



**HAL**  
open science

## **A modeling approach to design a software sensor and analyze agronomical features - Application to sap flow and grape quality relationship**

Aurélie Thébaut, Thibault Scholash, Brigitte Charnomordic, Nadine Hilgert

### ► **To cite this version:**

Aurélie Thébaut, Thibault Scholash, Brigitte Charnomordic, Nadine Hilgert. A modeling approach to design a software sensor and analyze agronomical features - Application to sap flow and grape quality relationship. 2013. hal-00863992v1

**HAL Id: hal-00863992**

**<https://inria.hal.science/hal-00863992v1>**

Preprint submitted on 20 Sep 2013 (v1), last revised 6 Jan 2014 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A modeling approach to design a software sensor and analyze agronomical features - Application to sap flow and grape quality relationship

Aurélie Thébault<sup>a</sup>, Thibaut Scholasch<sup>b</sup>, Brigitte Charnomordic<sup>a</sup>, Nadine Hilgert<sup>a</sup>

<sup>a</sup>*INRA-SupAgro, UMR 729 MISTEA, F-34060 Montpellier, France*

<sup>b</sup>*Fruition Sciences, Montpellier, France*

---

## Abstract

This work proposes a framework using temporal data and domain knowledge in order to analyze complex agronomical features. 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 requires the use of advanced data analysis methods, such as functional regression. The methodology is applied to a case study involving an experimental design in French vineyards. The temporal data consist of sap flow measurements, and the goal is to explain fruit quality (sugar concentration and weight), using vine's water courses through the various vine phenological stages. The results are discussed, as well as the method genericity and robustness.

*Keywords:* vine water stress, functional data analysis, ontology, expert

---

*Email address:* [bch@supagro.inra.fr](mailto:bch@supagro.inra.fr) (Brigitte Charnomordic)

*Preprint submitted to Computer and Electronics in Agriculture Friday 20<sup>th</sup> September, 2013*

1 **1. Introduction**

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

12 Our objective in this paper is to show the interest of a formalized data  
13 and knowledge-based approach to study a complex agronomical phenomenon,  
14 namely the influence of vine water deficit on grape quality.

15 Any index of vine water deficit aims at evaluating the amount of water  
16 effectively used by the vine in order to determine if this amount falls short  
17 of some reference amount. Typically, the *reference* amount is the maximal  
18 amount of water a vine can use.

19 Various methods exist to characterize the level of water deficit experi-  
20 enced by the plant as reviewed by Jones (2004). Tissue water status can be  
21 assessed visually or by measurements of vine water potential. However, both  
22 methods have serious drawbacks. The lack of precision of visual observations  
23 often leads to yield reduction before visible symptoms occurs. The pressure

24 chamber method used to measure water potential is slow and labour inten-  
25 sive, especially for predawn measurement, and is unsuitable for automation.  
26 In addition, measurements done with pressure chambers are very dependent  
27 on atmospheric conditions and vine phenological stage Olivo et al. (2009);  
28 Williams and Baeza (2007); Rodrigues et al. (2012); Santesteban et al. (2009).

29 Sap flow sensors, that indirectly measure changes in conductance, have  
30 recently become available. This sensitive measurement method requires a  
31 complex instrumentation and technical expertise for the definition of irriga-  
32 tion control thresholds Ginestar et al. (1998). The main advantages of sap  
33 flow method is to allow automatic and continuous measurement of water  
34 flowing through the plant, which is directly related to transpiration Escalona  
35 et al. (2002); Jones (2004); Cifre et al. (2005); Zhang et al. (2011).

36 However, sap flow is a complex phenomenon and expert knowledge is nec-  
37 essary to convert raw data into useful transformed data, i.e. water courses,  
38 by designing a software sensor. Once these data transformations are vali-  
39 dated, it opens the way to a range of new studies, based on vine's water  
40 courses.

41 In this paper, we will first show how a formalized data and knowledge-  
42 based approach can be useful to design a software sensor. Knowledge formal-  
43 ization will be done by using ontologies, which take increasing importance  
44 in the field of Life Sciences Villanueva-Rosales and Dumontier (2008); Tho-  
45 mopoulos et al. (2013), for their ability to model and structure qualitative  
46 domain knowledge.

47 In a second step, water use trajectories will be put in relation to grape  
48 quality indicators such as Berry Weight or Sugar Concentration, using re-

49 cent data analysis tools and formalized knowledge. Innovative data analysis  
50 tools include functional data analysis that offers the possibility to use curve  
51 (functional) data instead of scalar data. Functional data analysis has not  
52 been much used in life sciences yet Ullah and Finch (2013), though it could  
53 be of particular interest in the Vine and Wine Industry, and more generally  
54 for modern Agronomy.

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

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

60 The paper is organized as follows: Section 2 presents the material and  
61 methods. It is divided into four parts. The first part gives some elements  
62 about data and the second one presents ontology-based formalization. The  
63 software sensor design, that relies on the use of mathematical models, data  
64 and formalized knowledge, is described in the third part. The selected exam-  
65 ple shows how it is possible to transform raw sap flow data into vine water  
66 deficit courses. The fourth part describes the methods used for analysing the  
67 software sensor output in relation to product quality. Section 3 presents and  
68 discusses the results and their relationship with grape composition (Sugar  
69 Concentration, Berry Weight). Some concluding remarks are given in Sec-  
70 tion 4.

## 71 **2. Material and methods**

72 In this section, we propose to follow four steps:

- 73 • to describe the experimental design with its input and output variables
- 74 • to formalize eco-physiological knowledge using an ontology
- 75 • to design a software sensor using formalised knowledge, a mathematical
- 76 model, and data
- 77 • to relate software sensor output to product quality using decision trees
- 78 and functional analysis.

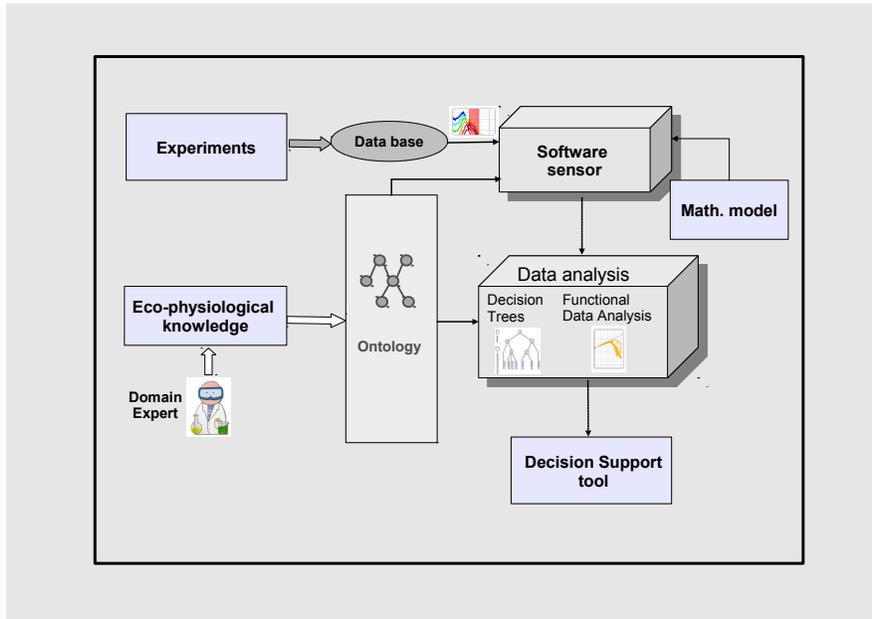


Figure 1: Outline of the proposed modeling approach.

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

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

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

### 100 *2.1. Experimental design*

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

107 Grenache or Chardonnay.

108 To get a wider range of vine water statuses during the season, an irrigation  
109 treatment was applied for two years on each of the eight site-variety com-  
110 binations. The irrigation treatment consisted of two modalities, replicated  
111 twice, yielding 32 experimental subplots. In the non irrigated subplots, vines  
112 only received natural precipitations during the growing season while in the  
113 irrigated subplots, vines received regular extra-amounts of water through  
114 drippers line (emission rate from 2 to 4 l h<sup>-1</sup>, 1 to 2 drippers per plant).

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

#### 118 *2.1.1. Meteorological data*

119 Hourly meteorological data on windspeed (km h<sup>-1</sup>), minimal, maximal  
120 and mean air temperature (°C), air humidity (%), solar radiation (W m<sup>-2</sup>)  
121 and amounts of precipitations (mm) were extracted from local meteorological  
122 stations for each site.

#### 123 *Transformed data*

124 Hourly vapor pressure deficit (*VPD*) and reference atmospheric evapora-  
125 tive demand (potential evapotranspiration,  $ET_{ref}$ ) were calculated according  
126 to methodologies referred to as FAO-56 Allen et al. (1998). Calculation of  
127 reference atmospheric evaporative demand ( $ET_{ref}$  in mm d<sup>-1</sup>) is based on  
128 Penman-Monteith formula.

129 Daily meteorological data were obtained from hourly data after a trapeze  
130 integration. Thermal time, *i.e.* the accumulation of growing degree days

131 (GDD) from April 1<sup>st</sup> was calculated by daily integration of mean air temper-  
132 ature minus a base temperature of 10 °C which is considered as the simplest  
133 model to estimate vine phenology Parker et al. (2011).

#### 134 *2.1.2. Phenological data*

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

#### 141 *2.1.3. Vine water status data*

142 Vine water status was monitored by two kinds of measurements: discrete  
143 measurements of leaf water potential at predawn ( $\Psi_b$ , or predawn LWP) and  
144 continuous measurements of sap flow.

##### 145 *Leaf water potential at predawn*

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

##### 149 *Sap flow*

150 The energy balance method (Sakuratani, 1981) was used to measure sap  
151 flow with Sap IP system (Dynamax, Houston, TX, USA). There is one variety  
152 per vineyard site. The vineyard site is divided into 2 irrigation treatments.  
153 Two vineyard rows were selected. One row represents one irrigation treat-  
154 ment. In each selected row, 2 vines were equipped with one sensor. Each

155 sensor measured vine sap flow rate every 15 minutes. The 2 selected vines  
156 were within 25 meters of each other within the same row.

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

#### 168 *2.1.4. Fruit composition quality data*

169 Starting two weeks before harvest, fruit was sampled for each irrigation  
170 treatment in each vineyard. Fruit data was collected at three different dates.  
171 Fruit composition analysis focused on berry weight (g), sugar concentration  
172 ( $\text{g l}^{-1}$ ), acidity ( $\text{gH}_2\text{SO}_4 \text{l}^{-1}$ ), anthocyanes and assimilable nitrogen ( $\text{mg l}^{-1}$ ).

#### 173 *2.2. Formalizing knowledge*

174 In this section, our aim is to show how ontologies can be used to formalize  
175 domain knowledge and to design a software sensor.

176 In information science, an ontology formally represents knowledge as a set  
177 of concepts within a domain, and the relationships between pairs of concepts.

178 Ontologies are becoming increasingly popular, due to the great amount

179 of available (complex) data and to the need for modeling (qualitative) know-  
180 ledge and structural information. This need first arose out of the develop-  
181 ment of the World Wide Web. However, there are still very few attempts  
182 to combine ontologies and statistical or data-driven models. This could be  
183 particularly useful in Life Sciences and Agronomy Villanueva-Rosales and  
184 Dumontier (2008); Thomopoulos et al. (2013); Destercke et al. (2013).

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

- 187 1. To share a common understanding of structured information Musen (1992);
- 188 2. To explicit the specificities of domain knowledge;
- 189 3. To identify ambiguous or inappropriate model choices.

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

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

198 Let us comment the main concepts and relations.

### 199 2.2.1. Concepts

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

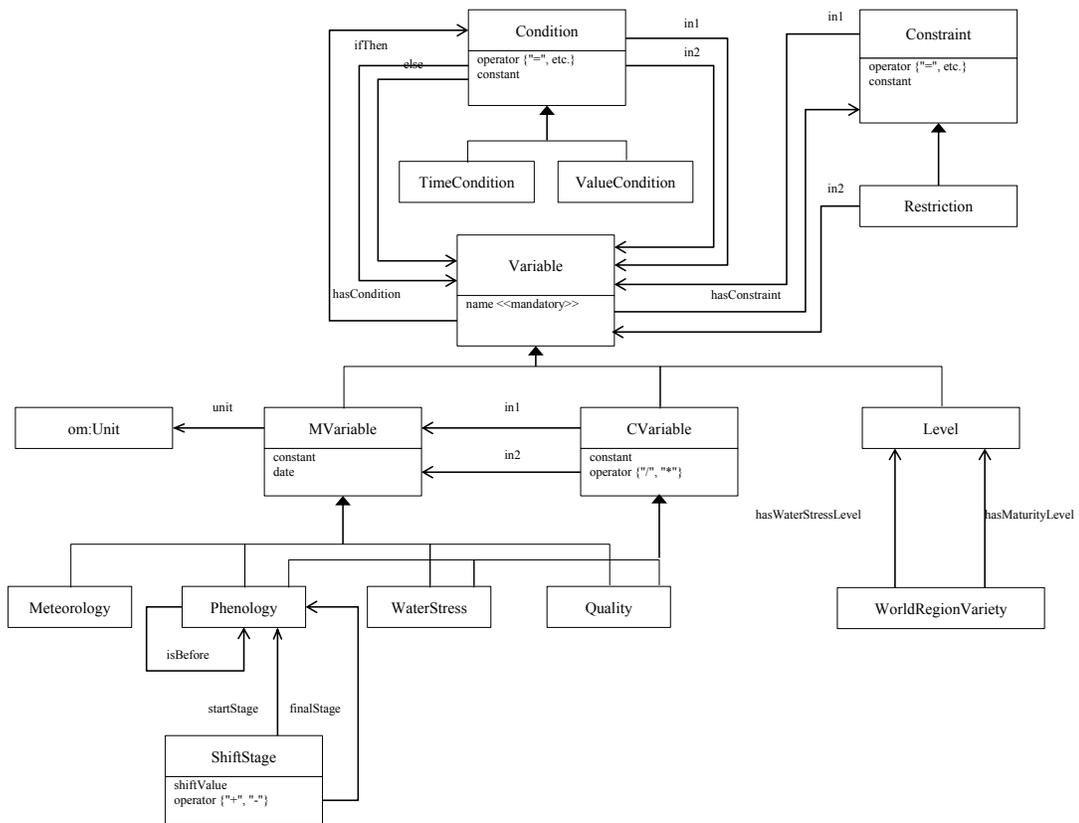


Figure 2: Class diagram of the ontology of Vine Water Stress

203 plained in Section 2.2.2. For instance, in Figure 2, *Meteorology*, *Phenology*,  
204 *WaterStress* and *Quality* are subconcepts of *Variable*.

- 205 • All *variables* must have a name, they can have a date, a unit and a  
206 default value. The units are taken from *OM*, an ontology of units of  
207 measures and related concepts Rijgersberg et al. (2013). .
- 208 • The *Condition* concept is defined with a comparison operator and two  
209 operands. It will be used together with the *hasCondition* ( $\mathcal{HCO}_c$ ) re-  
210 lation, defined in Section 2.2.2.
- 211 • The *Constraint* concept is defined with a comparison operator and one  
212 operand. It will be used together with the *hasConstraint* ( $\mathcal{HCS}_c$ ) rela-  
213 tion, defined in Section 2.2.2. The *Restriction* concept is a subconcept  
214 of *Constraint*, and is a specific two-fold constraint.
- 215 • The *ShiftStage* concept is proposed in order to determine a phenological  
216 stage from another one. This is the case for the *Nouaison* stage, which  
217 is not generally observed. Its date can be estimated by shifting the  
218 *Bloom* date by  $k$  *GDD*, where  $k$  can be variety-dependent. *Nouaison*  
219 and *Bloom* are instances of the *Phenology* concept.

### 220 2.2.2. Relations

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

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

$$\mathcal{C}_c = \{c' \in \mathcal{C} | c' \preceq c\}. \quad (1)$$

228 For example, in Figure 2, let us consider the concept  $c = Variable$ . We  
229 have  $\mathcal{C}_{Variable} = \{MVariable, CVariable, Level\}$ , where *MVariable* rep-  
230 represents a measurement available in a data base, *CVariable* a variable  
231 calculated following a given method, and *Level* a constant value de-  
232 pending on some other concepts.

- 233 • The *subsumption* relation can be multiple. For instance a *Phenological*  
234 *concept* can be such as  $c \preceq CVariable$  or  $c \preceq MVariable$ .
- 235 • The *isBefore* relation allows to represent temporal precedence. It is  
236 very important for checking the consistency of the phenological stage  
237 dates, where *bloom* has to occur before *veraison*, and so on.
- 238 • The HasCondition ( $\mathcal{HCO}_c$ ) relation, where  $c$  represents the concept on  
239 which the condition is to be applied, is used together with a condition.
- 240 • Similarly, the HasConstraint ( $\mathcal{HCS}_c$ ) relation allows the application of  
241 a *constraint* on the  $c$  concept.

242 Examples of use will be given in Section 2.3.2.

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

245 on the World Wide Web, which is specified using W3C<sup>1</sup> recommendations.

246 The use of OWL allows reusing ontologies developed elsewhere, such as *OM*.

247 Note that each modeling component: data base, ontology and mathe-  
248 matical modelling or data analysis, can be modelled independently, using its  
249 favorite language (sql, R, OWL). The communication between the various  
250 components can be implemented by a high level interface, written in Python,  
251 PHP or Java.

### 252 *2.3. Design of the software sensor for vine water stress estimation*

253 Based on the knowledge formalized in the ontology given in Fig.2 and on  
254 a mathematical model Ferreira et al. (2012), the software sensor is designed  
255 to transform the calibrated transpiration measurements from sap flow sensors  
256 into a vine water deficit estimator, denoted by  $Ks(t)$ .

257 The different steps of the methodology used to design  $Ks$  are summarized  
258 in Fig.3 and detailed in the following subsections.

#### 259 *2.3.1. Sap flow under limiting soil water condition : computation of $Ks$*

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

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

262 It accounts for the decline in vine water use due to soil moisture deficit. Even  
263 if some authors Allen et al. (1998) present a general proposal for estimat-  
264 ing  $Ks$ , specific  $Ks$  functions for vineyards, from field experiments, are not  
265 generally available.

---

<sup>1</sup><http://www.w3.org/TR/>

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

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

270  $ET_{ref}$  is the reference evapotranspiration and  $K_{cB}$  a coefficient linearly  
 271 related to the leaf area index (LAI) or to the fraction of ground coverage  
 272 Williams and Ayars (2005); Picón-Toro et al. (2012). As  $K_{cB}$  is dependent  
 273 upon crop type and management practices, which will influence the rate of  
 274 canopy development and the ultimate canopy size, *i.e.* amount of ground  
 275 cover (Allen et al. (1998)), a site-specific determination of  $K_{cB}$  is necessary  
 276 for each vineyard.

277

278 *2.3.2. Sap flow under non limiting soil water condition : computation of dry*  
 279 *soil  $K_{cB}$*

280 No direct measurements are available to determine  $K_{cB}(t)$  (see Eq.3),  
 281 which is vine specific and varies with leaf area. We propose to use formalized  
 282 concepts and relations based on expertise, all of them implemented in the  
 283 OVWS ontology.  $K_{cB}$  profile is divided into two growth stages :  $L_{dev}$  and  
 284  $L_{mid}$  as presented in Fig.4. To determine  $K_{cB}(t)$ , two hypotheses on the  
 285 curve shape are assumed:

$$K_{cB}(t) = f(t) \text{ for } t < t_{K^*} \quad (4)$$

$$K_{cB}(t) = K^* \text{ for } t \geq t_{K^*} \quad (5)$$

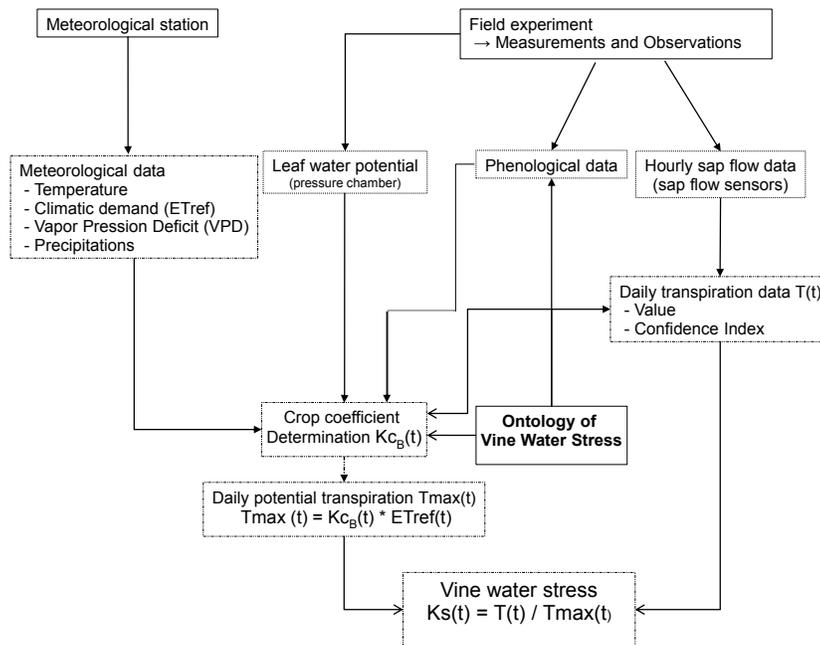


Figure 3: Methodological steps for the vine water deficit ( $K_s$ ) estimation.

286 where  $f(t)$  is assumed to be linear in  $t$ , and  $t_{K^*}$  is the breakpoint for which  
 287  $Kc_B$  reaches the plateau  $K^*$ . The key point is to set  $t_{K^*}$ , or indifferently  $K^*$ .

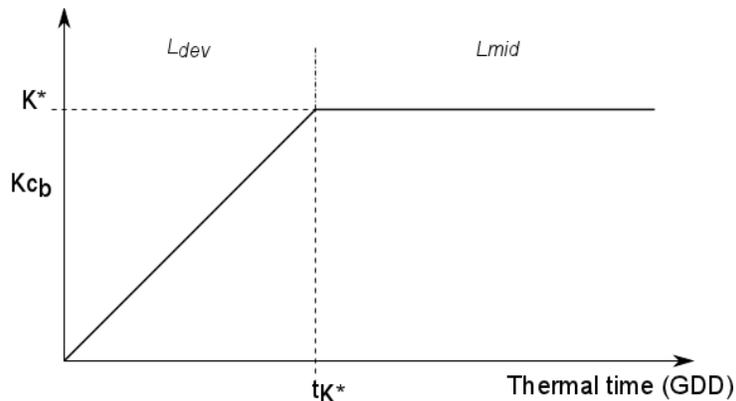


Figure 4: Theoretical curve of  $Kc_B$  evolution during the season.

287

288 According to Eq.3, we make the hypothesis that, in the absence of water  
 289 deficit,  $K^*$  is defined as:

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

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

295 1. *Selection based on phenology*

296 A linear relationship exists between  $Kc_B$  variations and leaf area index  
 297 (LAI) or the fraction of ground covered by the vine Ferreira et al.  
 298 (2012). We thus assume that peak  $Kc_B$  (i.e.  $K^*$ ) is reached when

299 LAI stops increasing. Consequently, the search period for  $K^*$  has been  
300 limited to the period between budbreak and veraison.

301 These two concepts are defined as subconcepts of the Phenology con-  
302 cept, itself being a subconcept of *MVariable*. The period limitation is  
303 intanciated by two *TimeConditions*, applied onto the  $Kc_B \preceq CVariable$   
304 concept, the  $\mathcal{HCO}_c$  relation where the *Condition* is characterized by a  
305 comparison operator  $\leq$  (resp.  $\geq$ ) and the *Veraison* (resp. *Budbreak*)  
306 concept.

### 307 2. *Selection based on predawn leaf water potential*

308 Conditions of maximal soil moisture availability could be inferred from  
309 predawn leaf water potential measurements, associated with a confi-  
310 dence interval derived from VPD. A rule was set so that  $K^*$  has to  
311 be reached before the first day at which predawn LWP measurement  
312 reveals a water stress. The levels to which predawn LWP characterises  
313 water stress can be defined by the stakeholder, or else set in agreement  
314 with a standard level, based on a region or/and variety.

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

### 317 3. *Selection based on meteorology*

318 Transpiration measurement through sap flow is sensitive to climatic  
319 conditions, mainly light and VPD. To account for sensitivity of tran-  
320 spiration measurements to  $VPD(t)$ , a filtering rule was set to remove  
321 computed  $Kc_B(t)$  obtained in situations of heat spikes, defined as pe-  
322 riod with VPD greater than a given level, set to 3.5 kPa in the present  
323 case study.

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

#### 328 4. Selection based on curve shape

329 By definition,  $K^*$  is reached when the ratio  $\frac{T(t)}{ET_{ref}(t)}$  reaches a maxi-  
 330 mum during a few days (as  $T(t)=T_{max}(t)$  and  $Kc_B(t)=K^*$ ) and then  
 331 decreases (as  $T(t)<T_{max}(t)$  due to limiting soil water conditions while  
 332  $Kc_B(t)=K^*$ ). As such, potential options for  $t_{K^*}$  have been defined at  
 333 points with a null first derivative and a negative second derivative.

334 This selection is implemented using two concepts:  $\dot{K}^*$  and  $\ddot{K}^*$ , both  
 335 such as  $\preceq WaterStress \preceq CVariable$ , and a  $\mathcal{HCS}_c$  relation with a  
 336 *Constraint* characterized by a comparison operator  $\leq \epsilon$  or  $\leq 0$ .

#### 337 User selection based on expert knowledge.

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

341 The final choice is left to the stakeholder who is the best aware of the  
 342 management practices or particular uncontrolled events that could have in-  
 343 terfered with vine growth (irrigation, leaf removal, treillis system, ...) and  
 344 therefore  $Kc_B$  curve.

#### 345 2.4. Relating software sensor output to product quality

346 The software sensor output consists of temporal data, and various meth-  
 347 ods can be used to study the relationships between these data and product

348 quality. Two complementary lines of work, based on statistical methods, are  
349 explored in the present work. The first one consists of extracting significant  
350 scalar parameters from the temporal data and using them as input to deci-  
351 sion trees, in order to provide the most discriminant features. The second  
352 one uses functional data analysis, that gives the possibility to model the tem-  
353 poral data impact on product quality as a whole. However, curve analysis is  
354 a recent research topic, with relatively few methods available, in comparison  
355 with classical data analyses.

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

#### 361 *2.4.1. Extracting scalar parameters from software sensor output*

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

367 Three periods were first defined according to phenological stages: the  
368 whole season, the pre-veraison period, which goes from the nouaison stage  
369 to the veraison stage, and the post-veraison period which ranges from the  
370 veraison stage to the harvest date. In a second step, the post-veraison period  
371 was divided by taking into account the maturity stage of berries (see Section  
372 2.1.2), which allowed to add a fourth period ranging from veraison to matu-

373 rity. Maturity stage is reached when the ratio between Sugar Concentration  
374 and acidity in grapes yields a given threshold, defined according to variety

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

#### 382 2.4.2. Decision trees as interpretable models

383 Decision tree algorithms are well established learning methods in super-  
384 vised data mining and statistical multivariate analysis. They allow to display  
385 non linear relationships between features and their impact on a response vari-  
386 able, in a compact way. A low complexity Ben-David and Sterling (2006) is  
387 essential for the model to be interpretable, as confirmed by Miller’s conclu-  
388 sions Miller (1956) relative to the *magical number seven*.

389 Decision trees can handle classification problems or regression cases, de-  
390 pending on the nature of the response variable. We present here the regres-  
391 sion case, where the response variable is numerical.

392 Input to regression decision trees consists of a collection of  $N$  training  
393 cases, each having a tuple of values for a set of  $P$  input variables, and one  
394 output variable  $(\mathbf{x}_i, \mathbf{y}_i) = (x_{1,i}, x_{2,i} \dots x_{P,i}, y_i)$ . An input  $X_p$  ( $p = 1 \dots P$ ) is  
395 continuous or discrete and takes its values  $(x_{p,i})_{i=1 \dots N}$  on a domain  $\mathcal{X}_p$ . The  
396 goal is to learn from the training cases a recursive structure (taking the shape  
397 of a rooted tree) consisting of (i) leaf nodes labeled with a mean value and

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*.

398 a standard deviation, and (ii) test nodes (each one associated to a given  
399 variable) that can have two or more outcomes, each of these linked to a  
400 subtree.

401 Decision trees are easily interpretable for a non-expert in statistical or  
402 learning methods, and facilitate exchanges with the domain expert.

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

407 Note that the CART family Breiman et al. (1984), based on binary splits.  
408 is mostly used by statisticians. There is another tree family Quinlan (1986),  
409 called ID3 Quinlan (1993), allowing non binary splits and mostly used by  
410 artificial intelligence researchers.

411 In this work, we used CART-based trees. In that case, the splitting  
412 criterion is based on finding the one predictor variable (and a given threshold  
413 of that variable) that results in the greatest change in explained deviance  
414 (for Gaussian error, this is equivalent to maximizing the between-group sum  
415 of squares, as in an ANOVA). This is done using an exhaustive search of  
416 all possible threshold values for each predictor. The implementation used for  
417 decision trees is the R Development Core Team (2009) software with the *rpart*  
418 package<sup>2</sup>. Specifying variety, *NouVer*, *VerHarv* and *VerMat* as explanatory  
419 variables, we computed decision trees on maximum values of grape quality  
420 features over the season.

---

<sup>2</sup><http://cran.r-project.org/web/packages/rpart/index.html>

421 *2.4.3. Functional data analysis*

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

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

425 where  $\varepsilon_i$  is a random error,  $\beta_0$  is the intercept of the model and  $\beta(t)$  is the  
426 coefficient function, both unknown and to be estimated from independent  
427 observations  $(X_i(t), Y_i)_{i=1, \dots, n}$ . In this model,  $\beta(t)$  determines the effect of  
428  $X_i(t)$  on  $Y_i$ . For example,  $X_i(t)$  has a greater effect on  $Y_i$  over regions of  $t$   
429 where  $|\beta(t)|$  is large. On the opposite,  $X_i(t)$  has no effect on  $Y_i$  over regions  
430 of  $t$  where  $\beta(t)$  is zero. Estimating  $\beta(t)$  in model (7) has given rise to an  
431 increasing litterature during this last decade, see for example Ramsay and  
432 Silverman (2005). A common approach involves projecting  $\beta$  and the  $X_i$ 's  
433 in a  $p$ -dimensional basis function where  $p$  is large enough to capture the  
434 unknown variations of  $\beta$ , but small enough to regularize the fit.

435 Interpretation of such estimators is not that easy. Recently, James et al.  
436 (2009) introduced new estimators that are both interpretable, flexible and  
437 accurate. The method, called "Functional Linear Regression That's Inter-  
438 pretable" (FLRTI), is based on a particular basis function and model selec-  
439 tion principles. The  $\beta$  function is either assumed to be zero or to have a  
440 simple linear form. The reason behind the first assumption is that the  $X_i(t)$   
441 observations are not of equal importance to explain the response  $Y_i$ . As said  
442 above,  $X_i(t)$  has no effect on  $Y_i$  when  $\beta(t) = 0$ . The second assumption is  
443 made for technical reasons to simplify the model. All together, this model  
444 will have a high explanatory capability, contrary to a pure predictive model.

445 These assumptions will constraint the estimation of the regression model (7),  
446 which corresponds to a penalized regression in sparse models. Two tuning  
447 parameters have to be fixed, a penalty term  $\sigma$  and a weight  $\omega$ . The penalty  
448 term is part of the model selection process (by lasso or Dantzig selector).  
449 The largest the  $\sigma$ , the more the form-related constraint is enforced. The  
450 weight  $\omega$  impacts the relative number of zeros of the  $\beta$  function. A weight of  
451 0 indicates that only the linear form constraint is respected.

452 The FLRTI method is implemented in an R function available on J.  
453 Gareth's web page<sup>3</sup>. A cross-validation algorithm is also proposed to opti-  
454 mize the choice of  $\sigma$  and  $\omega$ .

455 Using the FLRTI method, we analysed the effects of water stress curve  
456 over the season ( $Ks(t)$ ) on Berry Weight and Berry Sugar Concentration at  
457 harvest.

### 458 3. Results

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

#### 464 3.1. Vine water stress course $Ks(t)$ estimation

465 Sap flow data require a pre-treatment, including sensor selection and  
466 signal smoothing. Sap flow sensors have only been used recently in European

---

<sup>3</sup><http://www-bcf.usc.edu/gareth/research/flrti>

467 vineyards. Thus, calibration protocols are not established yet and therefore  
468 sensors can still be unreliable. Consequently, a selection step is required.

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

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

### 481 *3.1.1. $K^*$ determination*

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

488     Fig.5 illustrates the results for the Grenache variety at the Piolenc site.

489     The validity of the  $K^*$  determination procedure can be assessed according  
490 to different points. First, the results regarding  $K^*$  determination based on the  
491 coupling of mathematical algorithms and expert knowledge were consistent

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

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.

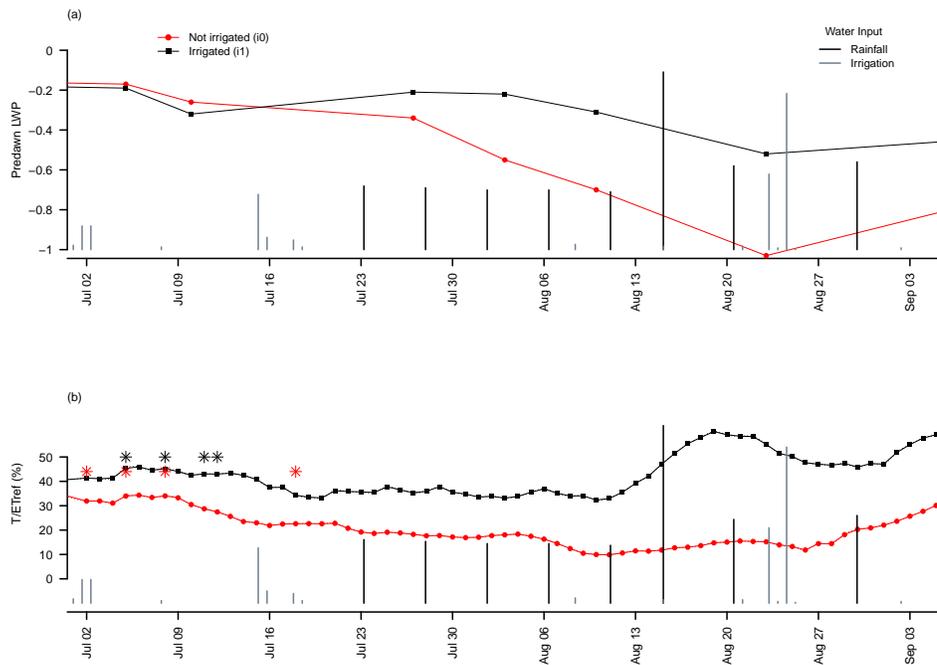


Figure 5: (a) Vine water status as indicated by leaf water potential and (b) determination of  $K^*$  and  $t_{K^*}$  (stars) based on expert knowledge following the mathematical algorithm suggestions for non irrigated (red bullets) and irrigated (black squares) treatments on the Grenache variety at Piolenc site.

492 with existing literature. Indeed, most of  $t_{K^*}$  occurred between 600 and 700  
493 GDD after budbreak (Table 2), which is in accordance with  $t_{K^*}$  reported in  
494 Picón-Toro et al. (2012) from a 3 year study in western Spain on Tempranillo,  
495 and in FAO-56 Allen et al. (1998), that respectively reported  $t_{K^*}$  around 650  
496 GDD and 555-592 GDD after budbreak.

### 497 3.1.2. Maximal transpiration and $Ks$ estimation

498 Following determination of  $K^*$  and  $t_{K^*}$ ,  $Kc_B(t)$  was calculated over all  $t$   
499 values (Eq.4). Its evolution for a Grenache variety is plotted on Fig.6, both  
500 in calendar time (a) and thermal time since budbreak (b).

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

505 Figure 7 shows vine water status according to both indicators: (a) Predawn  
506 LWP and (b)  $Ks$  during the season 2012 in a Grenache variety of the Languedoc-  
507 Roussillon region.

### 508 3.2. Relationships between vine water stress $Ks(t)$ and grape quality

509 As explained in Section 2.4,  $Ks(t)$  can be used in two different ways,  
510 either summarized as a series of scalar values, or as a whole. The way to  
511 summarize  $Ks(t)$  is detailed in Section 2.4.1. Scalar values and  $Ks(t)$  will be  
512 put in relation to grape quality, by the respective use of (i) regression trees  
513 and (ii) functional data analysis. The studied grape quality features include  
514 *i)* Berry Weight and *ii)* Sugar Concentration in berries.

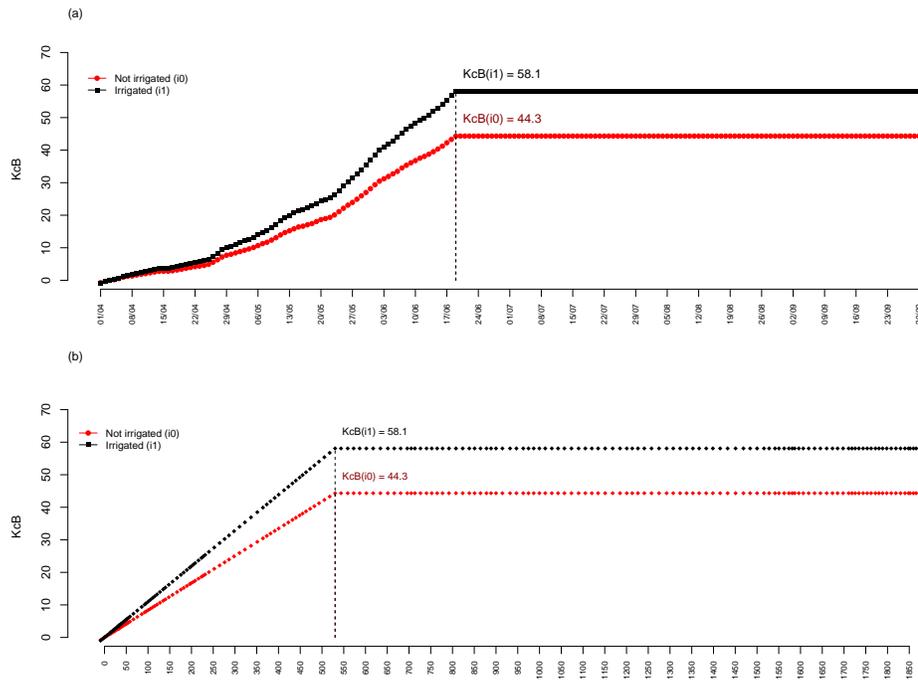


Figure 6: Evolution of vine basal crop coefficient ( $K_{cB}$ ) during the season at Piolenc site with Grenache variety, from budbreak to harvest. (a) x-scale in Julian days - (b) x-scale in GDD since budbreak.

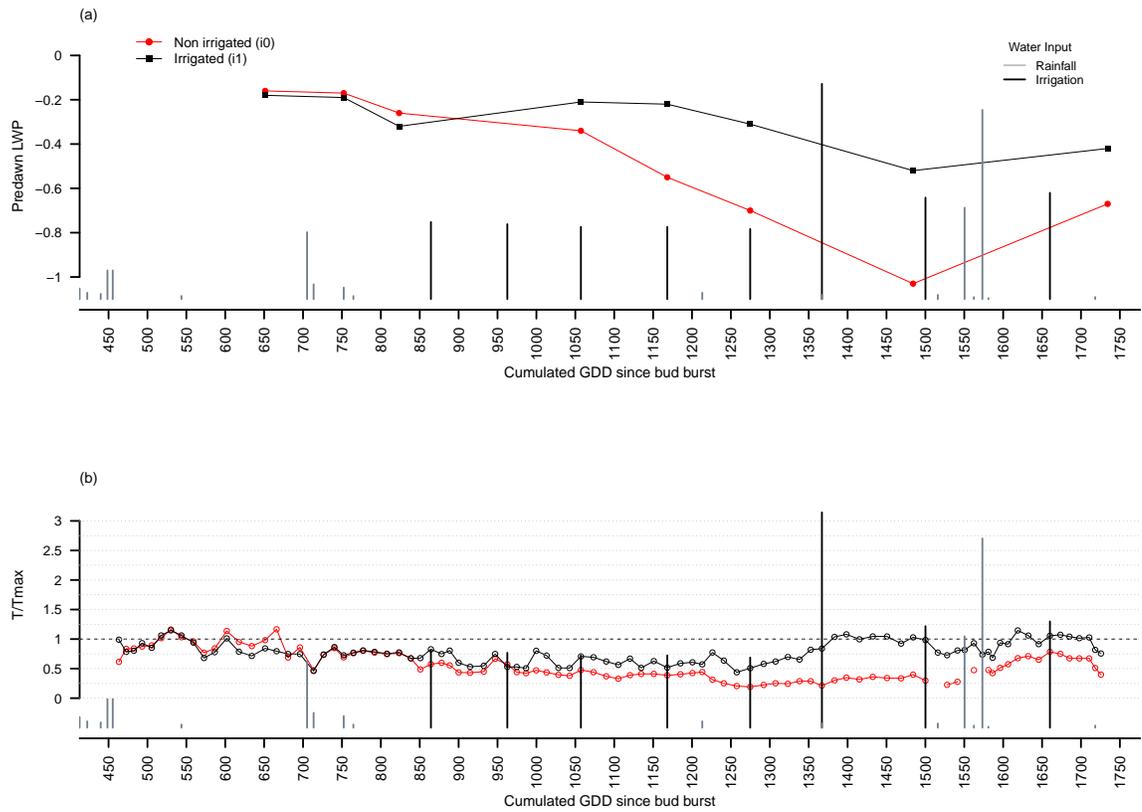


Figure 7: Water stress during 2012 millesim in Grenache variety at Piolenc site assessed by (a) Predawn LWP and (b) vine water stress indicator  $K_s$ .

515 *3.2.1. Regression trees*

516 Aggregated variables over periods can be used as explanatory variables  
517 in regression trees to detect and prioritize the periods critical to changes  
518 in grape quality. We studied the effects of *NouVer*, *VerHarv*, *VerMat* and  
519 variety on the two components of grape quality cited above.

520 According to Fig.8, Berry Weight seems to be mostly affected by the  
521 variety (Fig.8). Grenache variety significantly yields heavier berries. The  
522 second split for all varieties is done on the post-veraison water stress only  
523 (either *VerHarv* or *VerMat*). The higher the post-veraison water stress, the  
524 smaller the Berry Weight.

525 Regarding Sugar Concentration, regression trees show that it is affected  
526 by water stress in both pre-veraison *NouVer* and post-veraison *VerHarv* pe-  
527 riods (Fig.9). The first discriminant variable on Sugar Concentration is the  
528 post-veraison water stress (*VerHarv*, Fig.9). The first split shows that a  
529 higher post-veraison water stress leads to a lower Sugar Concentration.

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

535 *3.2.2. Functional data analysis*

536 Using a continuous indicator of water deficit enables the use of the whole  
537 season water stress curve to explain berry composition. This in turn will  
538 promote a more precise monitoring of vine water needs according to the  
539 targeted fruit composition. Using the FLRTI method James et al. (2009),

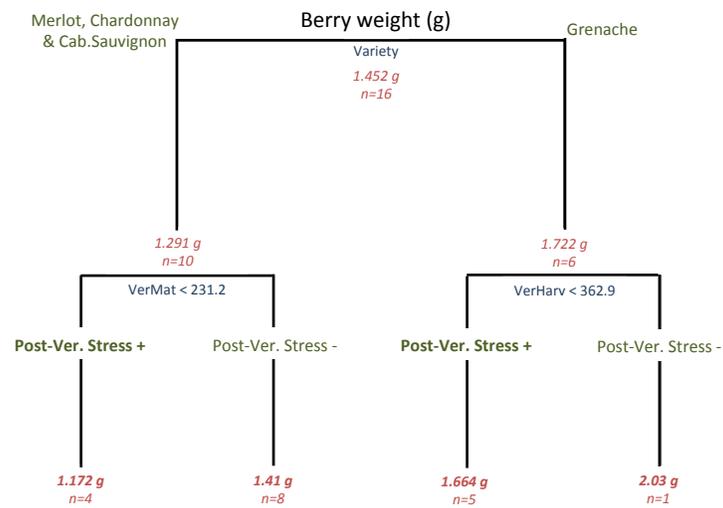


Figure 8: Regression tree explaining Berry Weight (g) using scalars summarizing the three periods, i.e. pre-veraison (*NouVer*), and post-veraison either until maturity (*VerMat*) or harvest (*VerHarv*).



540 we analyzed the effects of water stress curve over the season on Berry Weight  
541 and Sugar Concentration in berry at harvest.

542 Regarding Berry Weight, the results showed no significant effect of  $Ks(t)$ .  
543 This was confirmed by applying a specific testing procedure designed for  
544 functional linear models, see Hilgert et al. (2013).

545 Results of the functional data analysis on the effect of  $Ks(t)$  on Sugar  
546 Concentration are displayed on Fig.10. The tuning parameters are indicated  
547 in the legend. Parameters were obtained following a ten-fold cross validation.  
548 A sensitivity analysis to small  $\sigma$  and  $\omega$  variations showed a good robustness  
549 of the model, with three main peaks always located in the same time periods  
550 across the different varieties. These three main peaks are labeled (1), (2)  
551 and (3). Each of them corresponds to a significant effect of  $Ks$  on Sugar  
552 Concentration which can be positive (peaks (1) and (3)) or negative (peak  
553 (2)).

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

559 Regarding the time period, peak (1) appears to be located before pre-  
560 veraison whereas peak (3) occurred during pre-veraison. By contrast, peak  
561 (2) has a negative effect on Sugar Concentration. During this period located  
562 just before pre-veraison, a low water stress decreases the Sugar Concentra-  
563 tion. This can be reformulated as follows: the higher the stress during that  
564 period, the lower the decrease in Sugar Concentration.

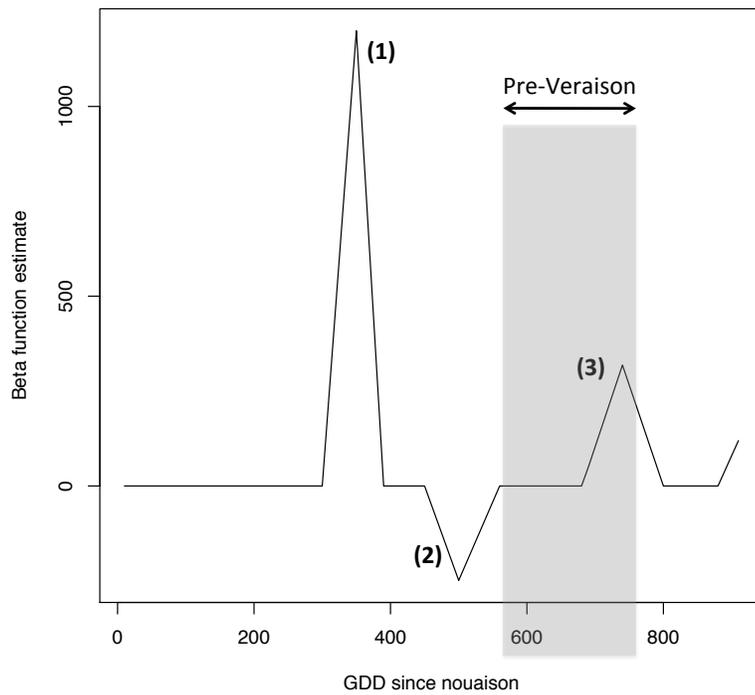


Figure 10: Beta Coefficient evolution over time (see Eq. 7), for explaining Sugar Concentration at harvest. Abscissa are in GDD. The values of  $\sigma = 0.05$  and  $\omega = 0.95$  have been found by cross-validation.

#### 565 4. Conclusion

566 The work presented in this paper used formalized knowledge and math-  
567 ematical models to design a software sensor from raw data and relate its  
568 temporal output to product quality. The proposed approach has been ap-  
569 plied to the case of a vine water stress indicator, and its relation to two grape  
570 quality variables: Berry Weight and Sugar Concentration.

571 Results provide a number of meaningful insights.

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

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

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

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

588 Second, the two-fold proposal for data analysis appears to be a good  
589 means of exploiting such temporal data as provided by the water stress in-

590 dicator  $Ks(t)$ . The results show that the water stress has an effect on grape  
591 quality. The analysis confirmed known results on the vine physiological re-  
592 sponse according to the variety and the irrigation effect. Thus our results are  
593 comforting the validity of the  $Ks$  indicator, and therefore the level of con-  
594 fidence and reliability in the software sensor design procedure. Functional  
595 data analysis highlighted critical periods for vine and berry development,  
596 regarding final quality features.

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

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

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

612 Applied perspectives of this work include the study of the relationship  
613 between vine water stress and other more complex quality features. In par-  
614 ticular new chemical analyses make it possible to follow the aroma develop-

615 ment in berries over time, which is assumed to be very sensitive to the grape  
616 water status.

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

#### 624 **Acknowledgments**

625 The research leading to these results has received funding from the Pi-  
626 lotype Program, funded by OSEO innovation and the Languedoc Roussillon  
627 regional council. The authors would like to thank all members of the project  
628 for their help and advices: Les grands Chais de France, Alliance Miner-  
629 vois, Les Vignerons du Narbonnais, INRA Pech Rouge, INRA SPO, INRA  
630 MISTEA, IFV Rhône Méditerranée, SupAgro Montpellier (ITAP), Fruition  
631 Sciences, Nyseos.

632 Finally, we wish to particularly thank Nicolas Saurin (INRA), Denis  
633 Caboulet, Jean-Christophe Payan and Elian Salançon (IFV) for providing  
634 the vine and wine-related data.

- 635 Allen, R. G., Pereira, L. S., Raes, D., Smith, M., 1998. Crop evapotranspi-  
636 ration: Guidelines for computing crop water requirements. Irrigation and  
637 Drainage Paper No. 56. FAO, Rome, Italy.
- 638 Ben-David, A., Sterling, L., 2006. Generating rules from examples of human  
639 multiattribute decision making should be simple. *Expert Syst. Appl.* 31 (2),  
640 390–396.
- 641 Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and  
642 regression trees. Wadsworth, Belmont, CA 1.
- 643 Cifre, J., Bota, J., Escalona, J., Medrano, H., Flexas, J., Apr. 2005.  
644 Physiological tools for irrigation scheduling in grapevine (*Vitis vinifera*  
645 L.)An open gate to improve water-use efficiency? *Agriculture, Ecosystems*  
646 *& Environment* 106 (2-3), 159–170.  
647 URL <http://linkinghub.elsevier.com/retrieve/pii/S0167880904002956>
- 648 Destercke, S., Buche, P., Charnomordic, B., jan. 2013. Evaluating data relia-  
649 bility: An evidential answer with application to a web-enabled data ware-  
650 house. *Knowledge and Data Engineering, IEEE Transactions on* 25 (1), 92  
651 –105.
- 652 Escalona, J., Flexas, J., Medrano, H., 2002. Drought effects on water flow,  
653 photosynthesis and growth of potted grapevines. *Vitis* 41, 57–62.
- 654 Ferreira, M. I., Silvestre, J., Conceição, N., Malheiro, A. C., Jun. 2012. Crop  
655 and stress coefficients in rainfed and deficit irrigation vineyards using sap  
656 flow techniques. *Irrigation Science* 30 (5), 433–447.  
657 URL <http://www.springerlink.com/index/10.1007/s00271-012-0352-2>

- 658 Ginestar, C., Eastham, J., Gray, S., Iland, P., 1998. Use of sap flow sensors  
659 to shedule vineyard irrigation. 1. effects of post-veraison water deficit on  
660 water relation, vine growth, and yield of shiraz grapevines. *Am. J. Enol.*  
661 *Vitic.* 49, 413–420.
- 662 Guarino, N., Oberle, D., Staab, S., 2009. Handbook on Ontologies, 2nd Edi-  
663 tion. Springer, Ch. What is an Ontology?
- 664 Hilgert, N., Mas, A., Verzelen, N., 2013. Minimax adaptive tests for the  
665 functional linear model. *The Annals of Statistics* 41 (2), 838–869.
- 666 James, G. M., Wang, J., Zhu, J., 2009. Functional linear regression that’s  
667 interpretable. *The Annals of Statistics* 37, 2083–2108.
- 668 Jones, H. G., Nov. 2004. Irrigation scheduling: advantages and pitfalls of  
669 plant-based methods. *Journal of experimental botany* 55 (407), 2427–36.  
670 URL <http://www.ncbi.nlm.nih.gov/pubmed/15286143>
- 671 Miller, G. A., 1956. The magical number seven, plus or minus two: Some  
672 limits on our capacity for processing information. *Psychological Review*  
673 63, 81–97.
- 674 Musen, M., 1992. Dimensions of knowledge sharing and reuse. *Computers*  
675 *and Biomedical Research* 25 (5), 435–467.
- 676 Olivo, N., Girona, J., Marsal, J., 2009. Seasonal sensitivity of stem water  
677 potential to vapour pressure deficit in grapevine. *Irrigation Science* 27,  
678 175–182.

- 679 Parker, A., De Cortázar-Atauri, I., Van Leeuwen, C., Chuine, I., 2011. Gen-  
680 eral phenological model to characterize the timing of flowering and verai-  
681 son of vitis vinifera l. Australian Journal of Grape and Wine Research 17,  
682 206–216.
- 683 Picón-Toro, J., González-Dugo, V., Uriarte, D., Mancha, L. a., Testi, L.,  
684 Jun. 2012. Effects of canopy size and water stress over the crop coefficient  
685 of a Tempranillo vineyard in south-western Spain. Irrigation Science  
686 30 (5), 419–432.  
687 URL <http://www.springerlink.com/index/10.1007/s00271-012-0351-3>
- 688 Quinlan, J., 1986. Induction of decision trees. Machine learning 1 (1), 81–106.
- 689 Quinlan, J., 1993. C4. 5: programs for machine learning. Morgan Kaufmann.
- 690 R Development Core Team, 2009. R: A Language and Environment for Sta-  
691 tistical Computing. R Foundation for Statistical Computing, Vienna, Aus-  
692 tria, ISBN 3-900051-07-0.  
693 URL <http://www.R-project.org>
- 694 Ramsay, J. O., Silverman, B. W., 2005. Functional data analysis, 2nd Edition.  
695 Springer Series in Statistics. Springer, New York.
- 696 Rijgersberg, H., van Assem, M., Top, J. L., 2013. Ontology of units of mea-  
697 sure and related concepts. Semantic Web 4 (1), 3–13.
- 698 Rodrigues, P., Pedroso, V., Gouveia, J. P., Martins, S., Lopes, C., Alves, I.,  
699 2012. Influence of soil water content and atmospheric conditions on leaf  
700 water potential in cv. touriga nacional deep-rooted vineyards. Irrigation  
701 Science 30, 407–417.

- 702 Santesteban, L. G., Miranda, C., Royo, J. B., 2009. Suitability of pre-dawn  
703 and stem water potential as indicators of vineyard water status in cv .  
704 Tempranillo. Society.
- 705 Thomopoulos, R., Destercke, S., Charnomordic, B., Iyan, J., Abécassis, J.,  
706 Feb. 2013. An iterative approach to build relevant ontology-aware data-  
707 driven models. *Information Sciences* 221, 452–472.
- 708 Ullah, S., Finch, C. F., 2013. Applications of functional data analysis: A  
709 systematic review. *BMC Medical Research Methodology* 13 (1), 43.  
710 URL <http://www.biomedcentral.com/1471-2288/13/43>
- 711 Villanueva-Rosales, N., Dumontier, M., 2008. Modeling life science knowledge  
712 with owl 1.1. In: *Proceedings of OWL'08*.
- 713 Williams, L., Baeza, P., 2007. Relationships among ambient temperature and  
714 vapor pressure deficit and leaf and stem water potentials of fully irrigated  
715 , field-grown grapevines. *Am. J. Enol. Vitic.* 58, 2.
- 716 Williams, L. E., Ayars, J. E., Oct. 2005. Grapevine water use and the crop  
717 coefficient are linear functions of the shaded area measured beneath the  
718 canopy. *Agricultural and Forest Meteorology* 132 (3-4), 201–211.  
719 URL <http://dx.doi.org/10.1016/j.agrformet.2005.07.010>
- 720 Zhang, Y., Kang, S., Ward, E. J., Ding, R., Zhang, X., Zheng, R., 2011.  
721 Evapotranspiration components determined by sap flow and microlysime-  
722 try techniques of a vineyard in northwest China: Dynamics and influential  
723 factors. *Agricultural Water Management*.  
724 URL <http://dx.doi.org/10.1016/j.agwat.2011.03.006>