



HAL
open science

A decision-support system to predict grape berry quality and wine potential for a Chenin vineyard

Nathalie Mejean-Perrot, Alberto Tonda, Ilaria Brunetti, Hervé Guillemain, Bruno Perret, Etienne Goulet, Laurence Guerin, Daniel Picque

► To cite this version:

Nathalie Mejean-Perrot, Alberto Tonda, Ilaria Brunetti, Hervé Guillemain, Bruno Perret, et al.. A decision-support system to predict grape berry quality and wine potential for a Chenin vineyard. *Computers and Electronics in Agriculture*, 2022, 200, pp.107167. 10.1016/j.compag.2022.107167 . hal-03815449

HAL Id: hal-03815449

<https://hal.science/hal-03815449>

Submitted on 14 Oct 2022

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.

1
2
3
4
5
6 **A Decision-Support System to Predict Grape Berry Quality and**
7 **Wine Potential for a Chenin Vineyard**
8

9
10 **Nathalie Mejean**^{*1¶}, **Alberto Tonda**¹, **Ilaria Brunetti**³, **Hervé Guillemain**², **Bruno Perret**³,
11 **Etienne Goulet**⁴, **Laurence Guerin**⁴, **Daniel Picque**³
12

13 ¹ Unité MIA-Paris, AgroParisTech, INRAE, Université Paris-Saclay, France.

14 ² Unité URTAL, INRAE, Unité de recherches en technologie et analyses laitières. Poligny, France.

15 ³ Unité SayFood-Grignon, AgroParisTech, INRAE, Unité de recherches en microbiologie et procédés alimentaires,
16 Université Paris-Saclay, France

17 ⁴ IFV Institut Français de la Vigne et du Vin, Unité de VINs, Innovations, Itinéraires, TERroirs et Acteurs, Amboise, France

18
19 **Abstract**

20 Grape berry ripening is a complex process, and predicting the quality of wine starting from the
21 ripening kinetics of grape berries is a challenging task. To tackle this problem, we present a
22 decision-support system based on coupling expert know-how with probability laws
23 encapsulated in a probabilistic model, a dynamic Bayesian network. The proposed approach
24 predicts the ripening kinetics of grape berries starting from initial measurements and weather
25 conditions, and then exploits the information to evaluate the potential of the wine that will
26 produced from them. The results show that the dynamic Bayesian network predicts the total
27 acidity concentration and the sugar content of the grape berries with a small amount of error

* Author to contact: nathalie.mejean@inrae.fr, UMR MIA, 16 rue Claude Bernard F-75231 Paris Cedex 05

28 (mean of 6% for total acidity concentration, 10% for sugar content) that is considered satisfying
29 by the experts, making it possible to predict the ideal moment for harvesting the grapes up to
30 two weeks in advance. Moreover, feeding the results from the probabilistic model to a fuzzy
31 expert model, the predicted trajectories are compared to an ideal trajectory described by wine
32 experts and formalized mathematically. From this comparison, it is possible to anticipate drifts
33 in wine sensory quality right from the step of grape ripening.

34

35 **Keywords:** Chenin Vineyard, Decision Support System, Dynamic Bayesian Network, Fuzzy
36 Logic, Grape Berry Ripening, Model Coupling, Wine Quality.

37

38 **Highlights**

- 39 • We propose a system to predict grape berries ripening and wine potential.
- 40 • The tool is based on a coupling between expert know-how and probability laws.
- 41 • Predictions from the system are considered satisfying by the experts.
- 42 • RMSE is less than 10% for all variables.
- 43 • The predictions help experts decide the date of harvesting.

44

45

46

47 **1. Introduction**

48 The ripening of grape berries is a complex process, including physicochemical and biochemical
49 reactions. Still more complex is predicting wine quality from the grape berries maturation
50 kinetics. Reactions during the ripening depend on multiple factors, with weather being the most
51 influential, especially in the last weeks preceding the harvest. Since berries ripeness plays a
52 major role in determining wine potentialities, correctly predicting the ripening process and
53 determining the ideal harvesting date is a significant challenge for the wine industry. While the

54 expert consensus is that choosing an appropriate moment for harvesting has a considerable
55 impact on the final quality of the wine, the exact effect has still not been exactly quantified
56 (Van Leeuwen, 2010). In this particular context, we define the wine potential of the grapes as
57 their capacity of producing a wine of at least acceptable quality, provided that the rest of the
58 processing is performed correctly. In other words, a good wine potential of the grapes is a
59 necessary, but not sufficient condition for obtaining high-quality wine.

60 The ripeness of grape berries can be evaluated resorting to different measurable quantities, for
61 example their sugar content, the color of their seeds, or their sensorial characteristics. Some of
62 these quantities can be measured exactly resorting to chemical means, while others require an
63 expert evaluation on a symbolic ordinated scale. In the last decade, new sensors have been
64 developed to easily measure grape characteristics such as color, sugar content, and aromatic
65 potentialities, (Ben Ghazlen et al., 2010), (Geraudie et al., 2010). Nevertheless, most of the
66 analyses are in practice still carried out in laboratory, with time-consuming and expensive
67 procedures required for a close monitoring of the grape berries ripeness and never used for on
68 line prediction. Literature reports relatively few contributions attempting to characterize the
69 link between the grape ripeness and a global indicator of wine quality. Interesting studies like
70 (Niimi et al., 2021) have worked on parameters determined by preprocessing techniques for
71 mid-infrared (MIR) spectra of grape berries to model sensory properties of different wines,
72 using Partial Least Squares (PLS) regression models. But the problem is also the cost and time
73 consuming of such approaches, considering the equipment needed to acquire the data. Some
74 studies like (Jensen et al., 2008) have linked grape phenolic composition to wine phenolic
75 composition and color of wines using multivariate analysis. Nevertheless it is in this case too
76 specific enough to be used directly for decision help directly at the field.

77

78 On the other hand, the development of mathematical models to predict or characterize different
79 mechanisms taking place during winemaking has been widely treated in literature. For example
80 the tool « Epicure » developed by the French Vineyard Institute (IFV) for managing
81 phytosanitary risks (Raynal, et al. 2010), or a work on the prediction of kinetics of fermentation
82 in wine processes like (Goelzer et al., 2009). For the grape ripeness prediction, a model has
83 been developed by (Baudrit, et al., 2015), (Perrot, et al., 2015) linking chemical indicators to
84 weather conditions on Cabernet Franc grape berries. [The approach presented in this paper is an](#)
85 [extension of this latter work, with several important differences: \(i\) the target grape berries are](#)
86 [for the Chenin wine, and this also reflects on the different structure of the dynamic Bayesian](#)
87 [network; \(ii\) in contrast with the approach outlined in \(Perrot, et al., 2015\), no expert system is](#)
88 [used to integrate the data set, as the available quantity of data is larger in this case; \(iii\) for the](#)
89 [first time, a link between variables related to grape berries ripening and wine quality is](#)
90 [provided, thanks to a fuzzy system.](#)

91 Few other works have been developed in this domain, except PLS approaches (Claverie, et al.,
92 2008) to link climatic and pedologic variables to chemical indicators of ripeness. Nevertheless,
93 those approaches used mathematical classifiers to link two different spaces: climatic and
94 pedologic ones but are not models of prediction of the kinetics of the physicochemical
95 compounds. An interesting study have been developed by (Petropoulos et al., 2017), based on
96 the development of a fuzzy tool linking different physicochemical and sensory parameters of
97 the grape berries to the wine quality. It was nevertheless developed for wine classification and
98 not for decision support during harvesting.

99

100 In this work, we present a novel decision-support system able to predict 15 days before the
101 maturation process, considering chemical indicators, by observing the weather conditions. The
102 predicted indicators' values are then used to evaluate the wine potential, according to the

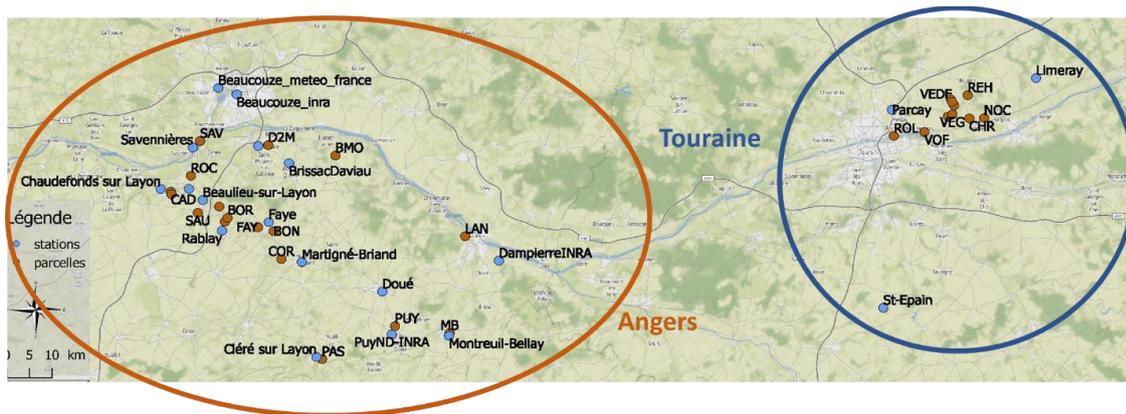
103 winemaker goal. This model relies on probability laws encapsulated in a dynamic Bayesian
104 network formalism learned on available data and a fuzzy expert model based on expertise.

105

106 2. Materials and Methods

107 2.1 Experimental data

108 Experimental data are gathered by the IFV institute from Chenin vineyards located in the French
109 region of “Vallee de la Loire” over several years, with weekly sample collections from July to
110 September before grape harvesting season. The data range from 1989 to 2017 (more precisely
111 1989, 1995-2001, 2016-2017) with land plots distributed between two geographical places,
112 “Anjou” and “Touraine” (see Figure 1), for a total of 30 vineyards and between 2 and 5 points
113 by kinetics for each vineyard according to the year of the experiment.



114

115 Figure 1: The geographical distribution of the Chenin vineyards considered in this work.

116 2.2 Variables of the model

117 The inputs used for the probabilistic model used in proposed approach are weather conditions
118 (see Table 1): Temperature (°C), rainfall (mm) and relative humidity (%) were supplied by
119 Meteo France meteorological stations located near and/or on the vineyards. Insolation (quantity
120 of solar radiation, in hours) was provided by one meteorological station located at Montreuil-
121 Bellay, in the middle of the geographical area of the study.

122

123

124 Table 1: Weather conditions, used as inputs of the proposed approach. With HR_{\min} is the
 125 lowest air humidity observe during the week, HR_{\max} is the highest air humidity observe during
 126 the week, T_{\min} is the lowest air temperature observe during the week, T_{\max} is the highest air
 127 temperature observe during the week

Variable	Description	Calculation	Unit
HR	Relative humidity	$mean(\sum_{i=1}^7 \frac{HR_{min_i} + HR_{max_i}}{2})$	%
T	Temperature	$\sum_{i=1}^7 \frac{T_{min_i} + T_{max_i}}{2}$	°C
Pl	Rainfall	$\sum_{i=1}^7 Pl_i$	mm
Ins	Insolation	$\sum_{i=1}^7 Ins_i$	h

128

129

130

131 The outputs of the probabilistic models include both physicochemical and sensory
 132 measurements. The physicochemical measurements have been selected after a discussion with
 133 wine experts, as those considered essential: sugar content (s) measured in g/l, total acidity
 134 concentration (ac) in gH_2SO_4/l , and malic acid concentration (ac_m) in g/l, (Barbeau, 2003),
 135 (Riou, 1994). Their variations during a week (defined as the difference between data collected
 136 in two subsequent time points) are also considered: variation in sugar content (Var_s), variation
 137 in total acidity concentration (Var_ac), and variation in malic acid concentration (Var_ac_m).
 138 Each week, a lot of 200 berries of Chenin, with pedicels, were randomly selected from each
 139 vineyard according to the method of Vine and Wine French Institute (ITV-France) (Cayla et
 140 al., 2002) in order to limit the effects of the grape heterogeneity. With the set of 200 berries of
 141 each sampling, a crushing was realized with a blender, then the must was filtered through a
 142 Whatman paper filter. Sugars content (g/l) was measured with a refractometer; total acidity
 143 concentration (g/l eq. H_2SO_4) was measured by the titration method and malic acid (g/l).

145 **2.3 Expert knowledge**

146 Two types of experts were interviewed: 3 scientists and 2 winegrowers working on the two
147 areas considered in this study. Each expert was interviewed during one or two sessions
148 (spanning 2–3 hours each). Each of the elicitation sessions was attended by one expert and one
149 or two interviewers. Three hours were allocated for each interview. Notes were taken and
150 audiotapes were recorded. After each session, the tapes were re-played several times to make
151 sure the notes were accurate and complete. To build the interview, adapted methods proposed
152 by (Sicard et al., 2011) were applied. The elicitation process was based on a set of
153 predetermined structured open-ended questions used to direct the interviews. Questions were
154 designed according to techniques based on survey methods with the aim of optimizing the
155 expression of expert knowledge. The objective was to ask clear and simple questions without
156 ambiguities and that did not implicitly direct the expert towards a specific answer. Questions
157 were also asked in such a way to encourage simulation of the expert's situation using the
158 explicitation interview method developed by (Vermersch, 1994) (i.e. cognitive interviews,
159 (Moody, Will, & Blanton, 1996)). We paid particular attention to context reinstatement as
160 recommended in these methods. This involves having the expert think about and describe their
161 feelings during the episodes being recalled.

162 Previous works (Perrot et al., 2015) had access to a larger number of experts, but it is important
163 to remark that in the past the aim was to integrate the available data on grape berries ripening
164 with an expert system, while in the current approach the objective is to link grape berries
165 ripening kinetics to wine quality, creating a completely novel model. For this particular task,
166 the availability of domain experts is relatively scarcer.

168 **2.4 Mathematical formulation**

169 The predictive model used in this work is a coupling between two models: a probabilistic
170 graphical model, dynamic Bayesian networks in particular, and a fuzzy expert model. The first
171 one is relevant to formalize mathematically, the implicit information, contained in the available
172 experimental data. With the second one, it is possible to interact with the wine experts of Chenin
173 vineyards and ultimately formalize the explicit knowledge, encoded through expert interviews
174 and know-how.

175 **2.4.1 Introduction to the modeling approaches**

176 The first model used in our approach is a Dynamic Bayesian Network (DBN), a probabilistic
177 graphical model able to describe phenomena developing over time (Jensen & Nilsen,
178 2010)(Pearl, 1988). The structure of a DBN is an oriented graph, representing correlations
179 between variables, which in our case was created by interacting with human experts of the
180 Chenin wine. Once the structure of a DBN is fixed, it is then possible to compute its parameters
181 starting from a training dataset: the parameters are conditional probability tables, assessing the
182 probability for variables taking a specific value, knowing the values of the variables they
183 depend on. For our specific application, the values of the variables need to be discretized.
184 Differently from a classical Bayesian Network, a DBN makes it possible to estimate variable
185 values over several subsequent time steps. In our case, each time step is equivalent to one week
186 in the grapes ripening processes. DBNs have been successfully adopted for several agri-food
187 applications (Baudrit, et al., 2015) (Perrot, et al., 2015).

188 Coupled to this first model, a fuzzy expert model is set up to mathematically describe the
189 explicit knowledge of the experts concerning the complex link between the maturation of the
190 grape berries and the wine potential.

191

192

2.4.2 The DBN algorithm

193

More formally, a DBN is a graph-based model of a joint multivariate probability distribution,

194

capturing properties of conditional independence between variables. Like a BN, a DBN is a

195

directed acyclical graphs (DAG) where the nodes represent variables, and the missing arcs

196

represent conditional independences between variables. In DBNs in particular, nodes $X(t) =$

197

$(X_1(t), \dots, X_n(t))$, represent n discrete random variables, indexed by time t , providing a

198

compact representation of joint probability distribution P for a finite time interval $[1, \tau]$. In

199

other words, the joint probability P can be written as the product of the local probability

200

distribution of each node and its parents, as follows, Equation 1:

201

$$P(X(1), \dots, X(\tau)) = \prod_{i=1}^n \prod_{t=1}^{\tau} P(X_i(t)|U_i(t)) [1]$$

202

Where $U_i(\cdot)$ denotes the set of all parents of node $X_i(\cdot)$, and $P(X_i(\cdot)|U_i(\cdot))$ describes the

203

conditional probability function associated with random variable $X_i(\cdot)$ given the values of $U_i(\cdot)$

204

. $X_i(t)$ is termed “slice”, and it represents the set of all variables at time t . This factorization of

205

the joint probability distribution, based on information from the graph, makes it possible to

206

straightforwardly represent large models, and use them for practical applications. In other

207

words, DBNs represent the beliefs of possible trajectories of the variables involved in a dynamic

208

process.

209

In order to make the problem treatable, DBNs assume the first-order Markov property: the

210

parents of a variable in time slice t must appear in either slice $t - 1$ or t . As a consequence, for

211

the first-order homogeneous Markov property, the conditional probabilities are time-invariant,

212

meaning that $P(U(t)) = P(U(2)) \forall t \in (1, \tau)$. In order to fully specify a DBN, we will then

213

need to define the intra-slice topology (within a time slice), the inter-slice topology (between

214

two time slices), as well as the parameters (i.e. conditional probability functions) just for the

215 first two time slices. The structure of a model can be explicitly built on the basis of knowledge
216 available in the literature and parameters can be automatically learned without a priori
217 knowledge on the basis of a dataset, a process termed parameter learning. The techniques for
218 learning DBNs are generally extensions of the techniques for learning BNs. Specialized
219 literature reports several methods to learn the structure or the parameters of a DBN from
220 substantial and/or incomplete data (Geiger & Heckerman, 1997); (Heckerman, 1999). In our
221 work, the topology of the graph is obtained from expert knowledge; for parameter learning, we
222 consider the simplest and most commonly adopted methodology, simply evaluating the co-
223 occurrence rate of values of variables in the training data.

224 Once a DBN is fully specified, it can be used to estimate marginal probabilities for target
225 variables, through a process also known as Bayesian inference (Equation 2):

$$226 \quad P(X(t)|O(t') = o(t')), \forall t' \in [1, \tau] \quad [2]$$

227 Where X is a set of variables whose values we are interested in predicting, and O is a set of
228 variables whose values are known (for example, in food processing X might be the variables
229 representing the physicochemical properties of a product and O might be the variables
230 representing the observed environmental conditions). In general, given a way of calculating
231 $P(X(t)|O(t'))$ from the knowledge of $P(X(t')|O(t))$, inference in a DBN is performed using
232 recursive operators and Bayes' theorem, updating the belief state of the DBN as new
233 observations become (Murphy, 2002).

234 **2.4.3 The fuzzy expert algorithm**

235 The second model used in our approach comes from the theory of fuzzy sets. Fuzzy logic was
236 originally proposed by (Zadeh, 1965), and it is an extension of set theory by the replacement of
237 the characteristic function of a set by a membership function whose values range between 0 and
238 1. Soft transitions between sets are thus obtained and make it possible to represent gradual

239 concepts, as well as the representation and the inference of linguistic rules stemming from
 240 expertise. This type of formalism is particularly adapted for taking human linguistic and
 241 reasoning processes into account (Perrot et al., 2006). Fuzzy models can be written in an easy
 242 form to understand for an expert, ie linguistic rules. Similarly, an essential fuzzy notion is the
 243 fuzzy membership function. A fuzzy set E in universe of discourse U can be defined by
 244 Equation 3:

$$E = \{(u, \mu_E(u)) \mid u \in E\}$$

$$\mu_E: U \rightarrow [0,1] \quad [3]$$

246 μ_E is thus the membership function of set E, and it represents the set of membership grades
 247 ($\mu_E(u)$) of a numerical variable u mapped to a fuzzy set E. This function makes it possible to
 248 link real numerical variables to a given linguistic variable. The value of the membership grade
 249 is a real number within the interval [0;1], expressing the translation from one space X to another
 250 space Y. It was set up with the experts. This notion gives the way to link a numeric variable to
 251 a linguistic variable often manipulated by the operators. In fact, fuzzy memberships are used to
 252 describe how much an object belongs to a linguistic notion. Going back to an example:

253 Suppose a value of sugar of 195 g/l belonging to the symbol “not target” with a membership
 254 degree of 0.5 and to the symbol “target” with a membership degree of 0.5. It would mean that
 255 the maturation will be mitigated for this value of sugar.

$$\mu(x) = \begin{cases} 0 & (x < a_1) \\ \frac{x - a_1}{a_2 - a_1} & (a_1 \leq x < a_2) \\ \frac{a_3 - x}{a_3 - a_2} & (a_2 \leq x < a_3) \\ 0 & (a_3 \leq x) \end{cases} \quad [4]$$

256

257 Membership functions can be expressed through various representations. The representations
258 most widely used are triangular (Equation 4) for a given triplet series of parameters a_1, a_2, a_3 .
259 On the basis of the definition of the fuzzy subsets, the fuzzy Tnorm, is used in this paper to
260 aggregate the information associating the three input variables of the fuzzy algorithm: the sugar
261 content, the total acidity concentration and the malic acid concentration, to the output: the wine
262 potential. Those input variables are joined by a connector “AND” (a classical mathematical
263 logical interpretation of the join) using the fuzzy Tnorm (Equation 5).

$$264 \quad Tnorm(\mu_0, \mu_1, \dots, \mu_n) = \prod \mu_i \quad [5]$$

265 Where μ_i are the input variables (three in our case study: sugar content, total acidity and malic
266 acid concentrations).

267

268 **2.4.4 Models evaluation**

269 The DBN previously introduced will be evaluated with a leave-one-out cross-validation
270 (LOOCV), where the model is repeatedly trained on the whole dataset, minus one sample, and
271 the remaining sample is used for testing. The procedure is repeated until each sample has been
272 used for testing. Considering the mean and standard deviation on the results of a LOOCV
273 provides a better estimate of the model’s capabilities than just considering a random split of the
274 available data between a training set and a test set (Geisser, 1993).

275 For the choices made in this study, before training the model, it is necessary to discretize the
276 real-valued variables in the dataset (see subsection 3.1.1). However, in order to evaluate the
277 performance of the model’s predictions against the ground truth, the results of the model will
278 have to be converted back into real values. Recalling that the predictions of a DBN model for
279 variable x will consist in a series of probabilities P_i for each possible discrete class $i =$

280 0, 1, ..., n_x associated with variable x , the predicted outcome can be converted to a real value
281 using the following Equation [6]:

$$282 \quad x_i^{predicted} = \sum_{i=1}^{n_x} \bar{x}_i P_i \quad [6]$$

283 Where \bar{x}_i is the average value of all samples of variable x that fall under class i .

284 The first metric used to evaluate the quality of the predictions against the ground truth is the
285 root mean squared error (RMSE), Equation [7]:

$$286 \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^{predicted} - x_i^{observed})^2} \quad [7]$$

287 Where N is the number of predictions considered for target variable x , and $x^{observed}$ indicates
288 its observed value. In this study, we will also use the relative RMSE (RRMSE) that expresses
289 the RMSE as a percentage of the range of observed values for the target variable, and it is thus
290 more informative as an error metric (Equation 8):

$$291 \quad RRMSE = \frac{RMSE}{(x_{max}^{observed}) - (x_{min}^{observed})} \times 100 \quad [8]$$

292
293 Where $x_{max}^{observed}$ and $x_{min}^{observed}$ are the highest and lowest values observed for variable x ,
294 respectively.

295 For the fuzzy algorithm, the results are evaluated as classification accuracy in percentage,
296 expressing the number of points classified correctly within the range of precision fixed by the
297 experts, over the total amount of points classified.

298 **3 Results**

299 The results presented in this section include both the numerical assessment of the models'
300 predictions, and the trained models themselves. The models are considered an output of this
301 work, as they will be exploited by experts in the field to predict the best moment for harvesting
302 grapes, once they have reached the proper degree of ripeness.

3.1 Description of the Chenin vineyard global model

The Chenin vineyard model is based on a coupling between a DBN expressing the kinetics of ripening of the Chenin grape berries and a Fuzzy expert model expressing the expert knowledge on empirical laws linking kinetics of maturation and wine sensory potentialities. We first built a Chenin Dynamic Bayesian Network (Chenin_DBN), which makes it possible to obtain reliable dynamic predictions of sugar content, total acidity, and malic acid concentrations by measuring air temperature, rainfall, relative humidity, and insolation hours in the three weeks preceding the harvest.

Once we predict the maturation indicators, we evaluate the wine potentialities according to the winemaker expectation, by means of a Chenin Fuzzy expert model (Chenin_FEM), linking the maturity indicators to global wine quality trajectories (see Figure 2).

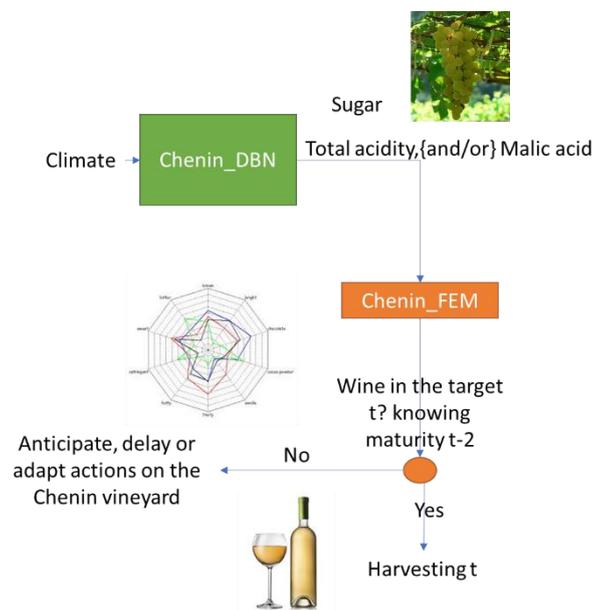


Figure 2: The Chenin vineyard model developed in this work consists of two coupled models: one for the prediction of the physicochemical variables using a DBN (Chenin_DBN), and another based on a Fuzzy expert theory (Chenin_FEM) dedicated to the prediction of the potentialities of the wine for being in the target required by the winemaker.

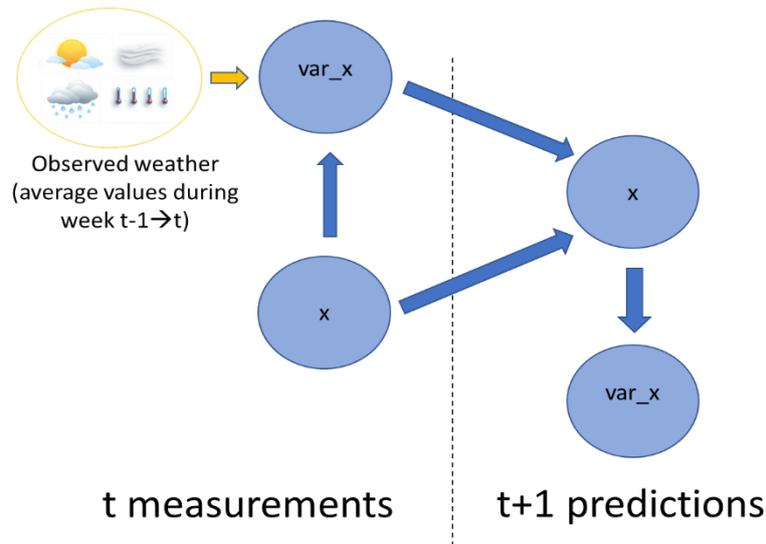
320 **3.2 The Chenin_DBN model**

321 The three key elements of the proposed Chenin_DBN model are: (i) its structure, defining the
322 relationships between the problem's variables, (ii) the choice of discretization for each variable,
323 (iii) the parameters, expressed by conditional probability tables (CPTs), which describe how
324 the probabilities of variables assuming a given value change, depending on the values of the
325 variables they depend on. The structure of the network has been defined on the basis of expert
326 knowledge, building upon previous works on different wines that led to the development of the
327 software PREVIMAT (Brousset et al., 2009; Baudrit et al., 2015).

328 **3.2.1 Network structure of the Chenin_DBN**

329 The first part of the model developed, inspired by previous work on Cabernet-Franc and Gamay
330 wines (Baudrit et al., 2015) predicts physicochemical indicators starting from weather
331 conditions (Figure 3). For each physicochemical ripeness indicator, the climatic variables
332 playing a key role in their kinetics are selected from expert knowledge and literature. In
333 particular, relative humidity only affects the two acidities (total acidity concentration and malic
334 acid concentration), sunshine influences sugar content, while temperature and rainfall have an
335 impact on the three variables considered: sugar content (s), total acidity concentration (ac),
336 malic acid concentration (ac_m).

337



338

339 Figure 3: Representation of the Chenin_DBN, in the form of two generic slices that can be
 340 unrolled on several slices representing the different step times. DBNs assume the first-order
 341 Markov property, which means that the parents of a variable in time slice t must occur in other
 342 slices and the conditional probabilities are time-invariant. The slice representing the time t (t
 343 measurements) is concerned at the beginning of the iterations by variables that are measured at
 344 time t_0 . The consecutive slice: time $t+1$ is dedicated to predictions. If several slices are added,
 345 for example t , $t+1$ and $t+2$, it starts at t_0 with an initialization where variables are measured,
 346 followed by two slices predicted $t+1$ and $t+2$, with $t+2$ predicted on the basis of the prediction
 347 of $t+1$.

348

349 As the Chenin_DBN needs to be able to capture dynamical variations of the values over time,
 350 to better predict the three variables it is necessary to define new intermediate state variables. In
 351 particular, having collected the data related to the last two to five weeks (according to the years
 352 of experiment) before the harvest, we can make the assumption that the trajectory in time of
 353 each variable is stable and its variation is constant, given identical meteorological conditions.
 354 In other words, we consider that a month before the harvest, only alterations in the weather can
 355 cause a significant deviation from an established trajectory in time.

356 More formally, considering each physicochemical variable $x \in \{ac, ac_m, s\}$ at time t and
 357 $t + 1$:

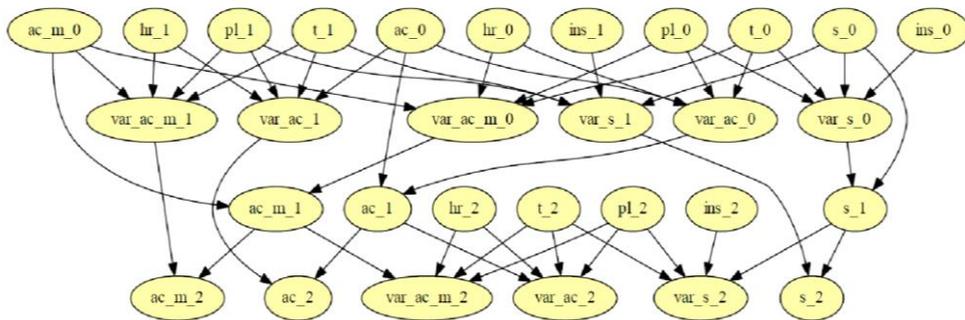
$$358 \quad x(t + 1) = var_x(t + 1) + x(t)$$

359 And consequently

$$360 \quad var_x(t) = x(t) - x(t - 1)$$

361 As already mentioned, the (absolute) value of a variable can be used as an indicator of the
 362 current stage of ripening, while its variation, as a function of the climatic variables, will dictate
 363 the ripeness trajectory.

364 Having taken into account the variation of the physicochemical variables over time and the
 365 limitation for certain years to two weeks of history, the Chenin_DBN can now be structured
 366 over the minimum time steps history of the whole kinetics, which is three time steps $t =$
 367 $\{0, 1, 2\}$, each one spanning a week before the time of the harvest. At time $t = 0$, the value of
 368 each variable is known; for the next two time steps, only the climatic variables are known
 369 (observed), while the physicochemical quantities and their variations are predicted by the
 370 model. The complete structure of the Chenin_DBN for the physicochemical variables, when
 371 considering all variables for the three time steps, is presented in Figure 4.



372
 373 Figure 4: The unrolled Chenin_DBN model over time steps $t = \{0, 1, 2\}$. A physicochemical
 374 variable ($t + 1$), considered at time step $t + 1$ is predicted as a function of its value at the
 375 previous time step t and its variation $var_x(t) = x(t) - x(t - 1)$. In turn, $var_x(t)$ is

376 predicted as a function of the climatic variables at time steps $t - 1$ and t . It is interesting to
377 notice that the values of the physicochemical variables are not influenced by each other.

378

379 **3.2.2 Variables discretization**

380 As previously described, to create the CPTs of our Chenin_DBN model, it is necessary to define
381 the discretization of the continuous variables in the problem. In this context, discretizing
382 variable x amounts to finding several intervals $\{[x_1, x_2), [x_2, x_3), \dots, [x_{n-1}, x_n)\}$ of continuous
383 values such that $x_1 < x_2 < \dots < x_n$, with each interval corresponding to a discrete class.

384 The definition of such intervals has a considerable impact on the performance of the
385 Chenin_DBN model, so this step is crucial for generalized the approach and to obtain satisfying
386 predictions. Specialized literature reports different established partial solutions for
387 discretization: for example, considering quantiles, nested averages, and amplitudes (Dougherty
388 et al., 1995).

389

390 In our case we opted for test the latter solution, since it makes it possible to create intervals of
391 similar amplitude for each category of variables in the problem: climatic variables, grape
392 sensory variables, and variables indicating instantaneous physicochemical quantities, or
393 variations of physicochemical quantities from a time step to the next. The number of classes
394 was chosen by selecting a limited but sufficiently representative number of intervals, following
395 the advice and recommendations of human experts.

396 For the climatic variables, the following intervals were defined (see Table 1 for the detailed
397 description of the variables):

- 398
- $Ins = [[15,30], [30,40], [40,55], [55,60], [60,75]]$

399 • $Pl = [[0,10], [10,20], [20,30], [30,45], [45,70], [70,100]]$

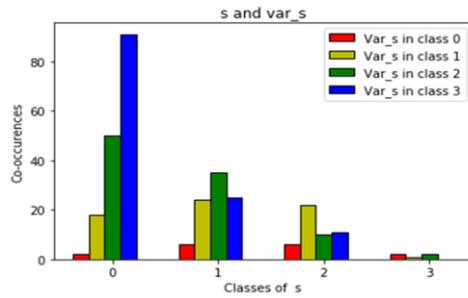
400 • $T = [[0,11], [11,15], [15,17], [17,19.5], [19.5,22]]$

401 • $HR = [[60,70], [70,75], [75,80], [80,90], [90,100]]$

402 For the physicochemical variables , an interactive semi-automated discretization approach was
403 developed, based on the notion of co-occurrence between variable values and their variations.

404 The methodology is based on a visualization software, EvoGraphDice, coupled with an
405 evolutionary optimization approach (Boukhelifa et al., 2017). Variations of the variable (for
406 example var_s for the sugar content) are fixed by the experts. Diagnostic and decisions of the
407 experts are indeed more based on the physicochemical variations during one week than on the
408 values themselves. It is then easier for them to describe the values in terms of variations. For
409 example, the variation var_s of the variable describing sugar content from one week to the next
410 is fixed on the basis of the expert description. The optimal discretization of the variable itself
411 like the sugar content for example, is then calculated by optimization, to ensure a repartition of
412 the var_s classes of interval as homogeneous as possible, for each sugar interval in the data.
413 Several iterations are performed to reach a good compromise between the values of
414 discretization proposed by the algorithm of optimization and what is considered to be coherent
415 according to the expert evaluation. Experts visualize the results of the optimization at each
416 iteration, and validate or reject the result.

417 An example of discretization obtained for sugar content is presented in Figure 5. We can see
418 that classes of s are created by the optimization algorithm after 5 runs of interaction with an
419 expert, with a good result in terms of repartition of the different classes of Var_s for each class
420 of s. This homogeneous repartition is required for a good learning of the probability laws in the
421 DBN model structure.



422

423 Figure 5: Result of the discretization of the variable s (classes equivalence: see Table 2) after 5
 424 iterations and expert's interactions, var_s being fixed by the expert. The automatic
 425 discretization was used for values ranging from 0 to 210 g/l, generating four classes. Points
 426 with values above 210 g/l are rare, but rather than assigning them all to the same class as the
 427 algorithm proposed, the experts decided to divide the space evenly in three more classes with
 428 an amplitude of 10 g/l, plus one final class for all values above 240 g/l.

429 All the results of the optimization and thus the discretization proposed for the physicochemical
 430 variables are presented in Table 2.

431 Table 2: Discretization of the physicochemical variables, fixed by experts for the variation
 432 var_X of variable X, obtained by the experts through interactive optimization.

	Discretization of X	Discretization of Var_X
S (Sugar)	<ul style="list-style-type: none"> • Class 0 = $[\infty, 156.9]$ • Class 1 = $[156.9, 182.86]$ • Class 2 = $[182.86, 201.8]$ • Class 3 = $[201.8, 210]$ • Class 4 = $[210, 220]$ • Class 5 = $[220, 230]$ • Class 6 = $[230, 240]$ • Class 7 = $[240, +\infty]$ 	<ul style="list-style-type: none"> • Class 0 = $[0, 12]$, • Class 1 = $[12, 20]$ • Class 2 = $[20, 35]$ • Class 3 = $[35, +\infty]$
ac (total acidity concentration)	<ul style="list-style-type: none"> • Class 0 = $[-\infty, 5.47]$ • Class 1 = $[5.47, 6.33]$ • Class 2 = $[33, 7.94]$ • Class 3 = $[7.94, +\infty]$ 	<ul style="list-style-type: none"> • Class 0 = $[-\infty, -1.5]$ • Class 1 = $[-1.5, -1]$ • Class 2 = $[-1, -0.6]$ • Class 3 = $[-0.6, 0]$
ac_m (malic acid concentration)	<ul style="list-style-type: none"> • Class 0 = $[-\infty, 3.66]$ • Class 1 = $[3.66, 4.6]$ 	<ul style="list-style-type: none"> • Class 0 = $[-\infty, -2.5]$ • Class 1 = $[-2.5, -1.5]$

	<ul style="list-style-type: none"> ● Class 2 = [4.6, 5.68] ● Class 3 = [5.68, 6.88] ● Class 4 = [6.88, +∞] 	<ul style="list-style-type: none"> ● Class 2 = [-1.5, -0.75] ● Class 3 = [-0.75, -0.5] ● Class 4 = [-0.5, 0]
--	---	---

433

434

435 3.2.3 Network parameters

436 Once the network structure has been defined, and the variables have been properly discretized,
 437 learning the parameters of the Chenin_DBN is a straightforward process. Each node represents
 438 a conditional probability table, describing the different probabilities for variable x to assume
 439 values x_1, x_2, \dots, x_N given the values of the parent variables it depends on, $pa(x)$. By reading
 440 the training data, the frequency of appearance of values for x together with the values for $pa(x)$
 441 for the same samples, can be directly used as a probability to fill the conditional probability
 442 tables, using the classic strategy of maximum likelihood (Redner & Walker, 1984).

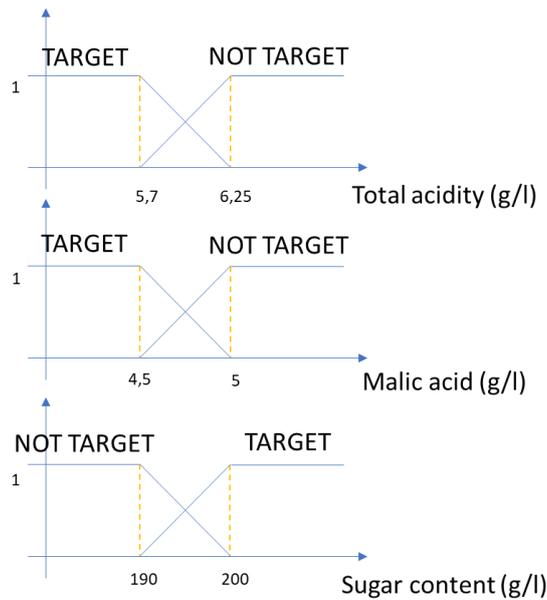
443

444 3.3 The Chenin Fuzzy Expert Model (Chenin_FEM)

445 The Chenin_FEM is constituted of a set of membership functions adapted to the Chenin
 446 characteristics of variation, and rules of aggregation leading to a quantification of the output of
 447 the model.

448 3.3.1 Membership functions

449 The membership functions (see Section 2) set up in the Chenin_FEM are presented in Figure 6
 450 for the three inputs considered: sugar content, total acidity concentration and malic acid
 451 concentration.



452
 453 Figure 6: Membership function and associated targets for the three variables manipulated by
 454 the expert to infer the quality of the Chenin vineyard: total acidity concentration, malic acid and
 455 sugar content.

456 3.3.2 Rules of aggregation

457 Two experts were interviewed in this study, to find a link between the ripeness of grape berries
 458 and their personal assessment of wine quality. Experts encode their knowledge as a symbolic
 459 trajectory towards ideal wine quality, and assess differences in terms of drift from this
 460 trajectory. The experts classify grape berry ripeness into three categories (that we named Class
 461 1, 2, and 3, see Table 3), in link with three steps of ripening that in their opinion should appear
 462 sequentially at given times during the ripening process. In case of drift from the perceived ideal
 463 category, experts usually perform their corrective actions after the harvesting, either by mixing
 464 grapes from different batches, or changing the parameters of the wine fermentation.
 465 Nevertheless, each expert has their own rules for managing this situation. In this work, we
 466 decided to use the rules of the expert with the longest track record in wine production.

467 The rules used by the experts for the aggregation of the three terms are presented in Table 3 and
 468 processed according to the fuzzy Tnorm methodology presented in Section 2.

469

470 Table 3: Expert rules of aggregation for the three input variables of the Chenin vineyard: sugar
 471 content (g/l), total acidity concentration (g/l) and malic acid concentration (g/l). The output is
 472 the projection of the class of wine inferred by the aggregation, knowing the values of the input
 473 variables: Class 3 represents the quality of the berries required to obtain wine that is at the very
 474 least of acceptable quality according to the expert. Class 2 means that the maturity of this
 475 vineyard plot is not enough to reach an acceptable wine quality, but it can potentially be still
 476 corrected by pursuing the ripening of the grapes, postponing the date of the harvest if possible;
 477 or, if not possible, by adapting the fermentation process or mixing the berries of different
 478 vineyard plots. Class 1 means that the maturity of the berries have, at the time of measurement,
 479 not the good potential to produce a good wine. It is generally associated with berries at the very
 480 beginning of the ripening step.

Sugar	TARGET		NOT TARGET	
	Malic acid TARGET	Malic acid NOT TARGET	Malic acid TARGET	Malic acid NOT TARGET
Total acidity concentration				
TARGET	3	2	2	Non existant
NOT TARGET	1	1	1	1

481

482 **4 Experimental results**

483 This section describes the experimental results obtained by comparing the complete models
 484 developed in Section 3 against the available data.

485 4.1 Chenin_DBN model predictions for the physicochemical variables

486 A LOOCV is performed on the dataset. At each iteration the network is trained on the whole
487 dataset minus one sample, and then tested on that sample. We obtained a mean RRMSE for
488 each predicted variable.

489 The results (Table 4) show that it is possible to predict with good results the total acidity
490 concentration and the sugar content in a range that is satisfying for the experts (10% mean error
491 for sugar, 6% mean error for the total acidity concentration) and so anticipate the maturation
492 date up to 15 days in advance. For the malic acid, it seems to be more complex to have a good
493 prediction with the only variables considered as inputs of the DBN. Other physicochemical
494 variables, that are harder to measure and were not included in this study, such as characteristics
495 of the soil, would probably be beneficial towards the aim of obtaining better predictions.

496 Table 4: Results of prediction for the three variables for two time steps: one week (1) or two
497 weeks (2) in advance.

Variable in the unrolled DBN model	RMSE	RRMSE % ac = [3.4,12.5]; ac_m = [1.7,10]; s = [144,271.8]	R2
ac_1 (g/l)	0.536	6	0.668
ac_2 (g/l)	0.648	7	0.583
ac_m_1 (g/l)	0.825	9	0.6673
ac_m_2 (g/l)	0.867	10	0.5302
s_1 (g /l)	11.37	8	0.702
s_2 (g/l)	12.87	10	0.67

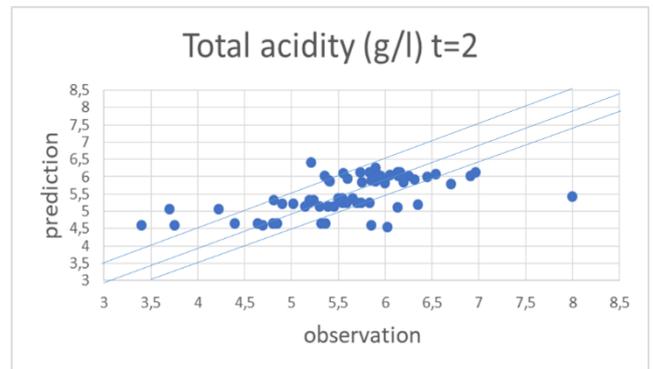
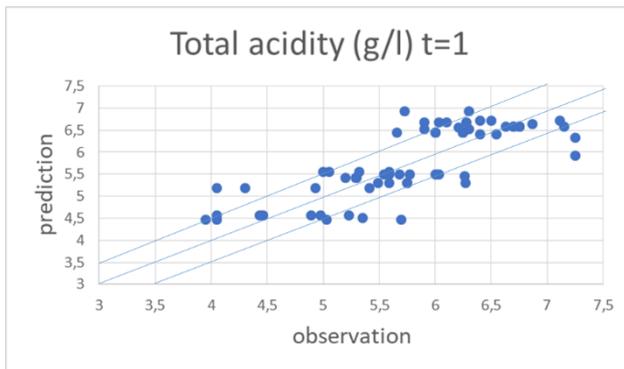
498

499 In Figure 7, the model's predictions of the physicochemical data are compared with the
500 observed values, for each variable and for the two considered time steps 1 and 2. Even if the R²
501 value are moderately satisfactory, varying from 0.702 for the sugar to 0.53 for the malic acid,
502 with cumulative errors for two time steps, scatterplots are relatively well aligned around the
503 bisector, and most of the errors of the predicted points are within the range of uncertainty of the

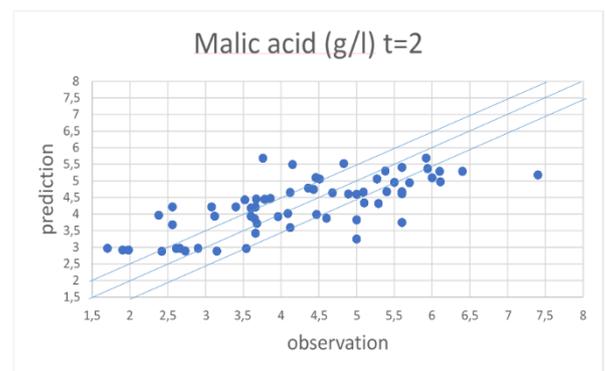
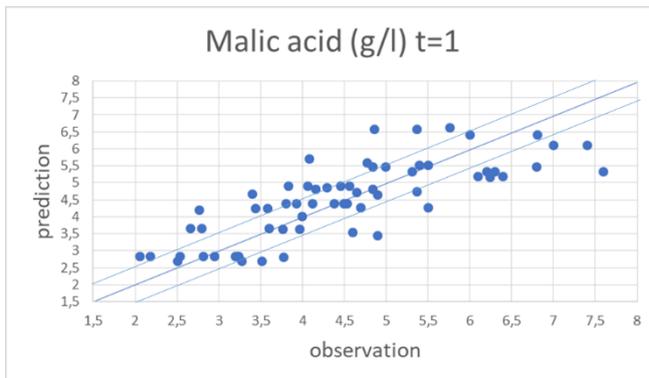
504 measurements defined by the experts. Essentially, 5 vineyards out of 30 were not well predicted
505 and were not considered for the computation of the correlation coefficients. It is a possible
506 effect of the lack of data on those land plots with particular behaviors and an increasing
507 uncertainty. As described above, the proposed approach faces more difficulties for the malic
508 acid variable, but the results are still acceptable.

509

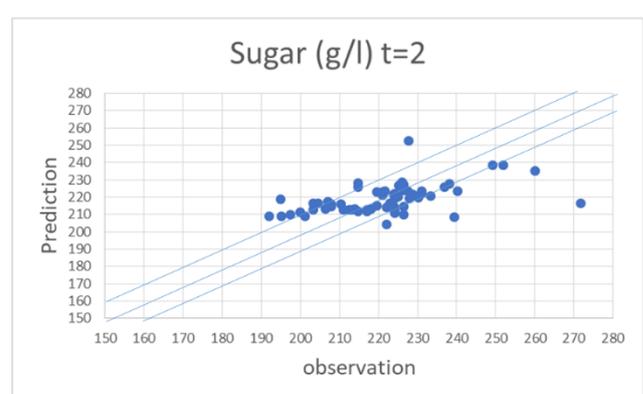
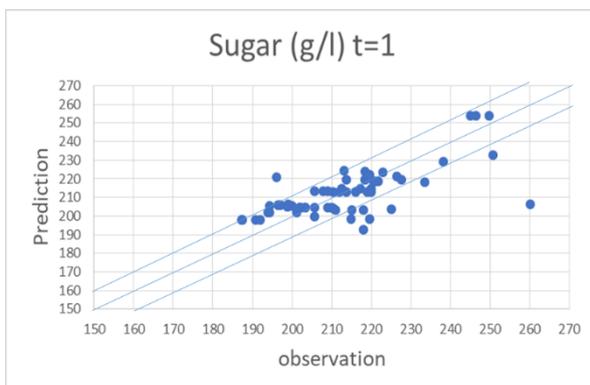
510



511



512



513

514 Figure 7: Scatter plots for the predictions of the physicochemical variables by the Chenin_DBN
515 model, for the two time steps $t=1$ and $t=2$. The error range considered acceptable by the experts
516 for those variables is represented by the dotted lines in the plots ($\pm 0.5g/l$ for the total acidity
517 concentration and malic acid concentration, and $\pm 10g/l$ for the sugar concentration).

518

519 **4.2 Validation of the Chenin_FEM**

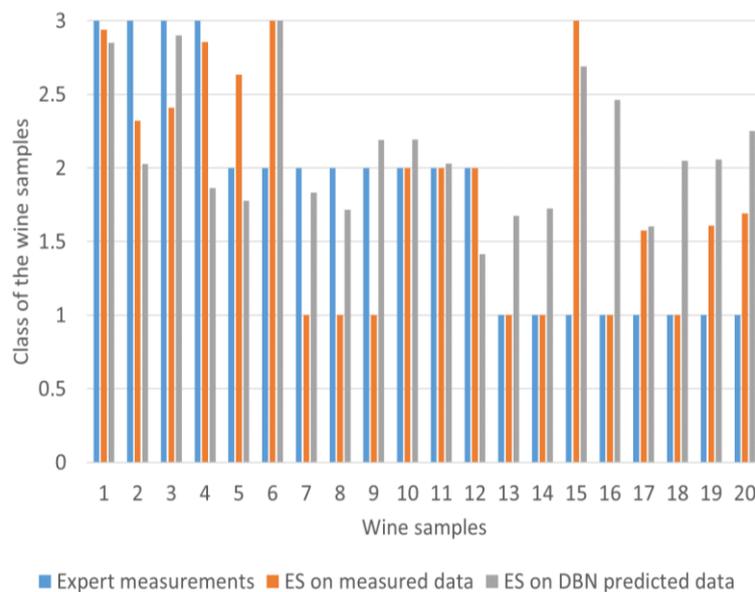
520 The Chenin_FEM model is tested in two separate experiments: Firstly, the predictions of the
521 fuzzy expert module are tested with inputs observed in the data; Secondly, the outputs of the
522 DBN presented above are used as inputs of the expert system. In this last case, the uncertainties
523 pertaining the prediction of the two models are cumulated.

524 The proposed approach is tested on twenty harvests performed in different batches, equivalent
525 to twenty samples in the database, are tested from 2016 and 2017, at different time steps of
526 maturation, and evaluated by a Chenin sensory expert.

527 The results are presented in Figure 8 for the predictions of Chenin_FEM based (1) on measured
528 input data and (2) on the Chenin_DBN predictions. The inputs are the physicochemical values
529 measured or predicted on the berries batch just before the grape harvest: the sugar
530 concentration, the total acidity concentration and the malic acid concentration. The output is
531 the prediction of the wine quality, considering those inputs. Inputs are data measured just before
532 the harvesting of the batch used to produce the wine that is later evaluated by the expert, or
533 classified by the Chenin_FEM. As the output of the Chenin_FEM is a fuzzy value rather than
534 a class, we consider the classification correct if the fuzzy value is within 1.0 of the class assigned
535 by the expert, a classical sensory threshold used by the experts in this experiment. Using this
536 metric, the classification accuracy is 75% for predictions based on measured data, and 60% for

537 respectively predictions based data predicted by the Chenin_DBN model. This result is
 538 considered acceptable, as predictions in this uncertain context, trying to establish a global link
 539 between grape berries physicochemical measurements and wine quality, are challenging. Errors
 540 of classification based on measured data appear more often for Class 2 predictions. Those for
 541 classification based on the Chenin_DBN predictions appear more often for Class 1, except one
 542 for Class 3. This second class would probably need to be redefined and optimized, if more data
 543 were available. The results of classification using data predicted by the Chenin_DBN model
 544 are lower than the previous one, which can be explained by a cumulative effect of the error of
 545 the two models (Chenin_FEM and Chenin_DBN). It is particularly true for predictions of Class
 546 1.

547



548

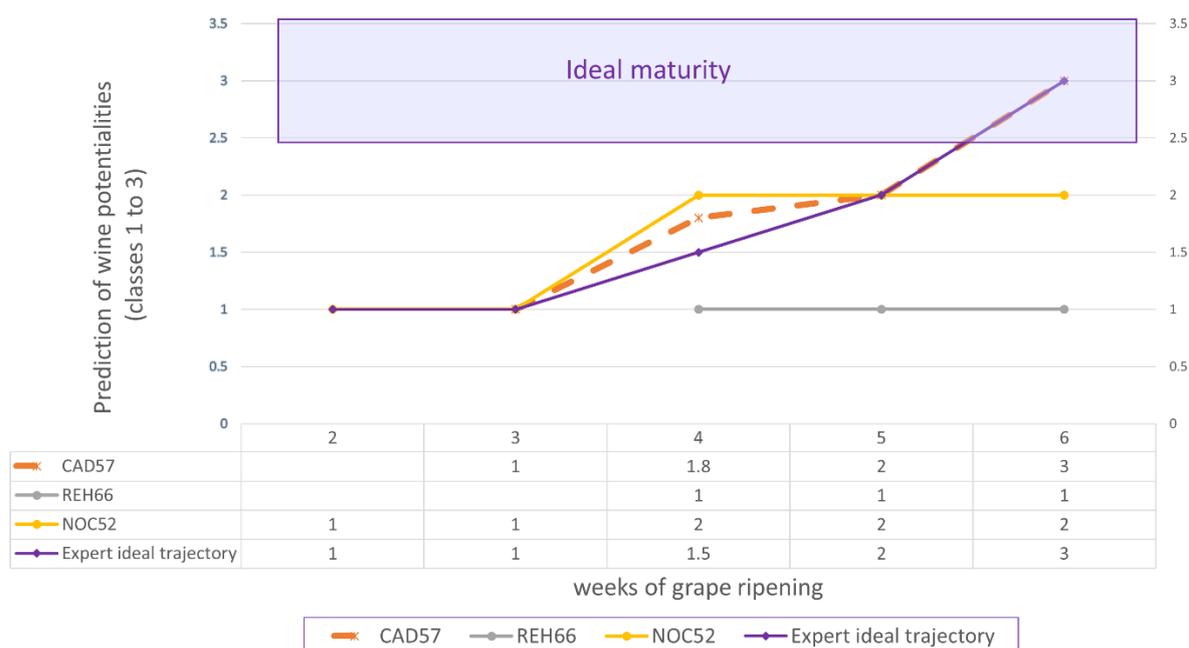
549 Figure 8: Predictions of the classes of wine for 20 samples by the Chenin_FEM. The expert
 550 evaluations are in blue, the classifications based on inputs measured on the samples are in
 551 orange, and the classifications based on physicochemical variable values predicted by the
 552 Chenin_DBN are in grey.

554 **4.3 Coupling the Chenin_DBN model to the Chenin_FEM: towards a decision support** 555 **system for wine quality prediction**

556 Coupling the two models has a strong interest for the winemakers, as together they can be used
557 to take a decision on the best date for harvesting grapes. It is also possible for them to intervene
558 on the vineyards in the earlier steps of ripening to correct the drift from the ideal trajectory.
559 Moreover, the experts' ability to predict the wine potential of the grapes allows them to manage
560 their action directly on the plants or during the fermentation, with a possible prediction of the
561 quality of the wine. Their reasoning in this uncertain space is achieved in terms of deviations
562 from an "ideal" trajectory that they have in mind. This type of reasoning has already been
563 observed in other traditional food production processes (Sicard et al., 2011). We have elicited
564 with the expert this "ideal trajectory" and have compared it to the trajectories of three samples
565 from the available database. Fixed fermentation process conditions were applied. The
566 trajectories of those three batches are shown in Figure 9, and compared to the ideal expert
567 trajectory. Thus, we can see that the deviation of the batch REH66 is growing all along the
568 weeks, with a maximum of a 2-class difference in week 6. It is important to notice that for this
569 batch, the first measurements were only available at week 4, but previous weeks are very likely
570 classified in Class 1. As predicted, after processing, the grapes of REH66 produced a wine that
571 was judged too acid and not aromatic enough by experts, so unsuitable to be a Chenin wine. In
572 this specific case, corrections could have been applied either before the step of grape ripening
573 or after, by mixing the grapes with other batches. For the batch NOC52, the trajectory is almost
574 the ideal one, but a drift is observed during the last week of the ripening. This was due to
575 unexpected extreme climatic conditions during the week. In this case, the corrective action is
576 generally performed during the fermentation, and the wine potential is recoverable: in this case,

577 the wine resulting from the NOC52 grapes was considered of acceptable quality. Lastly, for the
 578 batch CAD57, we can observe that this trajectory is quite the same as the one desired by the
 579 expert, and predictably, after fermentation has led to a wine having the ideal characteristics
 580 required by the experts for Chenin.

581



582

583 Figure 9: Ripening trajectories of 3 batches (labeled CAD57, REH66, NOC52) expressed under
 584 the form of a prediction by the fuzzy algorithm of classes of potential towards wine.
 585 Physicochemical measurements are obtained since the beginning of the maturation step.
 586 Predictions of the fuzzy set of expert rules are calculated on these measurements.
 587 Measurements and predictions start at week 4 for the batch REH66, because the ripening
 588 process of those vineyard plots was judged to be late. Week 6 is the one just before the harvest.
 589 The wines produced from these batches were also evaluated by an expert of Chenin. For CAD57
 590 and NOC52 the wine was evaluated to be at a good sensory quality type for a Chenin, better for
 591 CAD57 with more aromatic potentialities and a good equilibrium between aroma and acidity

592 than NOC52. For REH66, the wine was evaluated not to be at the level of quality attempted for
593 a Chenin, with too much acidity and not enough aromatic liveliness.

594

595

596 **Conclusions**

597 This study shows that it is possible and of valuable interest to propose computing tools able to
598 formalize and capitalize on the knowledge of a food domain, based on data as well as human
599 know-how and expertise. Even if for these complex processes the uncertainty is high, and
600 amount of data is relatively small to describe each ripening kinetics, this combination makes it
601 possible to develop relevant decision support systems based on artificial intelligence.

602 Interestingly, when compared to (Perrot et al., 2015), the accuracy of the predictive model for
603 acidity is noticeably lower. While the previous work was carried out on a different quality of
604 grapes, and considering less variables, the variability in weather for the data in the 2016-2017
605 period might also have played a role in the lower performance of the model presented.

606 Part of the difficulties in obtaining reliable predictive models is likely due to the growing impact
607 of climate change on local weather. Future works will investigate the possibility of integrating
608 climate-weather models into the proposed approach, in order to better take into account the
609 evolution of this global phenomenon.

610

611 **Acknowledgements**

612 We would like to thank all the experts and winegrowers of Tour and Angers who have
613 participated to this study. We also thank InterLoire, the Institut Français de la Vigne et du vin
614 – IFV Tours, the Institut Français de la Vigne et du vin – IFV Angers, the Chambre

615 d'Agriculture d'Indre et Loire – Groupement de Développement Viti-Vinicole (GDVV), the
616 Cellule Terroir Viticole, the INRA Angers – Unité Vigne et Vin, the ESA Angers – Laboratoire
617 GRAPPE.

618

619 **References**

620

621 Barbeau, G. B.-H. (2003). Comportement de quatre cépages rouges du Val de Loire en
622 fonction des variables climatiques. *J Int Sci Vigne Vin*, 38, 35-40.

623 Baudrit, C., Perrot, N., Brousset, J.-M., Guillemain, H., Perret, B., Picque, D., Barbeau, G.

624 (2015). A probabilistic graphical model for describing the grape berry maturity.

625 Computers and Electronics in Agriculture. *Computers and Electronics in Agriculture*,
626 118, 124-135.

627 Baudrit, C., Wuillemin, P.-H., & Perrot, N. (2013). Parameter elicitation in probabilistic

628 graphical models for modelling multi-scale food complex systems. *Journal of Food*
629 *Engineering*, 115(1), 1-10.

630 Ben Ghazlen, N., Moise, N., Latouche, G., Martinon, V., Mercier, L., Besancon, E., &

631 Cerovic, Z. (2010). Assessment of grapevine maturity using a new portable sensor:

632 Non-destructive quantification of anthocyanins. *Journal International des Sciences de*
633 *la Vigne et du Vin*, Vol. 44, pp. 1-8.

634 Boukhelifa, N., Tonda, A., Trelea I.-C., Perrot, N., & Lutton, E. (2017). Interactive

635 Knowledge Integration in Modelling for Food Sustainability: Challenges and
636 Prospects. *ACM CHI Workshop on Designing Sustainable Food Systems*.

637 Brousset, J. (2009). *Caractérisation multifactorielle et modélisation de la maturité de baies*

638 *de Cabernet Franc en moyenne vallée de Loire*. Rapport InterLoire.

639 Claverie, M., Prud'Homme, P., Mongendre, J., Zabollone, E., Raynal, M., Coulon, T., Forget,
640 D. (2008). Modélisation statistique de la qualité en viticulture par la méthode PLS
641 Spline. *VIIe Congrès International des terroirs viticoles*.

642 Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization
643 of continuous features. In *Machine learning proceedings 1995* (pp. 194-202).

644 Geiger D., Heckerman D. (1997) A characterization of the Dirichlet distribution through
645 global and local parameter independence. *The Annals of Statistics* 25: 1344–1369.

646 Geisser, S. (1993). *Predictive Inference*. New York: Chapman and Hall.

647 Geraudie, V., Roger, & J. M., O. H. (2010). Développement d'un appareil permettant de
648 prédire la maturité du raisin par spectroscopie proche infra-rouge(PIR). *Revue*
649 *Française d'Oenologie*, 240, 2-8.

650 Goelzer, A., Charnomordic, B., Colombié, S., Fromion, V., & Sablayrolles, J. (2009).
651 Simulation and optimization software for alcoholic fermentation in winemaking
652 conditions. *Food Control*, 20(7), 635-642.

653 Heckerman D. (1999) A Tutorial on Learning with Bayesian Networks. MIT Press,
654 Cambridge, MA, USA, 301–354.

655 Jensen Finn V. and Nielsen Thomas D.. (2010) *Bayesian Networks and Decision Graphs*,
656 Springer-Verlag. 464p.

657 Jensen, Jacob S., Demiray, S., Egebo, M., Meyer, A.S. (2008). Prediction of Wine Color
658 Attributes from the Phenolic Profiles of Red Grapes (*Vitis vinifera*). *J. Agric. Food*
659 *Chem.* 2008, 56, 3, 1105–1115. <https://doi.org/10.1021/jf072541e>

660 Moody, J, Blanton, J.E., Augustine M.A. (1996). Enhancing End-User Mental Models of
661 Computer Systems through the Use of Animation. *Proceedings of the 29th annual*

662 Hawaii International Conference on System Sciences.
663 <https://www.researchgate.net/publication/221181855>

664 Murphy, K. P. (2002). Dynamic bayesian networks: representation, inference and learning.
665 *PhD Dissertation*. Berkeley: University of California.

666 Niimi, J., Liland, K.H., Tomic, O., Jeffery D.W., Bastian S., Boss P.K. (2021). Prediction of
667 wine sensory properties using mid-infrared spectra of Cabernet Sauvignon and
668 Chardonnay grape berries and wines. *Food Chemistry* 344 (2021).

669 Pearl J. (1988). Probabilistic Reasoning in Intelligent systems: Networks of Plausible
670 Inference. Morgan Kaufmann, San Diego. 552p.

671 Perrot, N., Trelea, I., Baudrit, C., Trystram, G., & Bourguine, P. (2011). Modelling and
672 analysis of complex food systems: state of the art and new trends. *Trends in Food
673 Science & Technology*, 22(6), 304-314.

674 Perrot, N., Baudrit, C., Brousset, J. M., Abbal, P., Guillemin, H., Perret, B., & Picque, D.
675 (2015). A Decision support system coupling fuzzy logic and probabilistic graphical
676 approaches for the agri-food industry: prediction of grape berry maturity. *PloS one*,
677 10(7), e0134373.

678 Perrot, N., Ioannou, I., Allais, I., Curt, C., Hossenlopp, J., Trystram, G. (2006). Fuzzy
679 concepts applied to food product quality control:A review. *Fuzzy Sets and Systems*
680 157 special edition, 1145–115. doi:10.1016/j.fss.2005.12.013

681 Petropoulos, S., Karavas C.S., Balafoutis A.T., Paraskevopoulos, I., Kallithraka S.,
682 Kotseridis, Y. (2017). Fuzzy logic tool for wine quality classification. *Computers and
683 Electronics in agriculture* 142, 552-562.

684 Raynal, M., Debord, C., Guittard, S., & Vergnes, M. (2010). Epicure, a geographic
685 information decision support system risk assessment of downy and powdery mildew
686 epidemics in Bordeaux vineyards. *Sixth international workshop on the grapevine
687 downy and powdery mildew* (p. 144-146). Bordeaux: INRA-ISV.

688 Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM
689 algorithm. *SIAM review*, 26(2), 195-239.

690 Riou, C. (1994). *Le déterminisme climatique de la maturation du raisin: application au
691 zonage de la teneur en sucre dans la Communauté Européenne*. Office des
692 Publications Officielles des Communautés Européennes.

693 Sicard, M., Perrot, N., Leclerc-Perlat, M.N., Baudrit, C., Corrieu, G., 2011. Towards 703
694 integration of experts skills and instrumental data to control food processes. 704
695 Application to Camembert-type cheese ripening. *Journal of Dairy Science* 94 (1), 705
696 1–13.

697 Vermersh, P. (2006). *L'entretien d'explicitation*. Paris: ESF.
698

699 Van Leeuwen, C. (2010). Terroir: the effect of the physical environment on vine growth,
700 grape ripening and wine sensory attributes *Viticulture and Wine Quality*. « *Managing
701 wine quality* » *Woodhead Publishing Series in Food Science, Technology and
702 Nutrition. Chapter 9*, Pages 273-315. <https://doi.org/10.1533/9781845699284.3.273>
703

704 Zadeh , L. (1965). ‘ Fuzzy Sets .’ *Information and Control* 8 , 338 –353
705

