Combining a sensor software with statistical analysis for modeling vine water deficit impact on grape quality

Aurélie Thébault, Thibault Scholash, Brigitte Charnomordic, Nadine Hilgert

Combining a sensor software with statistical analysis for modeling vine water deficit impact on grape quality

Aurélie Thébault[a], Thibaut Scholasch[b], Brigitte Charnomordic[a], Nadine Hilgert[a]

[a]*INRA-SupAgro, UMR 729 MISTEA, F-34060 Montpellier, France*
[b]*Fruition Sciences, 34 Montpellier, France*

Abstract

This work proposes a methodology using temporal data and domain knowledge in order to analyze a complex agronomical feature, namely the influence of vine water deficit on grape quality. Raw temporal data are available but they are not directly usable to estimate vine water deficit. The methodology associates advanced techniques in computer science and statistics. A preliminary step is required to determine if the amount of water effectively used by the vine is sufficient or not. This step necessitates an ecophysiological model, based on expertise. The expertise is first formalized in an ontology, under the form of concepts and relationships between them, and then used in conjunction with raw data and mathematical models to design a software sensor. Next the software sensor outputs are put in relation to product quality, assessed by quantitative measurements. This relation is analyzed by regression trees and advanced data analysis methods, such as functional data regression. The methodology is applied to a case study involving an experimental design in French vineyards. The temporal data consist of sap

Email address: `bch@supagro.inra.fr` (Brigitte Charnomordic)

Preprint submitted to Computer and Electronics in Agriculture Tuesday 10th December, 2013

flow measurements, and the goal is to explain fruit quality parameters (sugar concentration and weight), using vine's water variations at key stages of vine phenological development. The results are discussed, as well as the method genericity and robustness.

Keywords: vine water stress, functional data analysis, ontology, expert knowledge, grape quality, regression tree, temporal data

1. Introduction

In modern Agronomy, the recent progress of sensors provides a lot of data, among them many temporal data. This opens new challenges, such as the proper calibration of these sensors, and the use of temporal data to establish relationships with product characteristics and quality. These relationships are not easy to determine because of the high variability of biological material. This can be compensated by the integration of expertise, as Agronomy is a domain that has always relied as much on experience than on science. Nevertheless, for domain knowledge to be effectively used in collaboration with mathematical models and data, an expertise formalization step is required.

Our objective in this paper is to show the interest of a formalized data and knowledge-based approach to study a complex agronomical phenomenon, namely the influence of vine water deficit on grape quality. Grape quality analytical measurements are available and well established. In contrast, vine water deficit cannot be directly measured and requires a preliminary step to relate the amount of water effectively used by the vine, in order to determine if if this amount falls short of some reference amount. Typically, the *reference* amount is the maximal amount of water a vine can use.

Various methods exist to characterize the level of water deficit experienced by the plant as reviewed by Jones (2004). Tissue water status can be assessed visually or by measurements of vine water potential. However, both methods have serious drawbacks. The lack of precision of visual observations often leads to yield reduction before visible symptoms occurs. The pressure chamber method used to measure water potential is slow and labour intensive, especially for predawn measurement, and is unsuitable for automation. In a production context, collecting predawn leaf water potential is not a practical solution. It is a destructive method, which must be performed before sunrise and is sensitive to vapor pressure deficit, making interpretation difficult, see Rodrigues et al. (2012). In addition, measurements done with pressure chambers are very dependent on atmospheric conditions and vine phenological stage, see Olivo et al. (2009); Williams and Baeza (2007); Rodrigues et al. (2012); Santesteban et al. (2011). Other plant sensor-based monitoring approaches for estimating water deficit, like trunk diameter fluctuations, have been reported as unsuccessful for irrigation scheduling, see Montoro et al. (2011).

Thus, as of today, sap flow sensors are the only commercially available method to measure automatically and continuously systemic plant water use, see Ferreira et al. (2012). Sap flow sensors indirectly measure changes in stomata conductance and have recently become available. The main advantage of sap flow measurements is to allow automatic and continuous measurement of water flowing through the plant, which is directly related to transpiration, see Escalona et al. (2002); Jones (2004); Cifre et al. (2005);

Zhang et al. (2011). However, sap flow is a complex phenomenon. The sensitive measurement technique requires a complex instrumentation and technical expertise for the definition of irrigation control thresholds, see Ginestar et al. (1998). Expert knowledge is necessary to convert raw data into useful transformed data, i.e. water courses, by designing a software sensor. To the best of our knowledge, no such attempt to design a sap flow software sensor has been done yet.

Once these data transformations are validated, it is possible to study the influence of vine water deficit on grape characteristics. The existence of relationships between vine water deficit and fruit composition has already been reported in the literature, see des Gachons et al. (2005); Koundouras et al. (2006); Van Leeuwen et al. (2009). These studies are limited to the study of vine water status scalar measurements. In the present paper, a proposal is made to use water courses, that opens the way to a range of new studies.

We will first show how a formalized data and knowledge-based approach can be useful to design a software sensor. Knowledge formalization will be done by using ontologies, which take increasing importance in the field of Life Sciences, see Villanueva-Rosales and Dumontier (2008); Thomopoulos et al. (2013), for their ability to model and structure qualitative domain knowledge.

In a second step, water use trajectories will be put in relation to grape quality indicators such as Berry Weight or Sugar Concentration, using recent data analysis tools and formalized knowledge. Innovative data analysis tools include functional data analysis that offers the possibility to use curve (functional) data instead of scalar data. Functional data analysis has not been much used in life sciences yet, see Ullah and Finch (2013), though it could be of particular interest in the Vine and Wine Industry, and more generally for modern Agronomy.

The modeling task is divided into two independent parts: software sensor design and temporal data analysis. If the sensor design procedure were different, this would not affect the validity of the data analysis methodology.

The methodological work is illustrated by a case study, involving an experimental design on several vineyards in the Languedoc region (France).

The paper is organized as follows: Section 2 presents the material and methods. It is divided into four parts. The first part gives some elements about data and the second one presents ontology-based formalization. The software sensor design, that relies on the use of mathematical models, data and formalized knowledge, is described in the third part. The illustrative example shows how it is possible to transform raw sap flow data into vine water deficit courses. The fourth part describes the data analysis methods used for analyzing the software sensor output in relation to product quality. Section 3 presents and discusses the results for vine water deficit estimation and its relationship with grape composition (Sugar Concentration, Berry Weight). Some concluding remarks and perspectives are given in Section 4.

2. Material and methods

In this section, we propose to follow four steps:

- to describe the experimental design with its input and output variables;
- to formalize ecophysiological knowledge using an ontology;

- to design a software sensor using formalized knowledge, a mathematical model, and data;

- to relate software sensor output to product quality using decision trees and functional data analysis.

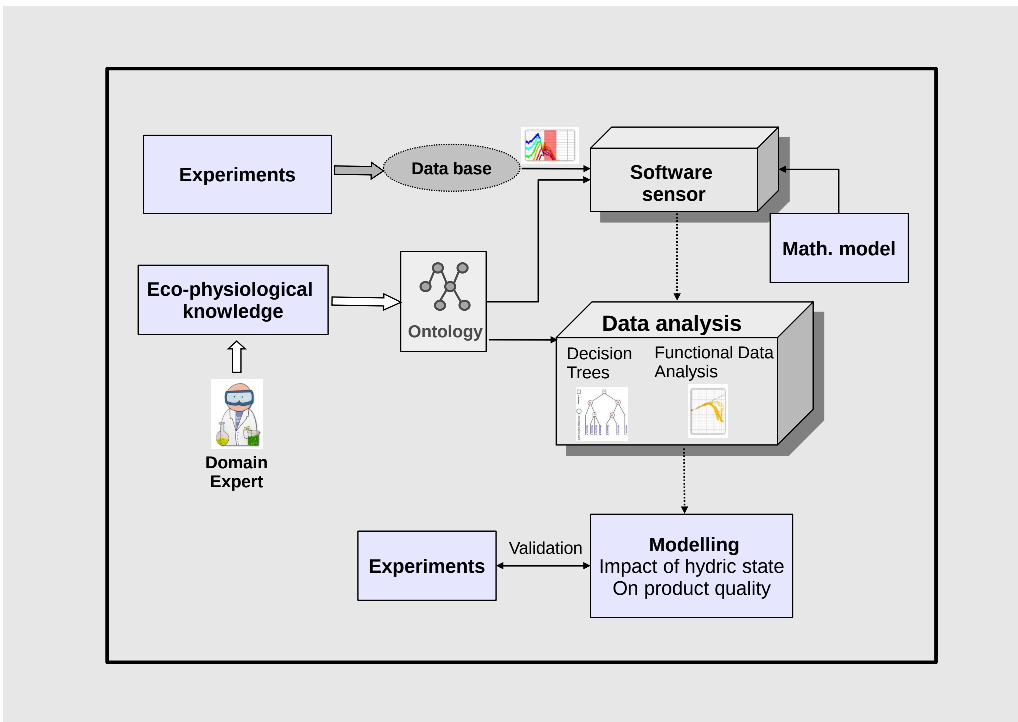


Figure 1: Outline of the proposed modeling approach.

Expert knowledge plays an essential part in the modeling process, and we focus on providing an efficient way to separate the data-based statistical procedures from the qualitative knowledge-based assumptions.

The outline of the approach is given in Fig.1. Experiments feed a data base. The software sensor integrates data from the data base, an ontology and a mathematical model. Its outputs can be analyzed using data analysis. This analysis also calls for ecophysiological knowledge, essentially about the phenological stages. Therefore, the ontology is used at two different levels: software sensor design and data analysis supervision. Data analysis is performed on the basis of two complementary perspectives for determining relations between software sensor temporal output and product quality. The first line of work is to design scalar explanatory variables, by summarizing a period of interest in compliance with the ecophysiological knowledge. These variables can then be used as input to decision trees. The second line of work is to use recent advances in functional data analysis, so that the inputs to the statistical model are the temporal data as a whole.

In the following, the approach is illustrated with a software sensor to estimate vine water courses and their relation to grape quality. Nevertheless, the proposed methodology is generic in many aspects, and could be useful in the development of other decision support tools in Agronomy, provided that expertise and temporal data are available.

2.1. Experimental design

Data used in this paper come from a multi-site experiment located in the south of France. The same experimental design was set up in seven sites across the Languedoc Roussillon region in order to test for the effects of vine water deficit status on grape potential and wine quality in contrasted environmental conditions. In total, vine water deficit status was followed over 16 vine plots, each planted with one of the following varieties: Merlot, Cabernet-Sauvignon, Grenache or Chardonnay.

To get a wider range of vine water statuses during the season, an irrigation treatment was applied for two years on each of the eight site-variety combinations. The irrigation treatment consisted of two modalities, replicated twice, yielding 32 experimental subplots. In the non irrigated subplots, vines only received natural precipitations during the growing season while in the irrigated subplots, vines received regular extra-amounts of water through drippers line (emission rate from 2 to 4 $\mathrm{l.h^{-1}}$, 1 to 2 drippers per plant).

Several kinds of data, collected according to an experimental design, are available: local meteorological data, vine water deficit related measurements, phenological state assessments, as well as grape quality analyzes.

2.1.1. Meteorological data

Hourly meteorological data on wind speed ($\mathrm{km.h^{-1}}$), minimal, maximal and mean air temperature (°C), air humidity (%), solar radiation ($\mathrm{W.m^{-2}}$) and amounts of precipitations (mm) were extracted from local meteorological stations for each site.

Transformed data

Hourly vapor pressure deficit (VPD) and reference atmospheric evaporative demand (potential evapotranspiration ET_{ref}) were calculated according to methodologies referred to as FAO-56, see Allen et al. (1998). Calculation of reference atmospheric evaporative demand (ET_{ref} in $\mathrm{mm.d^{-1}}$) is based on Penman-Monteith formula.

Daily meteorological data were obtained from hourly data after a trapeze integration. Thermal time, *i.e.* the accumulation of growing degree days (GDD) from April 1^{st}, was calculated by daily integration of mean air temperature minus a base temperature of 10°C, which is considered as the simplest model to estimate vine phenology, see Parker et al. (2011).

2.1.2. Phenological data

The main phenological phases (budbreak, bloom, nouaison, veraison) were estimated visually in each experimental plot when 50% of the plants reached the stage. Bloom was observed when 50% of the clusters had the cap off. Nouaison was defined using the bloom stage, according to local expert knowledge (see Section 2.2.1). Veraison dates were recorded when 50% of the fruit had turned red.

2.1.3. Vine water status data

Vine water status was monitored by two kinds of measurements: discrete measurements of leaf water potential at predawn (Ψ_b, or predawn LWP) and continuous measurements of sap flow.

Leaf water potential at predawn

LWP measurements were conducted every week from the end of June to the end of August with a pressure chamber at predawn (between $\approx$ 3.00 am and $\approx$ 5.00 am).

Sap flow

The energy balance method (Sakuratani, 1981) was used to measure sap flow with Sap IP system (Dynamax, Houston, TX, USA). There is one variety per vineyard site. The vineyard site is divided into 2 irrigation treatments. Two vineyard rows were selected. One row represents one irrigation treatment. In each selected row, 2 vines were equipped with one sensor. Each sensor measured vine sap flow rate every 15 minutes. The 2 selected vines were within 25 meters of each other within the same row.

Sap flow rates measured on each vine were averaged on an hourly basis within each row. Total sap flow of each vine was calculated as the product of sap flux density and cross sectional sap wood area at the measurement point. Various expert methods were applied to filter out nighttime, weak and erroneous signals. Sap flow measurements were scaled at the plant level according to plant leaf area estimates corresponding to each sensor. The daily sap flow assumed to measure daily vine transpiration was computed by adding all hourly sap flow rates measured during the day. The volumetric flux per vine (g.h^{-1}) was converted into mm.h^{-1} taking into account the respective area of ground per vine. Daily vine transpiration will be noted $T(t)$.

2.1.4. Fruit composition quality data

Starting two weeks before harvest, fruit was sampled for each irrigation treatment in each vineyard. Fruit data was collected at three different dates. Fruit composition analysis focused on berry weight (g), sugar concentration (g.l^{-1}), acidity (g(H2SO4).l^{-1}), anthocyans and assimilable nitrogen (mg.l^{-1}).

2.2. Formalizing knowledge

In this section, our aim is to show how ontologies can be used to formalize domain knowledge. In information science, an ontology formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts.

Ontologies are becoming increasingly popular, due to the great amount of available (complex) data and to the need for model (qualitative) knowledge and structural information. This need first arose out of the development of the World Wide Web. However, there are still very few attempts to combine ontologies and statistical or data-driven models. This could be particularly useful in Life Sciences and Agronomy, see Villanueva-Rosales and Dumontier (2008); Thomopoulos et al. (2013); Destercke et al. (2013).

The main incentives for using ontologies, see Guarino et al. (2009), are the following ones:

1. To share a common understanding of structured information, as advocated in Musen (1992);
2. To explicit the specificities of domain knowledge;
3. To identify ambiguous or inappropriate model choices.

For the present work, a specific ontology has been built, in order to formalize the concepts and relations required to design a vine water deficit indicator and to analyze its impact on grape quality.

The general class diagram of the ontology, called Ontology of Vine Water Stress (OVWS), is shown on Figure 2 as a Unified Model Language (UML) diagram. It is composed of concepts, represented as rectangular boxes, and of relations, represented by arrows. Formally, the ontology Ω is defined as a tuple $\Omega = \{\mathcal{C}, \mathcal{R}\}$ where $\mathcal{C}$ is a set of concepts and $\mathcal{R}$ is a set of relations.

Let us comment the main concepts and relations.

2.2.1. Concepts

In this ontology, four kinds of primary concepts were defined: *Variable*, *Condition*, *Constraint* and *ShiftStage*. All other concepts are sub-concepts of these primary ones and linked to them by a *subsumption* relation, as ex-

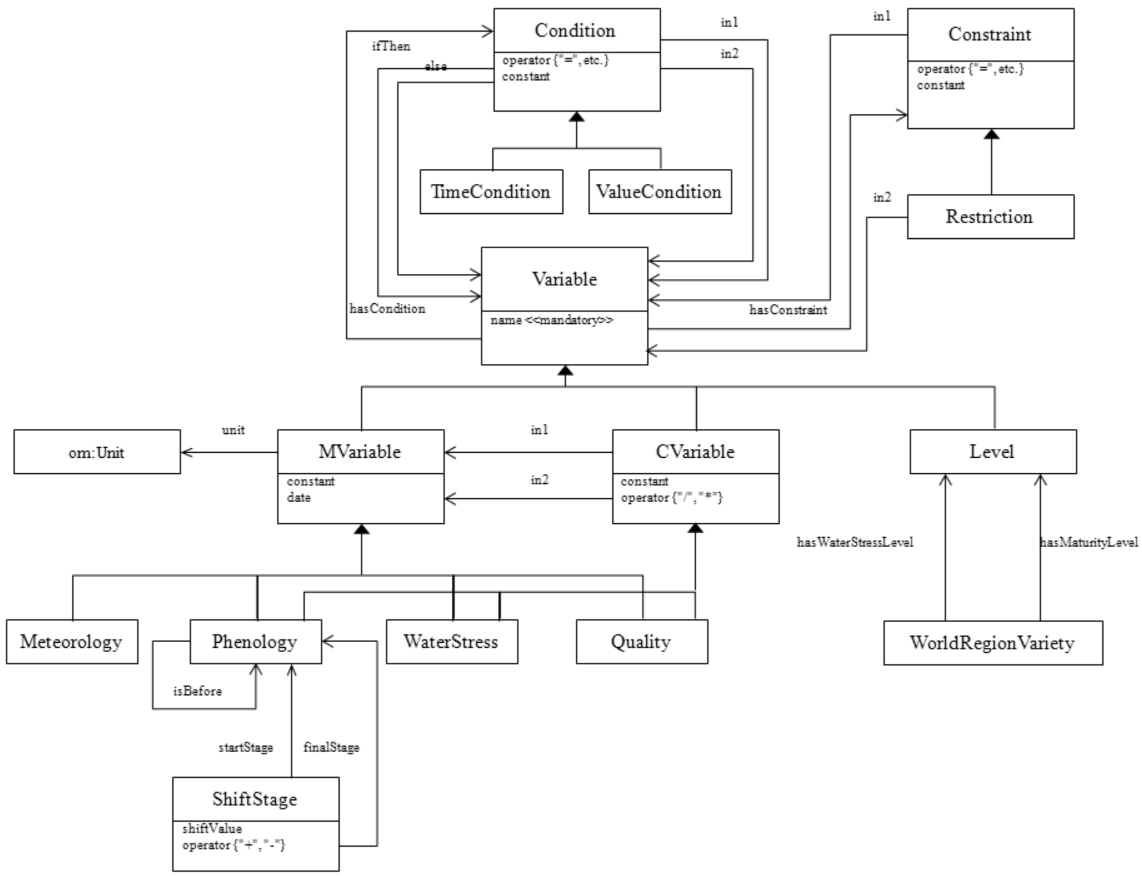


Figure 2: Class diagram of the ontology of Vine Water Stress.

plained in Section 2.2.2. For instance, in Figure 2, *Meteorology*, *Phenology*, *WaterStress* and *Quality* are sub-concepts of *Variable*.

- All *variables* must have a name, they can have a date, a unit and a default value. The units are taken from *OM*, an ontology of units of measures and related concepts, see Rijgersberg et al. (2013).

- The *Condition* concept is defined with a comparison operator and two operands. It will be used together with the *hasCondition* ($\mathcal{HCO}_c$) relation, defined in Section 2.2.2.

- The *Constraint* concept is defined with a comparison operator and one operand. It will be used together with the *hasConstraint* ($\mathcal{HCS}_c$) relation, defined in Section 2.2.2. The *Restriction* concept is a sub-concept of *Constraint*, and is a specific two-fold constraint.

- The *ShiftStage* concept is proposed in order to determine a phenological stage from another one. This is the case for the *Nouaison* stage, which is not generally observed. Its date can be estimated by shifting the *Bloom* date by k GDD, where k can be variety-dependent. *Nouaison* and *Bloom* are instances of the *Phenology* concept.

2.2.2. Relations

On Fig. 2, there are two kinds of arrows: thick-headed arrows and regular ones. The former correspond to the *subsumption* relation, and the latter to the other relations. In that last case, the arrow label gives the relation name, for instance *hasCondition*.

- The *subsumption* relation, also called the 'kind of' relation and denoted by $\preceq$, defines a partial order over $\mathcal{C}$. Given a concept $c \in \mathcal{C}$, we denote by $\mathcal{C}_c$ the set of sub-concepts of c, such that:

$$\mathcal{C}_c = \{c' \in \mathcal{C} | c' \preceq c\}. \tag{1}$$

  For example, in Figure 2, let us consider the concept $c = \textit{Variable}$. We have $\mathcal{C}_{\text{Variable}} = \{\textit{MVariable}, \textit{CVariable}, \textit{Level}\}$, where *MVariable* represents a measurement available in a data base, *CVariable* a variable calculated following a given method, and *Level* a constant value depending on some other concepts.

- The *subsumption* relation can be multiple. For instance a *Phenological concept* can be such as $c \preceq CVariable$ or $c \preceq MVariable$.

- The *isBefore* relation allows to represent temporal precedence. It is very important for checking the consistency of the phenological stage dates, where *bloom* has to occur before *veraison*, and so on.

- The HasCondition ($\mathcal{HCO}_c$) relation, where c represents the concept on which the condition is to be applied, is used together with a condition.

- Similarly, the HasConstraint ($\mathcal{HCS}_c$) relation allows the application of a *constraint* on the c concept.

In Section 2.3.2, examples will be given to illustrate the interest of the ontology for designing the software sensor.

The ontology is modeled using the Web Ontology Language (OWL). OWL is a semantic markup language for publishing and sharing ontologies on the

World Wide Web, which is specified using W3C[1] recommendations. The use of OWL allows reusing ontologies developed elsewhere, for instance the Ontology of units of Measure *(OM)*[2].

2.3. Design of the software sensor for vine water deficit estimation

Based on the knowledge formalized in the ontology given in Fig.2 and on a mathematical model, established by Ferreira et al. (2012), a software sensor is required to transform raw data from sap flow sensors into a significant vine water deficit estimator, denoted by $Ks(t)$.

The software sensor is a relatively complex information system that performs different functions and associates various technologies. Its design can benefit from using a conceptual framework including several viewpoints, such as the ones proposed by a *"4+1"* viewpoint set, introduced by Kruchten (1995) or RDM-OP approach (Reference Model for Open Distributed Processing), described in Raymond (1995).

For instance, the RDM-OP framework defines a set of five viewpoints: enterprise, information, computational, engineering and technology. Among them, the information view describes the way that the architecture stores, manipulates, manages, and distributes information. The computational view, presented in Fig.3, contains an object-oriented model of the functional structure of the system, with a particular focus on interfaces and interactions. Each component (rectangle) is a modular part of the system, interfaces are represented by connectors with circles, and dependency between components

[1]http://www.w3.org/TR/

[2]http://www.wurvoc.org/vocabularies/om-1.6/

is illustrated by dashed arrows.

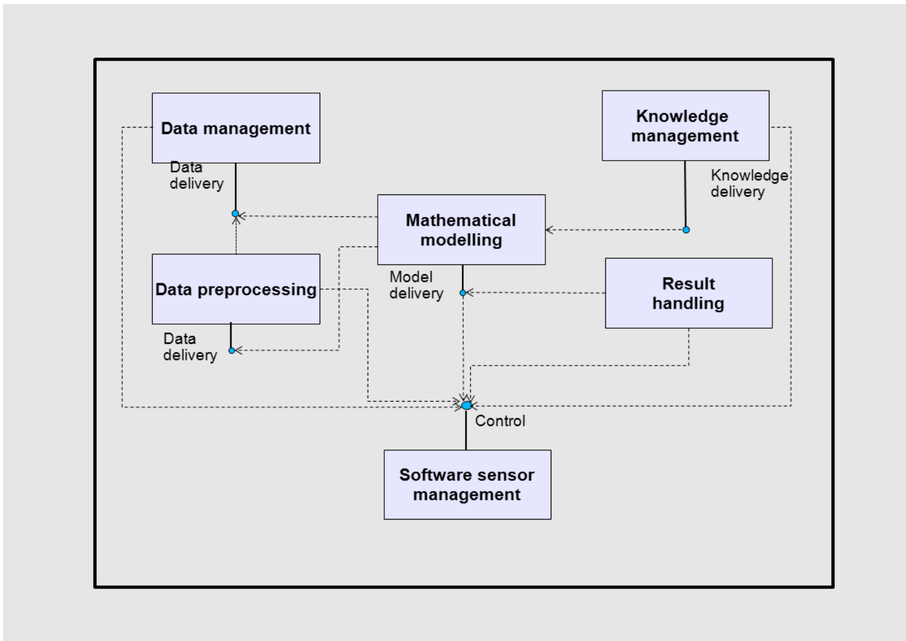


Figure 3: Software sensor computational view.

Note that each component can be modeled independently, using a suitable language (sql, OWL, R). The communication between the various components can be implemented by a high level interface, written in Python, PHP or Java, according to the technology viewpoint. In its present implementation, the software sensor is available as a desktop application to the members of the *Pilotype* project (see Acknowledgments).

More details about the components of the desktop application are given in Fig.4 and the various steps to follow for estimating the vine water deficit (Ks) are detailed below.

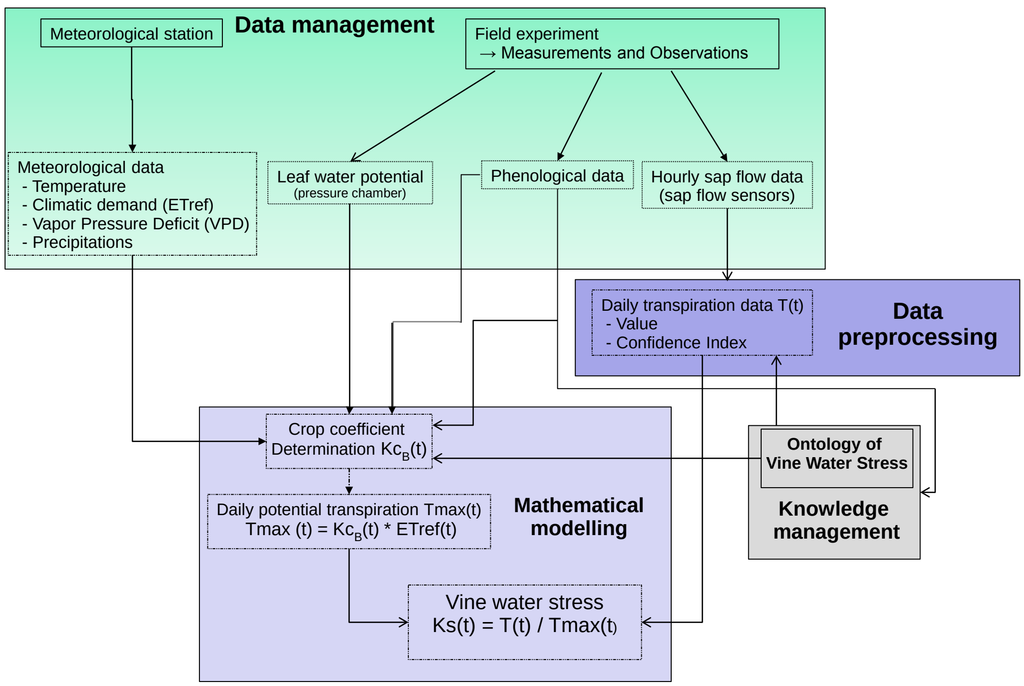


Figure 4: Software sensor component details for the vine water deficit (Ks) estimation.

2.3.1. *Sap flow under limiting soil water condition: computation of Ks*

Ks is the ratio between actual and maximum crop transpiration, defined as:

$$Ks(t) = \frac{T(t)}{T_{max}(t)} \quad (2)$$

It accounts for the decline in vine water use due to soil moisture deficit. Ks represents the level of daily vine water use by reference to its maximal level. $Ks = 1$ reflects a situation when maximal level of vine water use is fully reached/ satisfied. When $Ks < 1$, the maximal level of vine water use is not reached. Daily vine water use is limited and Ks level indicates some level of water deficit. Arbitrarily we characterize this situation as a 'stress'. When $Ks = 0$, stress is maximal.

Allen et al. (1998) have presented a general proposal for estimating Ks. Ferreira et al. (2012) have reported results showing the variations of specific Ks in vineyard subjected to contrasted soil moisture regimes. Functions for vineyards, from field experiments, are not generally available.

In the vine context, in Eq.2, T is the daily measured transpiration from sap flow and T_{max} is the daily maximal vine transpiration obtained under dry soil condition (meaning no cover crop) when soil moisture is non limiting, defined as in Allen et al. (1998).

$$T_{max}(t) = Kc_B(t)\ ET_{ref}(t) \quad (3)$$

ET_{ref} is the reference evapotranspiration and Kc_B a coefficient linearly related to the leaf area index (LAI) or to the fraction of ground coverage, see Picón-Toro et al. (2012). Consequently a site-specific determination of Kc_B

is necessary for each vineyard to account for differences due to canopy size and planting density.

2.3.2. Sap flow under non limiting soil water condition : computation of dry soil Kc_B

$Kc_B(t)$ is vine specific and varies with leaf area development. When $Kc_B(t)$ is multiplied by $ET_{ref}(t)$, it yields an estimate of plant maximal transpiration, which is the volume of vine water use in absence of soil moisture deficit (see Eq.3). We propose to use formalized concepts and relations based on expertise, all of them implemented in the OVWS ontology. We divided the $Kc_B(t)$ profile into two main growth stages: L_{dev} and L_{mid} as presented in Fig.5. This profile is derived from the FAO segmented crop profile for 2 growing stages (development period and mid-season period) as reported by Allen and Pereira (2009). L_{dev} corresponds to the period during which leaf area is growing at a fast rate, linearly with thermal time (the *grand growth period*). L_{mid} corresponds to the period during which leaf area does not grow anymore (because of natural shoot growth cessation or due to mechanical hedging cutting away the growing points).

To determine $Kc_B(t)$, two hypotheses on the curve shape are assumed:

$$Kc_B(t) = f(t) \text{ for } t < t_{K^*} \tag{4}$$

$$Kc_B(t) = K^* \text{ for } t \geq t_{K^*} \tag{5}$$

where $f(t)$ is assumed to be linear in t, and t_{K^*} is the breakpoint for which Kc_B reaches the plateau K^*. The key point is to set t_{K^*}, or indifferently K^*. According to Eq.3, we make the hypothesis that, in the absence of water deficit, K^* is defined as:

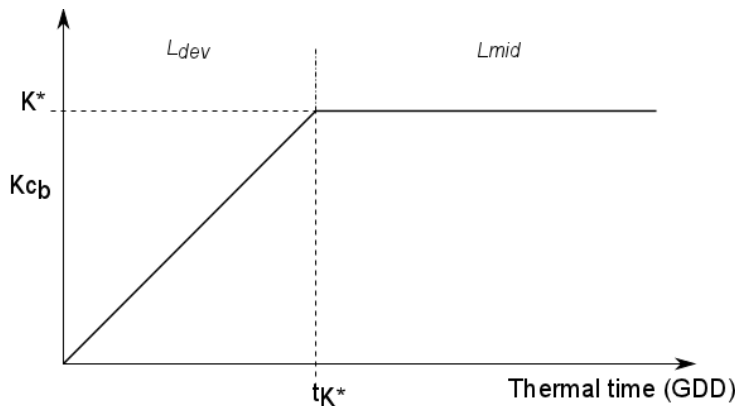


Figure 5: Theoretical curve of Kc_B evolution during the season.

$$K^* = \frac{T(t_{K^*})}{ET_{ref}(t_{K^*})} \tag{6}$$

Using the OVWS ontology defined in Section 2.2, the following rules are set up to automatically define a limited number of potential options for t_{K^*}. The interest of having the rules and concepts defined in an ontology is two-fold: *i*) they have to be completely explicit, *ii*) they can evolve independently of the numerical procedures.

1. *Selection based on phenology*

   A linear relationship exists between Kc_B variations and leaf area index (LAI) or the fraction of ground covered by the vine, see Ferreira et al. (2012). We thus assume that peak Kc_B (i.e. K^*) is reached when LAI stops increasing. Consequently, the search period for K^* has been limited to the period between budbreak and veraison.

   These two concepts are defined as sub-concepts of the Phenology concept, itself being a sub-concept of *MVariable*. The period limitation is instantiated by two *TimeConditions*, applied onto the $Kc_B \preceq CVariable$

concept, the $\mathcal{HCO}_c$ relation where the *Condition* is characterized by a comparison operator $\leq$ (resp. $\geq$) and the *Veraison* (resp. *Budbreak*) concept.

2. *Selection based on predawn leaf water potential*

   Conditions of maximal soil moisture availability could be inferred from predawn leaf water potential measurements, associated with a confidence interval derived from VPD. A rule was set so that K^* has to be reached before the first day at which predawn LWP measurement reveals a water deficit level limiting shoot elongation. The levels to which predawn LWP characterizes that limiting effect can be defined by the stakeholder, or else set in agreement with a standard level, based on a region or/and variety.

   This is implemented in the ontology by the *WorldregionVariety* and $level \preceq Variable$ concepts.

3. *Selection based on meteorology*

   Transpiration measurement through sap flow is sensitive to climatic conditions, mainly light and VPD. To account for sensitivity of transpiration measurements to $VPD(t)$, a filtering rule was set to remove computed $Kc_B(t)$ obtained in situations of heat spikes, defined as period with *VPD* greater than a given level, set to 3.5 k.Pa in the present case study.

   The rule is implemented using a $\mathcal{HCS}_c$ relation, applied onto the $Kc_B \preceq CVariable$ concept, where the *Constraint* is characterized by a comparison operator $\leq$ and the $VPD \preceq Meteorology \preceq MVariable$ concept.

4. *Selection based on curve shape*

By definition, K^* is reached when the ratio $\dfrac{T(t)}{ET_{ref}(t)}$ reaches a maximum during a few days (as $T(t)=T_{max}(t)$ and $Kc_B(t)=K^*$) and then decreases (as $T(t)<T_{max}(t)$ due to limiting soil water conditions while $Kc_B(t)=K^*$). As such, potential options for t_{K^*} have been defined at points with a null first derivative and a negative second derivative. This selection is implemented using two concepts: $\dot{K}^*$ and $\ddot{K}^*$, both such as $\preceq WaterStress \preceq CVariable$, and a $\mathcal{HCS}_c$ relation with a $Constraint$ characterized by a comparison operator $\leq \epsilon$ or ≤ 0.

User selection based on expert knowledge.

The analysis of $\dfrac{T(t)}{ET_{ref}(t)}$ curve shape, associated with all previous rules based on phenology, meteorology and predawn leaf water potential leads to the proposal of a small finite set of t_{K^*} candidates.

The final choice is left to the stakeholder who is the best aware of the management practices or particular uncontrolled events that could have interfered with vine growth (irrigation, leaf removal, trellis system. . .) and therefore with the Kc_B curve.

2.4. Relating software sensor output to product quality

The software sensor output consists of temporal data, and various methods can be used to study the relationships between these data and product quality. Two complementary lines of work, based on statistical methods, are explored in the present work. The first one consists of extracting significant scalar parameters from the temporal data and using them as input to decision trees, in order to provide the most discriminant features. The second one uses functional data analysis, that gives the possibility to model the temporal data impact on product quality as a whole. However, curve analysis is a recent research topic, with relatively few methods available, in comparison with classical data analysis.

This section is divided into three parts. The first part describes how to use the formalized knowledge for extracting significant scalar parameters from the temporal data. The other two parts give some elements necessary to understand the statistical methods that will be used: decision trees and functional data analysis.

2.4.1. Extracting scalar parameters from software sensor output

Meaningful scalar parameters can be extracted from temporal courses determined by the software sensor outputs. In many cases, expert knowledge can be the support of such extraction procedures. In the case of vine water courses, this can be achieved by taking into account important phenological periods, which are defined as concepts in the ontology (see Fig.2).

Three periods were first defined according to phenological stages: the whole season, the pre-veraison period, which goes from the nouaison stage to the veraison stage, and the post-veraison period which ranges from the veraison stage to the harvest date. In a second step, the post-veraison period was divided by taking into account the maturity stage of berries, which allowed to a add a fourth period ranging from veraison to maturity. Maturity stage is reached when the ratio between Sugar Concentration and acidity in grapes yields a given threshold, defined according to variety.

Using trapeze integration under Ks curves over these four periods, the continuous $Ks(t)$ curve was summarized into four new variables corresponding to the cumulative amount of stress encountered by the vine over these periods: *NouHarv*, *NouVer*, *VerHarv* and *VerMat*. Table 1 gives the summary of these four aggregated variables for each plot. Since all these aggregated variables are based on the area under the curve, the lower their value, the stronger the water stress over the considered period.

2.4.2. Decision trees as interpretable models

Decision tree algorithms are well established learning methods in supervised data mining and statistical multivariate analysis. They allow to display non linear relationships between features and their impact on a response variable, in a compact way.

Decision trees can handle classification problems or regression cases, depending on the nature of the response variable. Note that the CART family, see Breiman et al. (1984), based on binary splits, is mostly used by statisticians. There is another tree family, called ID3, see Quinlan (1986, 1993), allowing non binary splits and mostly used by artificial intelligence researchers.

We recall here the principle of the regression case, where the response variable is numerical.

Input to regression decision trees consists of a collection of N training cases, each having a tuple of values for a set of P input variables, and one continuous output variable $(\mathbf{x_i}, \mathbf{y_i}) = (x_{1,i}, x_{2,i} \dots x_{P,i}, y_i)$. An input X_p $(p = 1...P)$ is continuous or discrete and takes its values $(x_{p,i})_{i=1...N}$ on a domain $\mathcal{X}_p$. The goal is to learn from the training cases a recursive structure (taking the shape of a rooted tree) consisting of (i) leaf nodes labeled with a mean value and a standard deviation, and (ii) test nodes (each one associated to a given variable) that can have two or more outcomes, each

Site	Variety	Irrigation	NouHarv	NouVer	VerHarv	VerMat
LB-CS	Cabernet-S.	i_0	1165.6	704.6	432.6	432.6
		i_1	1117.8	713.8	375.0	347.6
OUV-Mer	Merlot	i_0	742.1	457.2	278.3	163.9
		i_1	1233.8	814.7	403.4	247.9
StGER-Mer	Merlot	i_0	608.9	470.9	131.6	129.4
		i_1	808.3	473.4	327.0	214.5
PR-Mer	Merlot	i_0	655.4	381.8	250.2	199.5
		i_1	722.7	398.3	313.4	266.2
StSAU-Char	Chardonnay	i_0	695.3	442.6	241.9	213.6
		i_1	693.8	414.0	265.3	253.7
RIE-Gre-Chm	Grenache	i_0	651.1	465.2	169.9	169.9
		i_1	620.0	390.9	209.5	209.5
RIE-Gre-Chp	Grenache	i_0	512.3	380.0	123.0	NA
		i_1	963.9	580.2	362.0	362.0
PIO-Gre	Grenache	i_0	677.2	458.4	212.0	152.3
		i_1	887.4	514.4	363.8	180.8

Table 1: Values of aggregated variables for each site-variety-irrigation treatment combination (i) over the entire season *NouHarv*, (ii) before veraison *NouVer*, (iii) after veraison *VerHarv* and (iv) from veraison to maturity *VerMat*.

of these linked to a sub-tree.

On a given node, the algorithm examines in turn all available variables, and selects the variable that most effectively splits the set of samples into subsets improving the separation between output values. Once (and if) a variable is selected, a new test node is created that splits on this variable, and the procedure is recursively applied on each (new) node child. At each node, the algorithm stops when no more variables are available, or if there is no improvement by splitting further: the node then becomes a leaf.

Decision trees are easily interpretable for a non-expert in statistical or learning methods, and facilitate exchanges with the domain expert. A low complexity, see Ben-David and Sterling (2006), is essential for the model to be interpretable, as confirmed by Miller's conclusions, see Miller (1956), relative to the *magical number* seven.

Well-known drawbacks of decision trees are the sensitivity to outliers and the risk of overfitting. To avoid overfitting, cross-validation is included in the procedure and to gain in robustness, a pruning step usually follows the tree growing step, see Quinlan (1986); Breiman et al. (1984); Quinlan (1993).

In this work, we used CART-based trees. In that case, the splitting criterion is based on finding the one predictor variable (and a given threshold of that variable) that results in the greatest change in explained deviance (for Gaussian error, this is equivalent to maximizing the between-group sum of squares, as in an ANOVA). This is done using an exhaustive search of all possible threshold values for each predictor. The implementation used for decision trees is the R software, described in R Development Core Team

(2009), with the *rpart* package[3].

Specifying variety, *NouVer*, *VerHarv* and *VerMat* as explanatory variables, we performed decision trees on maximum values of grape quality features over the season.

2.4.3. Functional data analysis

Functional linear regression is an approach to model the relationship between a scalar dependent variable Y and a functional predictor $X(t)$, a function of a real variable t (time for example). The model is written as

$$Y_i = \beta_0 + \int X_i(t)\beta(t)dt + \varepsilon_i\,, \quad i = 1, \ldots, n \tag{7}$$

where ε_i is a random error, β_0 is the intercept of the model and $\beta(t)$ is the coefficient function, both unknown and to be estimated from independent observations $(X_i(t), Y_i)_{i=1,\ldots,n}$. In this model, $\beta(t)$ determines the effect of $X_i(t)$ on Y_i. For example, $X_i(t)$ has a greater effect on Y_i over regions of t where $|\beta(t)|$ is large. On the opposite, $X_i(t)$ has no effect on Y_i over regions of t where $\beta(t)$ is zero. Estimating $\beta(t)$ in Eq.7 has given rise to an increasing literature in the last decade, see for example Ramsay and Silverman (2005). A common approach involves projecting β and the X_i's in a p-dimensional basis function where p is large enough to capture the unknown variations of β, but small enough to regularize the fit. Such techniques are not sufficient to produce estimates of $\beta(t)$ that are exactly zero in the regions of t where $X_i(t)$ has no effect on Y_i.

Recently, James et al. (2009) introduced new estimators that are both

[3] http://cran.r-project.org/web/packages/rpart/index.html

interpretable, flexible and accurate. The method, called "Functional Linear Regression That's Interpretable" (FLRTI), is based on a particular basis function and variable selection techniques. The time-period is divided into a fine grid of points $(t_j)_{j=1,..,p}$. The β function is assumed to be exactly zero over some time periods and exactly linear over the remaining periods, period location being unknown. The reason behind the first assumption is that all the $X_i(t)$ observations, for varying t, are not of equal importance to explain the response Y_i since $X_i(t_j)$ has no effect on Y_i when $\beta(t_j) = 0$. Hence the β function is assumed to be sparse. The second assumption is made for obtaining an easily interpretable β function, it is implicitly equivalent to the assumption that the second derivative of $\beta(t)$ will be zero over these regions of t, that is, the second derivative is assumed to be sparse.

These assumptions will constraint the estimation of the regression model (Eq.7), which corresponds to a penalized regression in sparse models, with a number of time grid points p much larger than the number of observations n. To estimate the β function at each point t_j, it is necessary to minimize the mean squared error criterion subject to a regularity constraint. The Dantzig selector is the solution to this problem used by the authors of FLRTI. The whole method is implemented in an R function available on J. Gareth's web page[4]. Finally, two tuning parameters have to be fixed, a penalty term σ and a weight ω. The penalty term is part of the Dantzig selector procedure. The largest the σ, the more the form-related constraint is enforced. The weight ω impacts the relative number of zeros of the β function. A weight equal to

[4]http://www-bcf.usc.edu/ gareth/research/flrti

0 indicates that only the linear form constraint is respected, no assumption is made on the sparsity of β.

A cross-validation algorithm is also proposed to optimize the choice of σ and ω. The cross-validation procedure aims at estimating optimal values for σ and ω, from two sets of possible values for them $(\sigma_k)_k$ and $(\omega_l)_l$. The principle is to divide the data set into N_f folds (typically 10). All folds but one are used to train the estimation process with each combination of (σ_k, ω_l). The excluded fold is used to test the estimated model, yielding an error for each (σ_k, ω_l). This is repeated until all folds have been used once for testing. At the end, we obtain N_f errors for each combination (σ_k, ω_l), whose mean yields a cross-validated error for each (σ_k, ω_l). The optimal choice of σ and ω is the couple with the smallest cross-validated error.

3. Results

In this section, we first present the results of the $Ks(t)$ estimation using the software sensor. In a second step, we study the relationship between $Ks(t)$ and grape quality features, using the methods described in Section 2. In the following, we will refer to irrigated treatments with i_1, and to non irrigated ones with i_0.

3.1. Vine water stress course $Ks(t)$ estimation

Sap flow data require a pre-treatment, including sensor selection and signal smoothing. Sap flow sensors have only been used recently in European vineyards. Thus, calibration protocols are not established yet and therefore sensors can still be unreliable. Consequently, a selection step is required.

Sensor reliability has been assessed on the basis of the number of incorrect hourly measurements resulting from expert filtering methods. A sensor was considered reliable when less than 5% of the hourly data were filtered out. For each variety-irrigation combination, the mean daily vine transpiration was calculated as the mean of daily measures from reliable sensors, which helped limiting the variability in plant transpiration measurements. However, one of the major drawback of sensor selection was the potential lack of reliable measurements on a daily basis.

To capture important patterns in daily sap flow data, while leaving out noise and extreme variations (daily peak), sap flow courses were smoothed with the central moving average method with a five day window. This smoothing allowed the removal of missing values and extreme peaks.

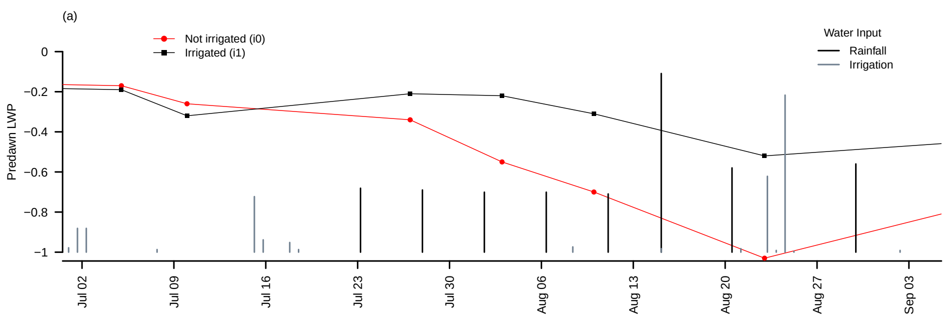
3.1.1. K^* determination

Regarding all site-variety combinations, the knowledge-based algorithm for t_{K^*} determination proposed from 5 to 9 candidates (resp. from 4 to 8 candidates) in the non-irrigated i_0 (resp. irrigated i_1) treatments. Most of the dates proposed by the mathematical algorithm were in accordance with expert knowledge, so allowing the expert to choose t_{K^*} within the algorithm suggestions (Fig.6). The results are given in Table 2. Fig.6 illustrates the results for the Grenache variety at the Piolenc site.

The validity of the K^* determination procedure can be assessed according to different points. First, the results regarding K^* determination based on the coupling of mathematical algorithms and expert knowledge were consistent with existing literature. Indeed, most of t_{K^*} occurred between 600 and 700 GDD after budbreak (Table 2), which is in accordance with t_{K^*} reported in

Site	Variety	Irrigation	K^*	t_{K^*} (GDD)	First irrigation(GDD)
La Baume	CS	i0	20.3	677.9	
		i1	32	698.6	1268.3
Pech Rouge	Merlot	i0	19.4	614.5	
		i1	26,6	614.5	610.4
St Gervasy	Merlot	i0	69.3	625.3	
		i1	85.1	669.4	844
Ouveillan	Merlot	i0	37.1	829.2	
		i1	21	1005.1	939.7
Piolenc	Grenache	i0	44.3	530.1	
		i1	58.1	530.1	864.4
Rieux	Grenache+	i0	43.4	600.5	
		i1	29.1	594.5	789.5
Rieux	Grenache-	i0	29.2	580	
		i1	46.1	580	789.5
St Sauveur	Chardonnay	i0	43.3	642	
		i1	54.4	749.7	777.2

Table 2: Values of basal crop coefficients K^* and dates (t_{K^*} in GDD) at which they were estimated in the site-variety-irrigation treatment combinations during season 2012.



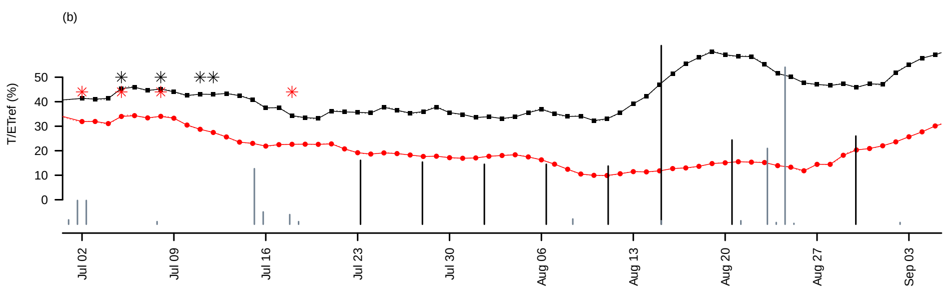


Figure 6: (a) Vine water status as indicated by leaf water potential and (b) determination of K^* and t_{K^*} (stars) based on expert knowledge following the mathematical algorithm suggestions for non irrigated (red bullets) and irrigated (black squares) treatments on the Grenache variety at Piolenc site.

Picón-Toro et al. (2012) from a 3 year study in western Spain on Tempranillo, and in FAO-56, see Allen et al. (1998), that respectively reported t_{K^*} around 650 GDD and 555-592 GDD after budbreak.

3.1.2. Maximal transpiration and Ks estimation

Following determination of K^* and t_{K^*}, $Kc_B(t)$ was calculated over all t values (Eq.4). Its variation for a Grenache variety is plotted on Fig.7, both in calendar time (a) and thermal time since budbreak (b).

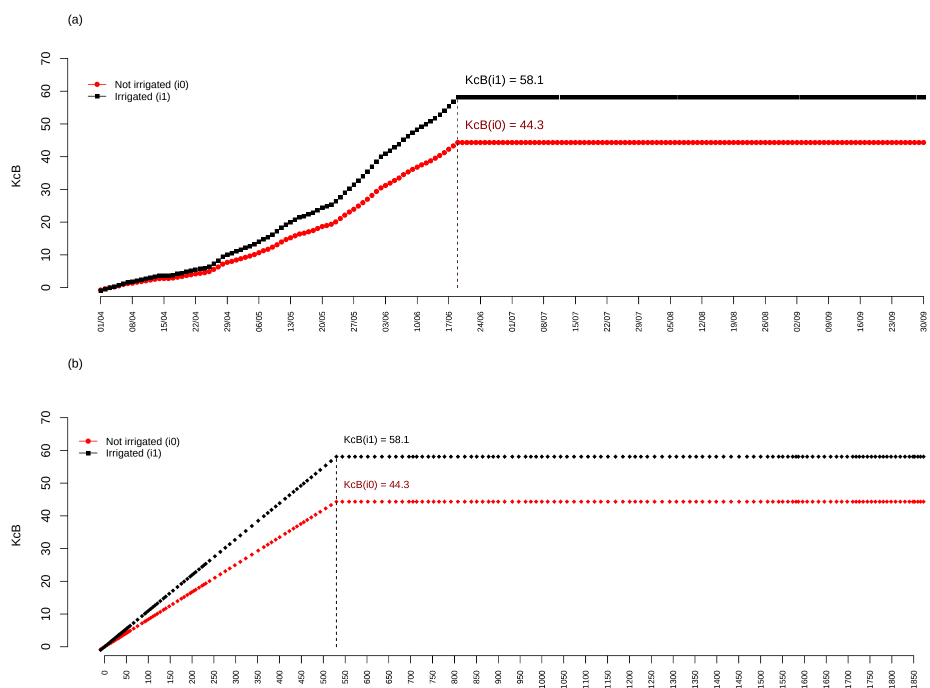


Figure 7: Evolution of vine basal crop coefficient (Kc_B) during the season at Piolenc site with Grenache variety, from budbreak to harvest. (a) x-scale in Julian days - (b) x-scale in GDD since budbreak.

$Kc_B(t)$ was then used to calculate the daily vine maximal transpiration (T_{max}), according to Eq.3. Finally, $Ks(t)$ was calculated as the daily ratio of measured transpiration by reliable sensors over potential transpiration (Eq.2).

Figure 8 shows vine water status according to both indicators: (a) Predawn LWP and (b) Ks during the season 2012 in a Grenache variety of the Languedoc-Roussillon region.

3.2. Relationships between vine water stress $Ks(t)$ and grape quality

As explained in Section 2.4, $Ks(t)$ can be used in two different ways, either summarized as a series of scalar values, or as a whole. The way to summarize $Ks(t)$ is detailed in Section 2.4.1. Scalar values and $Ks(t)$ will be put in relation to grape quality at harvest time, by the respective use of (i) regression trees and (ii) functional data analysis. The studied grape quality features include *i)* Berry Weight and *ii)* Sugar Concentration in berries. For interpreting the results, note that $Ks(t)$ is inversely related to water deficit.

3.2.1. Regression trees

Aggregated variables over periods can be used as explanatory variables in regression trees to detect and prioritize the periods critical to changes in grape quality. We studied the effects of *NouVer*, *VerHarv*, *VerMat* and variety on the two components of grape quality cited above. The corresponding regression trees are displayed in Fig.9 and Fig.10, together with the distribution of values at terminal nodes, represented by boxplots. Table 3 shows the gaining in deviance for each splitting step during the tree generation. The number of available samples being small (16), the minimum number of

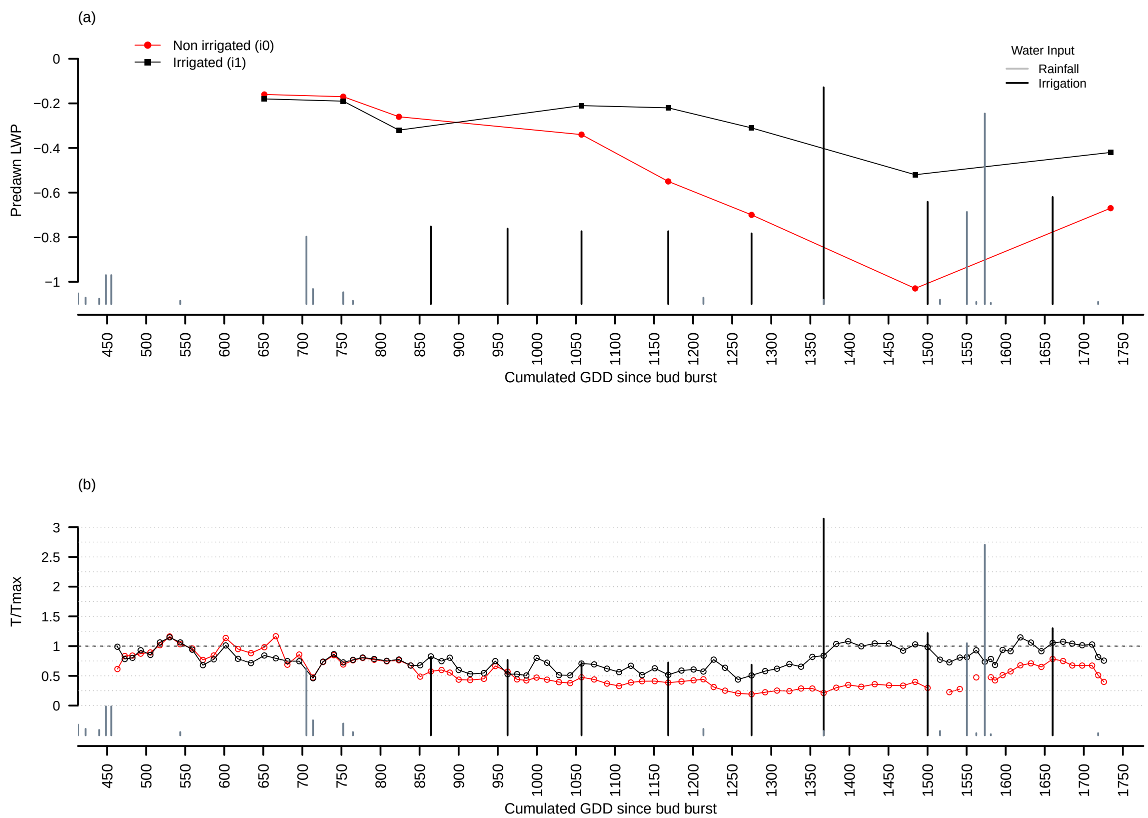


Figure 8: Water deficit during 2012 millesim in Grenache variety at Piolenc site assessed by (a) Predawn LWP and (b) vine water stress indicator Ks.

observations in any terminal node was set to 1. That is not sufficient to support prediction with a good confidence level, but is still interesting for summarizing the data.

Regression tree	split 1	split 2	split 3	split 4
Berry weight	0.64	0.63	0.69	
Sugar concentration	0.44	0.39	0.87	0.33

Table 3: Gain in deviance during regression tree generation.

According to Fig.9, Berry Weight seems to be mostly affected by the variety (Fig.9). Grenache variety significantly yields heavier berries. The second split for all varieties is done on the post-veraison water stress only (either *VerHarv* or *VerMat*). The more severe is water deficit *post veraison*, the smaller is the Berry Weight.

Regarding Sugar Concentration, regression trees show that it is affected by water stress in both pre-veraison *NouVer* and post-veraison *VerHarv* periods (Fig.10). The first discriminant variable on Sugar Concentration is the post-veraison water stress (*VerHarv*, Fig.10). The first split shows that a higher post-veraison water stress leads to a lower Sugar Concentration.

The left branch resulting from the first split shows that the next discriminant variable is again the post-veraison stress *VerHarv*, so enhancing the effect of the previous split. Lastly, pre-veraison water stress (*NouVer*) can exacerbate the decrease in Sugar Concentration (as shown at the tree bottom).

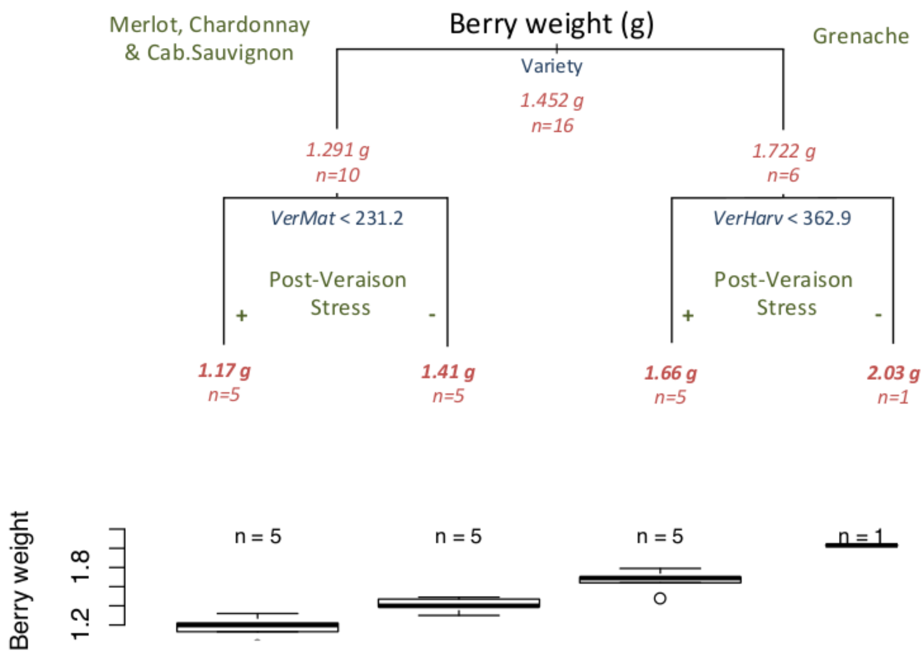


Figure 9: Regression tree explaining Berry Weight (g) using scalars summarizing the three periods, i.e. pre-veraison ($NouVer$), and post-veraison either until maturity ($VerMat$) or harvest ($VerHarv$). Boxplots showing the distribution of values at terminal nodes are displayed below the tree.

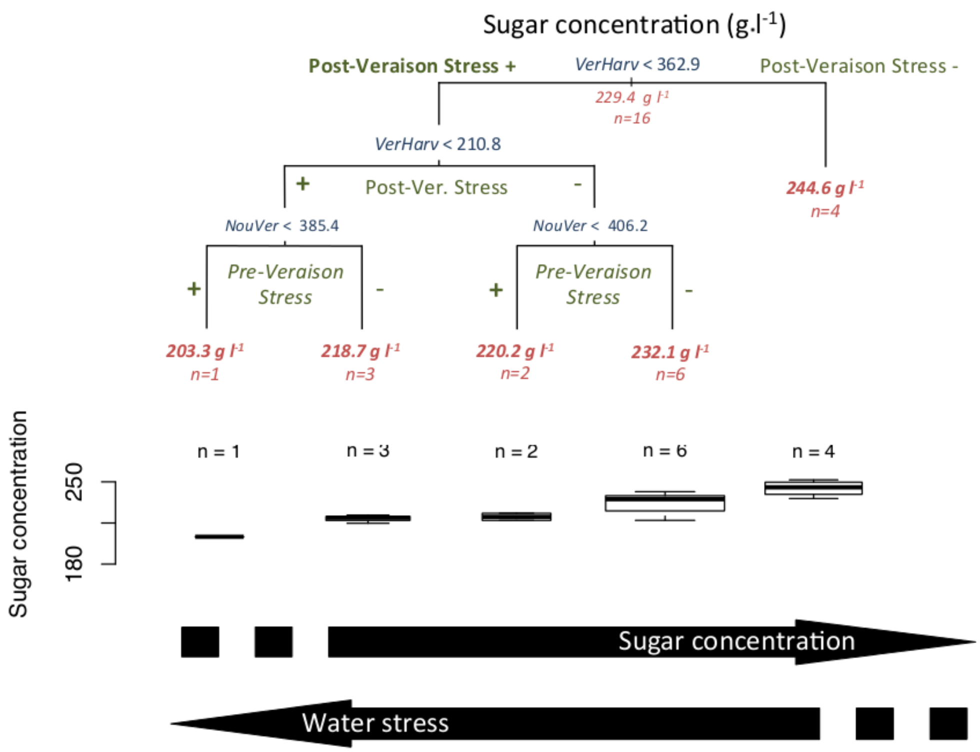


Figure 10: Regression tree explaining Sugar Concentration in berries (g.l^{-1}) using scalars summarizing the three periods, i.e. pre-veraison (*NouVer*), and post-veraison either until maturity (*VerMat*) or harvest (*VerHarv*). Boxplots showing the distribution of values at terminal nodes are displayed below the tree.

3.2.2. Functional data analysis

Using a continuous indicator of water deficit enables the use of the whole season water deficit curve to explain berry composition. This in turn is likely to promote a more precise monitoring of vine water needs according to the targeted fruit composition. Using the FLRTI method, described in James et al. (2009), we analyzed the effects of the vine water deficit over the season on Berry Weight and Sugar Concentration in berry at harvest.

Regarding Berry Weight, the results showed no significant effect of $Ks(t)$. This was confirmed by applying a testing procedure designed to test the nullity of the β function in a generic functional linear model with scalar output like the one given in Eq.7. The literature on such tests is scarce. We applied the one introduced in Hilgert et al. (2013), which has the particularity of not requiring any prior knowledge on the β function. A p-value of 0.7 of the procedure was estimated by Monte-Carlo simulations (with 10 000 runs). The fact that $Ks(t)$ has no significant effect on Berry Weight might be due to the non taking into account of the variety effect in the model, which is very important to explain Berry Weight. On top of that, it may be possible that the level of water deficit is not severe enough to induce changes in Berry Weight or that the timing of water deficit happens too late in the season to have an effect at limiting berry size, see Ojeda et al. (2002). Decision trees highlighted an effect of post veraison water deficit on Berry Weight, but as a minor effect compared to the variety influence.

The functional data analysis on the effect of $Ks(t)$ on Sugar Concentration yields an estimation of the $\beta(t)$ coefficient function, that is displayed in Fig.11. β_0, the intercept in Eq.7, is estimated at 178.3 g.l^{-1}. The tuning parameters are indicated in the legend. The goodness-of-fit of the estimated β curve is measured by a R^2 value, equal to 0.7. Since this coefficient measures the percentage of variation of the data explained by the fitted model, a R^2 equal to 0.7 is a rather high value in the context of penalized regression. A p-value of 0.02 of the testing procedure was estimated, in the same way than for Berry Weight. Residuals, plotted in Fig.12, showed a good repartition when plotted against predicted values, and no tendency. So the $Ks(t)$ curve appears to be a relevant variable to explain the Sugar Concentration. Let us also note that parameters were obtained following a ten-fold cross validation. A sensitivity analysis to small σ and ω variations showed a good robustness of the model, with three main peaks always located in the same time periods across the different varietals. These three main peaks are labeled (1), (2) and (3). Each of them corresponds to a significant effect of Ks on Sugar Concentration, which can be positive (peaks (1) and (3)) or negative (peak (2)).

Peaks (1) and (3) are positive, which implies a rise in Sugar Concentration. During these periods, the stronger the Ks value, the higher the rise. As Ks varies inversely with water deficit, it means that the lower the water deficit during these periods, the higher the rise in Sugar Concentration. The effect is twice as strong for the peak (1) than for the peak (3).

Regarding the time period, peak (1) appears to be located before pre-veraison whereas peak (3) occurred during pre-veraison. By contrast, peak (2) has a negative effect on Sugar Concentration. During this period located within the grand growth phase, a low water deficit decreases the Sugar Concentration. This can be reformulated as follows: the higher the deficit during

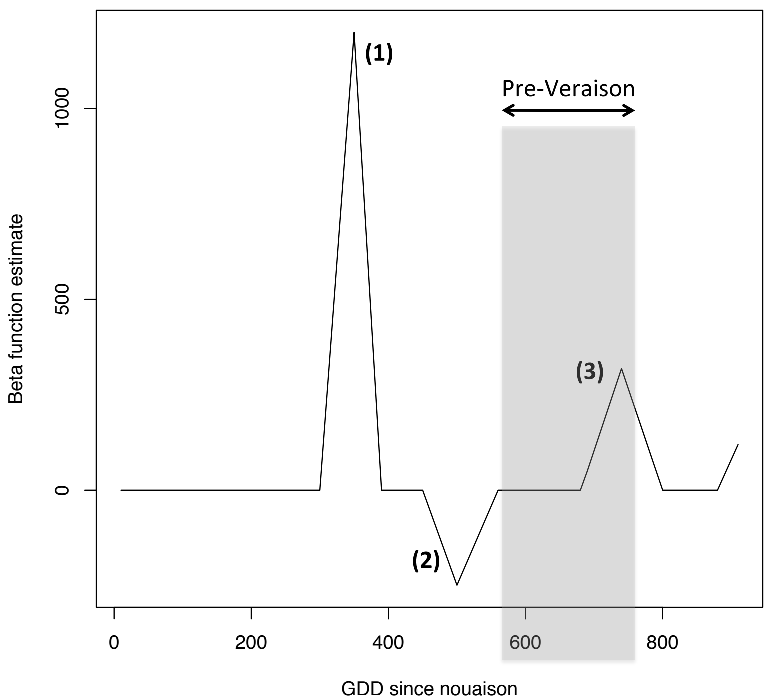


Figure 11: Beta function evolution over time (see Eq. 7), for explaining Sugar Concentration at harvest. Abscissae are in GDD. The values of $\sigma = 0.05$ and $\omega = 0.95$ have been found by cross-validation.

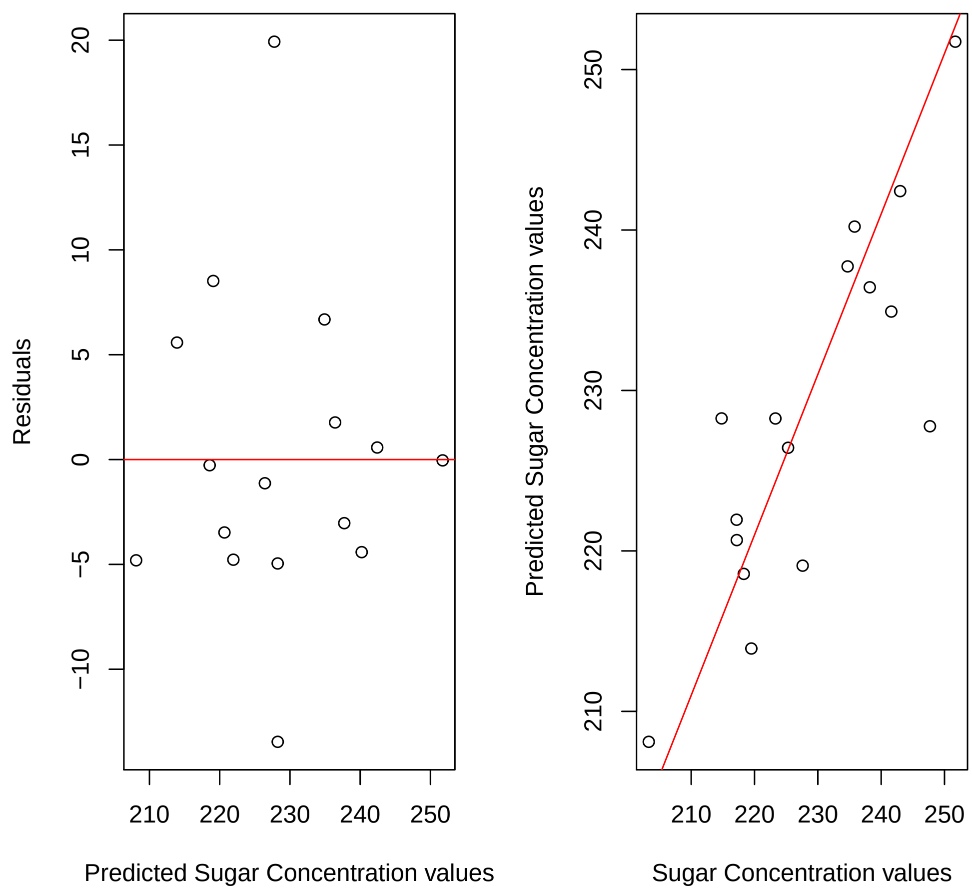


Figure 12: Plots of the residuals and the predicted Sugar Concentration values associated with the Beta function estimation. Abscissae are in g.l^{-1}.

that period of grand growth (i.e. before K^* is reached), the lower the Sugar Concentration. These results are consistent with the ones obtained by using decision trees (section 3.2.1, Fig.10), but more informative regarding the time period of interest.

4. Conclusion

The work presented in this paper used formalized knowledge and mathematical models to design a software sensor from raw data and relate its temporal output to product quality. The proposed approach has been applied to the case of a vine water deficit indicator, and its relation to two grape quality variables: Berry Weight and Sugar Concentration.

Results provide a number of meaningful insights.

First of all, the software sensor key point, which is the determination of K^*, seems reasonably consistent with the literature. From an agronomical point of view, this allows to effectively work at plot scale, and to offer decision support for irrigation, as a function of each plot characteristics.

The use of an ontology allows to separate expert knowledge and numerical models. It makes it much easier to build a generic model, that is both evolutive and adaptable over time as knowledge progresses or climate changes.

Contrary to a data base, an ontology schema adds semantics to the data structure, allowing automatic reasoning, using logical properties, such as reflexivity or transitivity.

The ontology presented here has a moderate complexity level: only four kinds of primary concepts, and five types of relations. This is still sufficient to express many mathematical conditions and dependencies, going well beyond the scope of the present case study. There may however be cases where new concepts and relations are necessary, and the ontology can easily be enriched when needed.

Second, the two-fold proposal for data analysis appears to be a good means of exploiting such temporal data as provided by the water deficit indicator $Ks(t)$. The results show that the water deficit has an effect on grape quality. Their analysis confirmed already known facts about the vine physiological response according to the variety and the irrigation effect. Thus our results are comforting the validity of the Ks indicator, and therefore the level of confidence and reliability in the software sensor design procedure. Functional data analysis highlighted critical periods for vine and berry development, regarding final quality features.

On one hand, the knowledge-based extraction of meaningful summary features over phenological periods of interest allowed to feed these features as input to decision trees. This confirmed the primordial effect played by the variety on Berry Weight determination. On the other hand, functional data analysis made it possible to use the water stress curve ($Ks(t)$), as a whole, to explain Sugar Concentration. This will in the future allow more precise monitoring of vine water needs according to a targeted product.

Note that we did not take account of the variety factor in functional data analysis. This would require a covariance analysis model adapted to functional data, which was not possible in this study as the number of data per variety was not sufficient.

These results show the complementarity of both approaches: the first one performs dimensional reduction by summarizing features which requires expert assumptions, the second one handles the continuous temporal data, without any reduction, but it needs more numerous data to be efficient.

Applied perspectives of this work include the study of the relationship between vine water stress and other more complex quality features. In particular new chemical analyses make it possible to follow the aroma development in berries over time, which is assumed to be very sensitive to the vine water status.

Our approach is innovative in more than one aspect. Even if the software sensor had a different design, the same advanced methodology could still be applied to analyze the temporal data. Beyond the present case study, the proposed methodology has a high genericity level, for the applied fields of Agronomy and Environment. It could be used in many cases when raw data have to be transformed by software sensors to be meaningful, or when temporal data have to be analyzed in depth.

Acknowledgments

The research leading to these results has received funding from the Pilotype Program, funded by OSEO innovation and the Languedoc Roussillon regional council. The authors would like to thank all members of the project for their help and advices: Les Grands Chais de France, Alliance Minervois, Les Vignerons du Narbonnais, INRA Pech Rouge, INRA SPO, INRA MISTEA, IFV Rhône Méditerranée, SupAgro Montpellier (ITAP), Fruition Sciences, Nyseos. Finally, we wish to particularly thank Nicolas Saurin (INRA), Denis Caboulet, Jean-Christophe Payan and Elian Salançon (IFV) for providing the vine and wine-related data.

References

Allen, R. G., Pereira, L. S., Sep. 2009. Estimating crop coefficients from fraction of ground cover and height. Irrigation Science 28 (1), 17–34. URL `http://link.springer.com/10.1007/s00271-009-0182-z`

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. Irrigation and Drainage Paper No. 56. FAO, Rome, Italy.

Ben-David, A., Sterling, L., 2006. Generating rules from examples of human multiattribute decision making should be simple. Expert Syst. Appl. 31 (2), 390–396.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Wadsworth, Belmont, CA 1.

Cifre, J., Bota, J., Escalona, J., Medrano, H., Flexas, J., Apr. 2005. Physiological tools for irrigation scheduling in grapevine (Vitis vinifera L.). An open gate to improve water-use efficiency? Agriculture, Ecosystems & Environment 106 (2-3), 159–170. URL `http://linkinghub.elsevier.com/retrieve/pii/S0167880904002956`

des Gachons, C. P., Leeuwen, C. V., Tominaga, T., Soyer, J.-P., GaudillÃÍre, J.-P., Dubourdieu, D., 2005. Influence of water and nitrogen deficit on fruit ripening and aroma potential of vitis vinifera l cv sauvignon blanc in field conditions. Journal of the Science of Food and Agriculture 85 (1), 73–85.

Destercke, S., Buche, P., Charnomordic, B., jan. 2013. Evaluating data reliability: An evidential answer with application to a web-enabled data warehouse. Knowledge and Data Engineering, IEEE Transactions on 25 (1), 92 –105.

Escalona, J., Flexas, J., Medrano, H., 2002. Drought effects on water flow, photosynthesis and growth of potted grapevines. Vitis 41, 57–62.

Ferreira, M. I., Silvestre, J., Conceição, N., Malheiro, A. C., Jun. 2012. Crop and stress coefficients in rainfed and deficit irrigation vineyards using sap flow techniques. Irrigation Science 30 (5), 433–447.
URL http://www.springerlink.com/index/10.1007/s00271-012-0352-2

Ginestar, C., Eastham, J., Gray, S., Iland, P., 1998. Use of sap flow sensors to shedule vineyard irrigation. 1. effects of post-veraison water deficit on water relation, vine growth, and yield of shiraz grapevines. Am. J. Enol. Vitic. 49, 413–420.

Guarino, N., Oberle, D., Staab, S., 2009. What is an ontology? In: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. International Handbooks on Information Systems. Springer Berlin Heidelberg, pp. 1–17.

Hilgert, N., Mas, A., Verzelen, N., 2013. Minimax adaptive tests for the functional linear model. The Annals of Statistics 41 (2), 838–869.

James, G. M., Wang, J., Zhu, J., 2009. Functional linear regression that's interpretable. The Annals of Statistics 37, 2083–2108.

Jones, H. G., Nov. 2004. Irrigation scheduling: advantages and pitfalls of plant-based methods. Journal of experimental botany 55 (407), 2427–36.
URL http://www.ncbi.nlm.nih.gov/pubmed/15286143

Koundouras, S., Marinos, V., Gkoulioti, A., Kotseridis, Y., van Leeuwen, C., 2006. Influence of vineyard location and vine water status on fruit maturation of nonirrigated cv. agiorgitiko (vitis vinifera l.). Effects on wine phenolic and aroma components. Journal of Agricultural and Food Chemistry 54 (14), 5077–5086.

Kruchten, P., Nov. 1995. The 4+1 view model of architecture. IEEE Softw. 12 (6), 42–50.
URL `http://dx.doi.org/10.1109/52.469759`

Miller, G. A., 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63, 81–97.

Montoro, A., Fereres, E., Lopez-Urrea, R., Manas, F., Lopez-Fuster, P., 2011. Sensitivity of trunk diameter fluctuations in vitis vinifera l. tempranillo and cabernet sauvignon cultivars. American Journal of Enology and Viticulture 63(1), 85–93.

Musen, M., 1992. Dimensions of knowledge sharing and reuse. Computers and Biomedical Research 25 (5), 435–467.

Ojeda, H., Andary, C., Kraeva, E., Carbonneau, A., Deloire, A., 2002. Influence of pre-and postveraison water deficit on synthesis and concentration of skin phenolic compounds during berry growth of vitis vinifera cv. shiraz. American Journal of Enology and Viticulture 53 (4), 261–267.

Olivo, N., Girona, J., Marsal, J., 2009. Seasonal sensitivity of stem water potential to vapour pressure deficit in grapevine. Irrigation Science 27, 175–182.

Parker, A., De Cortázar-Atauri, I., Van Leeuwen, C., Chuine, I., 2011. General phenological model to characterize the timing of flowering and veraison of vitis vinifera l. Australian Journal of Grape and Wine Research 17, 206–216.

Picón-Toro, J., González-Dugo, V., Uriarte, D., Mancha, L. a., Testi, L., Jun. 2012. Effects of canopy size and water stress over the crop coefficient of a Tempranillo vineyard in south-western Spain. Irrigation Science 30 (5), 419–432.
URL `http://www.springerlink.com/index/10.1007/s00271-012-0351-3`

Quinlan, J., 1986. Induction of decision trees. Machine learning 1 (1), 81–106.

Quinlan, J., 1993. C4. 5: programs for machine learning. Morgan Kaufmann.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Ramsay, J. O., Silverman, B. W., 2005. Functional data analysis, 2nd Edition. Springer Series in Statistics. Springer, New York.

Raymond, K., 1995. Reference model of open distributed processing (rm-odp): Introduction. In: Proceedings of the IFIP TC6 International Conference on Open Distributed Processing. pp. 3–14.

Rijgersberg, H., van Assem, M., Top, J. L., 2013. Ontology of units of measure and related concepts. Semantic Web 4 (1), 3–13.

Rodrigues, P., Pedroso, V., Gouveia, J. P., Martins, S., Lopes, C., Alves, I., 2012. Influence of soil water content and atmospheric conditions on leaf water potential in cv. touriga nacional deep-rooted vineyards. Irrigation Science 30, 407–417.

Santesteban, L. G., Miranda, C., Royo, J. B., 2011. Suitability of pre-dawn and stem water potential as indicators of vineyard water status in cv. Tempranillo. Australian Journal of Grape and Wine Research 17 (1), 43–51.
URL `http://dx.doi.org/10.1111/j.1755-0238.2010.00116.x`

Thomopoulos, R., Destercke, S., Charnomordic, B., Iyan, J., Abécassis, J., Feb. 2013. An iterative approach to build relevant ontology-aware data-driven models. Information Sciences 221, 452–472.

Ullah, S., Finch, C. F., 2013. Applications of functional data analysis: A systematic review. BMC Medical Research Methodology 13 (1), 43.
URL `http://www.biomedcentral.com/1471-2288/13/43`

Van Leeuwen, C., Tregoat, O., Choné, X., Bois, B., Pernet, D., Gaudillère, J.-P., et al., 2009. Vine water status is a key factor in grape ripening and vintage quality for red bordeaux wine. How can it be assessed for vineyard management purposes? J. Int. Sci. Vigne Vin 43 (3), 121–134.

Villanueva-Rosales, N., Dumontier, M., 2008. Modeling life science knowledge with owl 1.1. In: Proceedings of OWL'08.

Williams, L., Baeza, P., 2007. Relationships among ambient temperature and vapor pressure deficit and leaf and stem water potentials of fully irrigated field-grown grapevines. Am. J. Enol. Vitic. 58, 2.

Zhang, Y., Kang, S., Ward, E. J., Ding, R., Zhang, X., Zheng, R., 2011. Evapotranspiration components determined by sap flow and microlysimetry techniques of a vineyard in northwest China: Dynamics and influential factors. Agricultural Water Management.
URL `http://dx.doi.org/10.1016/j.agwat.2011.03.006`