, Spain. Representative daily data are were collected from five vineyard varieties (Cabernet Blanc, Merlot, Petit Verdot, Sauvignon Blanc and Syrah) from 2005 to 2009. These data were subsequently analysed by Data Mining techniques, which are oriented to extraction and rule systems classification.

It is possible to apply this methodology to the development of applied algorithms in software for decision tools in irrigation scheduling for vineyard production.

Daily irrigation scheduling is traditionally based on physical-mathematical models (soil water balance, radiation, drainage among others) as in model-driven methodology is proposed.

2 Problem definition and main objective

Irrigation scheduling is basically a method for calculating how much and when irrigation water has to be applied. In order to answer these questions, some data have to be collected, which are mainly soil-based, plant-based and meteorological-based. These data are processed in the water balance. Water balance is obtained by applying a mass conservation equation [7–10]:

$$ET = P + I - \Delta S - D - R \quad (1)$$

where, ET is evapotranspiration (1), P is rainfall, I is irrigation, ΔS is the variation of water stock between two dates, D is the drainage under the profile and R is the runoff [11, 12]. ET is obtained by multiplying reference evapotranspiration (E_{To}) by the crop coefficient (K_c). The first factor is related to meteorological data and the second depends on the crop. In the development of the experiment, daily data were obtained (E_{To} , P, I, ΔS). Drainage and runoff were not considered in this experiment. In addition, some data to detect plant water status were collected, as well as monitoring trunk diameter variations (TDV) or dendrometry.

In order to carry out the experiment on optimal daily irrigation scheduling for vineyard production, different variables have to be considered. Some of them are related to soil (soil moisture variation), other variables are related to crop (variety, dendrometry) and finally some climatic and agricultural variables are considered (reference evapotranspiration E_{To} , rainfall and irrigation dose). Finally a leaf wetness sensor is included. This is a meteorological parameter that describes the amount of dew and precipitation left on surfaces. It is only used for



monitoring leaf moisture for agricultural purposes, such as fungus and disease control, for control of irrigation systems, and for detection of fog and dew conditions, and early detection of rainfall. The latter is not included for irrigation scheduling but its measurement data have been tested in order to contrast them with meteorological data.

Site description and experimental design: The experiment was carried out during the seasons from 2005 to 2009 at several plots with different varieties: Cabernet Blanc, Merlot, Petit Verdot, Sauvignon Blanc and Syrah vines (*Vitis vinifera* L.). These plots are situated in a valley between two hills which has sandy franc soil. These vineyards were planted in 2001 with a spacing of $2.80 \times 0.90 \text{ m}^2$ (density is around $4,000 \text{ plant} \cdot \text{ha}^{-1}$).

The irrigation system is trickle irrigation with integrated emitters spaced 0.80 cm. The average flow of the emitter is $3.85 \text{ l} \cdot \text{hour}^{-1}$ (1.125 emitters/plant). The vineyard is near Villena ($38^\circ 34' 31.4'' \text{N}$, $0^\circ 56' 35.58'' \text{W}$, elevation 540 m), Alicante, Spain (Figure 1).



Figure 1: Aerial view of the experimental plots (Googlemaps).

Weather conditions during the experiment were measured with an automated meteorological station at the plot and reference evapotranspiration (ET₀) was calculated as in [13].

Irrigation frequency varied from 3 to 5 days per week. The previously mentioned water balance parameters were measured daily throughout the experimental period. ET₀ and cumulated daily rainfall (P) were measured by the automated meteorological station. Daily irrigation was measured with a pluviometer. Several FDR soil moisture gauges measured the amount applied to each monitored plant (one per variety) and at several depths (10, 30 and 60 cm deep). Using these data, the variation of water stock between two dates (ΔS) was obtained. The daily trunk diameter was measured in three plants per variety. The

daily measurement was obtained from the average data in order to detect water stress in the plant. Finally, the leaf wetness sensor measured the leaf moisture every day. These last two data are not related to irrigation scheduling but are associated to plant and meteorological status.

In this case study, the attributes to be considered are the following:

- DATE: Is the date of data collection throughout the experimental period.
- SOIL_MOIST_10: Is the soil moisture measurement obtained with a FDR soil moisture probe 10 cm deep in percentage % (0-100).
- SOIL_MOIST_30: Is the soil moisture measurement obtained with a FDR soil moisture probe 30 cm deep in percentage % (0-100).
- SOIL_MOIST_60: Is the soil moisture measurement obtained with a FDR soil moisture probe 30 cm deep in percentage % (0-100).
- CF_AVG_INCR: Is the daily average trunk diameter (mm).
- LEAF_HUMIDITY: Is the leaf moisture in percentage % (0-100).
- ETo: Is the reference evapotranspiration obtained by Hargreaves method (mm).
- RAINFALL: Is the daily cumulated rainfall (mm).
- IRRIGATION: Is the daily cumulated irrigation (mm).

After the data collection, a specific data mining methodology was applied in order to detect possible wrong data and determine the weight of all data in the water balance. This methodology is used in several applications and, in this case, it was used to obtain algorithms with a low data quantity in order to apply them to a hardware design for automatic irrigation.

3 Methodology

In contrast to the classical model driven approach, where mathematical or physical models are used to describe or forecast problems, in this paper the data driven perspective defended by Solomatine [14] is suggested as the most accurate and efficient approach for classification tasks from historical data, without consideration of previous context knowledge.

Decision Trees are widely applied in this framework, where a target attribute must be classified. There are different tree implementations depending on the nature of the variables to be considered, the branch cardinality and some other constraints. In this paper the authors chose the ID3 algorithm to extract classification rules responding the structure: $A_1, A_2, \dots, A_n \rightarrow C$.

A_i being the categorical antecedent variables and C the categorical consequent or target variable, as shown in Table 1, in section 2.

After applying ID3 algorithm [15] with the original data set D , a rule set R is obtained. The reduction algorithm RBS [16] reduces R to R' that contains only significant rules to predict the consequent as shown in Fig.2.



Table 1: Attribute description.

Attribute	Original values	Pre-processed values
DATE	Dates typed as dd/mm/yyyy	Jan-Apr: Winter May-Jun: Spring Jul-Aug: Summer Sep-Oct: Autumn Nov-Dec: Winter
SOIL_MOIST_10 SOIL_MOIST_30 SOIL_MOIST_60	Positive numerical values	< 8.5: Default [8.5..16.4]: OK > 16.4: Exceed
CF_AVG_INCR	Positive numerical values	< 300: Decrease (*) [-300..300]: Constant >300: Increase
LEAF_HUMIDITY	Positive numerical values in range 0..10	0..3: Low 4..6: Medium 7..10: High
ETo	Positive numerical values	[0..3]: Low (3..6]: Medium (6..10]: High >10: Very High
RAINFALL	Positive numerical values normalized	0: Null (0..5): Low [5..15]: Medium [15..25]: High >25: Very High
IRRIGATION	Positive numerical values	0: Null (0..3): Low [3..10): Medium >=10: High

(*)All attributes were discretized exclusively attending expert criteria except CF_AVG_INCR obtained from the density function around Average $\pm 2 \cdot \text{StdDeviation}$.

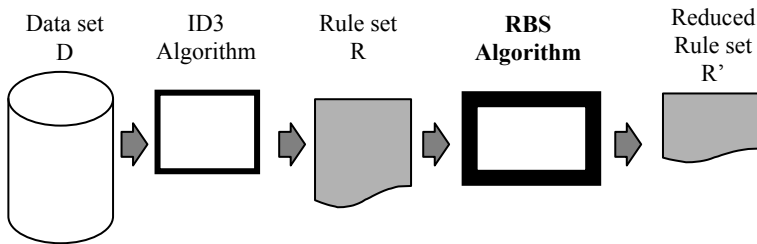


Figure 2: From Data set to Reduced Classification Rule set.

4 Computational experiment

In this section the results of the computational experiment are shown exclusively under the data analysis approach, without expert interpretation.

After the RBS algorithm had been applied to the 5 scenarios, the 5 vineyard varieties have very similar correlation between the studied antecedents and the irrigation variable, with minor differences.

Trying to minimize the water consumption, the irrigation target variable was instanced to “Null” and the most important patterns leading to no-irrigation instance are the following:

- Attributes SOIL_MOIST_10, SOIL_MOIST_30 and SOIL_MOIST_60 have a very low correlation with the target variable.
- Climatic data correlate more with Irrigation if they are considered on the previous day, instead of the current day.
- Rules in Spring and Autumn have a uniform distribution, so there are no supportable patterns to describe the vineyards behaviour in those seasons.
- The most repeated pattern in Winter is CF_AVG_INCR=Null, ETo=Low, RAINFALL=Null. With this antecedent, the IRRIGATION is NULL with confidence close to 98%.
- In Summer (with dry conditions ETo=Low, RAINFALL=Null/Low), while CF_AVG_INCR do not decrease, the Irrigation is NULL with confidence close to 92%, but when CF_AVG_INCR decreases, IRRIGATION variable turns to MEDIUM.

5 Conclusions

The data obtained for Cabernet Blanc and Merlot could be extrapolated to Petit Verdot, Sauvignon Blanc and Syrah vines. The main variation is the irrigation dose and timing.

Because climatic data correlate more with irrigation if they are considered on the previous day, instead of the current day, it is necessary to consider the automation of irrigation scheduling in order to apply the specific irrigation dose.

This methodology could be used to obtain algorithms with a low quantity of data in order to obtain cheaper irrigation hardware devices. With low data, the designed hardware spends less memory and space. It could be reached through Feature Selection techniques.

The methodology developed could be applied to improve irrigation scheduling in vineyards. From a similar field of data treatment, it is possible to adapt some irrigation practices (deficit irrigation, partial root drying, etc.) to the specific local conditions of the vineyard and use adequate water quantity and no more than this. This is one of the main objectives for water resource management in an arid zone with water scarcity.



References

- [1] Williams, L.E. & Matthews, M.A., Grapevine (Chapter 34). *Irrigation of agricultural crops agronomy monograph no. 30*. eds. B.A. Stewart & D.R. Nielsen, ASA-CSSA-SSSA Madison, WI, USA pp 1019–1055, 1990.
- [2] Jackson, D.I. & Lombard, P.B., *Environmental and management practices affecting grape composition and wine quality-A review*. American Journal of Enology and Viticulture 44:409–430, 1993.
- [3] McCarthy, M.G., Loveys, B.R., Dry, P.R. & Stoll, M., *Regulated deficit irrigation and partial rootzone drying as irrigation management techniques for grapevines*. Deficit irrigation practices. FAO Water Reports No 22. Rome, Italy, pp 79–87, 2000.
- [4] Salón, J.L., Chirivella, C. & Castel, J.R., Response of Vitis Vinifera cv. Bobal to deWcit irrigation in Requena, Spain. Water relations, yield and wine quality. American Journal of Enology and Viticulture 56:1–18, 2005.
- [5] Goldhamer, D.A. & Fereres, E., Irrigation scheduling protocols using continuously recorded trunk diameter measurements. Irrigation Science 20:115–125, 2001.
- [6] Naor, A., Irrigation scheduling and evaluation of tree water status in deciduous orchards. Horticulture 32:111–166, 2006.
- [7] Faures, J.M., Goodrich, D.C., Woolhiser, D.A. & Sorooshian, S., *Impact of small-scale rainfall variability on runoff model*. Journal of hydrology. Amsterdam: Elsevier Science. B.V, 1995.
- [8] Kerkides, P., Michalopoulou, H., Papaioannou, G. & Pollatou, R., *Water balance estimates over Greece*. Agricultural Water Management, 32, 85–104, 1996.
- [9] Moreno, F., Cayuelas, J.A., Fernández, J.E., Boy, E., Murillo, J.M. & Cabrera, F., *Water balance and nitrate leaching in an irrigated maize crop in SW Spain*. Agricultural Water Management, 32, 71–83, 1996.
- [10] Sánchez-Cohen, I., Lopes, V.L., Slack, D.C. & Fogel, M.M., *Water balance model for small scale water harvesting systems*. Journal of Irrigation and Drainage Engineering, 132 (2), 123–128, 1997.
- [11] Maraux, F. & Lafolie, F., *Modeling soil water balance of a maize-sorghum sequence*. Soil Science Society of America Journal, 62 (1), 75–82, 1998.
- [12] Schiller, G. & Cohen, Y., *Water balance of Pinus halepensis Mill. afforestation in an arid region*. Forest Ecology and Management, 105 (1-3), 121–128, 1998.
- [13] Hargreaves, G.H. & Samani, Z.A., *Estimation of standard deviation of potential evapotranspiration*. Proceedings ASCE, Journal of Irrigation and Drainage Engineering, 114(1), 1988.
- [14] Solomatine, D.P., *Applications of data-driven modeling and machine learning in control of water resources*. Idea Group Publishing. Computational Intelligence in Control, 197–217, 2002.
- [15] Quinlann, J.R., *Induction of decision trees*. Machine Learning, vol. 1, 81–106, 1986.



- [16] Abadía, R., Almiñana, M., Escudero, L.F., Pérez, A., Rabasa, A. & Santamaría, L., *A Rules Reduction Algorithm Based On Significance Measure*. Ed. Zanasi, Alzamora, Ebecken, Brebbia. Cádiz. WIT Press. Proc. IX Data Mining and Information Engineering. 63-72, 2008.

