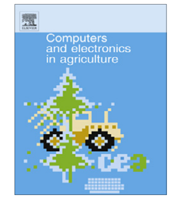
Computers and Electronics in Agriculture 118 (2015) 124–135

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



A probabilistic graphical model for describing the grape berry maturity

Cédric Baudrit [a,*], Nathalie Perrot [b], Jean Marie Brousset [b], Philippe Abbal [c], Hervé Guillemin [b], Bruno Perret [b], Etienne Goulet [d], Laurence Guerin [d], Gérard Barbeau [e], Daniel Picque [b]

[a] *Institut National de la Recherche Agronomique – Institut de Mécanique et d'Ingénierie, Talence, France*
[b] *Institut National de la Recherche Agronomique, Unité Génie et Microbiologie des Procédés Alimentaires, Thiverval-Grignon, France*
[c] *Institut National de la Recherche Agronomique, Unité Sciences Pour l'Œnologie, Montpellier, France*
[d] *Institut Français de la Vigne et du Vin, Unité de VINs, Innovations, Itinéraires, TERroirs et Acteurs, Amboise, France*
[e] *Institut National de la Recherche Agronomique, Unité Vigne et Vin, Beaucouzé, France*

ARTICLE INFO

Article history:
Received 2 March 2015
Received in revised form 17 August 2015
Accepted 17 August 2015
Available online 9 September 2015

Keywords:
Dynamic Bayesian networks
Knowledge integration
Uncertainty
Wine
Modeling

ABSTRACT

Grape berry maturation depends on complex and coupled physiological and biochemical reactions which are climate dependant. Moreover one experiment represents one year and the climate variability could not be covered exclusively by the experiments. Consequently, harvest mostly relies on expert prediction. A big challenge for the wine industry is nevertheless to be able to anticipate the reactions for sustainability purposes. We propose to implement a robust mathematical model able (1) to capitalize the heterogeneous fragmented available knowledge including data and expertise by means of probabilistic graphical approaches; and (2) to predict sugar, acidity and anthocyanin concentrations over the maturity.

1. Introduction

Grape berry quality is the result of complex physiological and biochemical reactions taking place during all the vine vegetative cycle and more particularly from veraison up to the harvest of grapes. Grape maturity can be characterized by several indices mainly berry size, grape color, concentration of total soluble solids, acidity, phenolic compounds, anthocyanin contents. These indices of maturity are very important to decide the harvesting date which influence the quality of the wines (Pérez-Magarino and Gonzales-San José, 2006; Champagnol, 1984; Huglin, 1978; Coombe and McCarthy, 2000). The notion of phenolic ripeness, a major enological interest, is difficult to characterize precisely. The criterion used may be considered as the percentage of the contribution of seed to optical density at 280 nm. It allows to differentiate the polyphenols from the skin (easily extractable) of those from seed (extracted mainly at the end of alcoholic fermentation). A good knowledge of the phenolic characteristics of the grape berries leads to a better control of winemaking by defining different parameters linked to it. Must oxygenation periods, temperature and fermentation time are scalable factors, adjusted annually to the quality of the berries. One can thus make the most of the grapes while remaining moderate in winemaking techniques, respecting its real phenolic potential. In all cases, a ripe grape is characterized by a skin rich in anthocyanins with tannins easily extractable and with a relatively small contribution of seeds tannins. The tannins of the seeds are extracted during the alcoholic phase. These molecules are essential for the structure of the wine and the stability of color. If their amount is too small, the wine lacks of structure and its color may change quickly. If the quantity is too high (above 60%), the wine may have a real astringency, hard to reduce during wine raising. The phenolic maturity of seed (SM) is the percentage of tannins from the seeds. This indicator is relevant to determine the optimum conditions of fermentation. Overall, the tannins from seeds are less qualitative than those from the skin; consequently, the higher the SM, the higher the proportion of tannins extracted from the skin and the longer the extraction during maceration can last. Red wines contain about ten times more tannin than white wines, about 1 to 3 grams per liter. Different kind of maturities may be distinguished:

- technological maturity where a ripe grape is characterized by a high ratio sugar/acidity. This index corresponds either to a maximum sugar content or to a limit for acidity;
- aromatic maturity where an optimum concentration of varietal aromas is obtained without disturbance by unwanted flavors;

* Corresponding author.
E-mail address: cedric.baudrit@u-bordeaux.fr (C. Baudrit).

http://dx.doi.org/10.1016/j.compag.2015.08.019
0168-1699/

Nomenclature

T	Mean temperature on a week, °C	CPT	total phenolic content, g/kg of berry
RH	Mean relative humidity of the air on a week, %	P200B	weight of 200 berries, kg
Ins	Insolation duration (hours of sunshine per week), h	IM	maturity index = sugar/total acidity, Unit-less
Pl	Rainfall on a week, mm	*t*	Time span for modeling, week
S	Sugar concentration, g/L	*P*	Probability measure, [0, 1]
Ac	Total acididity, g/L of H_2SO_4		
An	Anthocyanin content, g/kg of berry		

– phenolic maturity corresponding to the maturity of the skin (anthocyanins, tannins) and the maturity of the seeds (tannins). It represents the maximum concentration of "qualitative tannins";
– enological maturity, used by the winemaker, integrating all the previous maturities depending on the type of wine he wants to produce.

The indices of maturity are climate dependant and weather conditions affect their evolutions (Van Leeuwen et al., 2004; Barbeau et al., 2003). In this context, we propose a probabilistic model to (1) simulate the behavior and (2) calculate reliable predictions of berry composition according to the meteorological conditions. Air temperature, rain fall, relative humidity, sunshine hours are well known to affect the grape ripening mainly sugar concentration (Riou, 1994), total acidity (Barbeau et al., 2003) and anthocyanin level (Kobayashi et al., 2011). Nevertheless, the implementation of predictive model remains difficult as regards to this specific domain relying on the integration and coupling of heterogeneous knowledge (Perrot et al., 2011; Baudrit et al., 2013).

In the case of grape maturity, above the complexity of the reactions involved, several factors are to be emphasized: Data handling is time consuming and limited (one year for one experimental condition), available knowledge is fundamental to handle but expressed in different forms (equations, expert opinions, databases, ...), different formats (numeric, symbolic, linguistic, ...) and at different scales (microbiological, physicochemical, organoleptic, ...). In addition, the heterogeneous character of knowledge is widely tainted with uncertainty such as randomness, incompleteness, imprecision, vagueness (Dubois, 2007; Ferson and Ginzburg, 1996; Helton et al., 2004). Uncertainty may arise from randomness due to the natural variability of observations resulting from heterogeneity or the fluctuations of variable over the time. It may also be caused by imprecision due to a lack of information resulting from a partial lack of data, either because collecting this data is too difficult or costly, or because only experts can provide it.

First works dedicated to help the expert decision in viticulture focused on the construction of indicators able to aggregate and characterize climate conditions for grape production and maturity. Several indicators have been proposed in the literature (Tonietto and Carbonneau, 2004). The Huglin index (Huglin, 1978) or the "heliothermic" index (Branas et al., 1946) or the growing degree-day index of Winkler (Winkler, 1962) are widely used by the experts. The first one is for example directly correlated to the sugar concentration in the grape berry at a given date (correlation 0.86%). Nevertheless they remain very global and not predictive. Mathematical models (Gutierrez et al., 1985; Williams et al., 1985a,b; Walker et al., 2005; Nendel and Kersebaum, 2004) have also been developed, more often focusing on wine-growth and development. Their level of details is very interesting for understanding purposes about the vine but often too specific. Moreover, they cannot be directly used at a more macroscopic level for decision support designing about grape maturity. PLS (Partial least square) approaches are also proposed for example by Claverie et al. (2008) linking climate and maturity indexes. Nevertheless those approaches only rely on a data basis and its completeness as regards to the climatic conditions, which are not particularly stable at present. Their nature exclusively interpolative, limits their robustness of prediction. Fernandez Martinez et al. (2011,2012) also led works about the modeling of certain attributes as the alcohol level or the grape berry weight during maturation process by using learning approaches stemming from artificial intelligence. Nonetheless, none model proposes a global view or reflection about the process of grape berry maturity. Few mechanistic approaches are available, due to the difficulty of the task in part impacted by a knowledge fragmented at very separated scales: from a global perception of the experts to a local understanding of the plant growing and maintenance. Dai et al. (2009) propose a mechanistic model so called SUGAR describing the biochemical reactions taking place in the grape and their link to the sugar concentration in the grape. The model prediction is nevertheless poor in precision with a RMSE (root mean square error of the model versus experiments) of 30 g/l for the sugar concentration.

Despite the number of areas involved in viticulture research, available knowledge of the grape berry maturation remains fragmented and tainted with uncertainty. None of the approaches or investigations carried out up to now makes it possible to provide an explicit overview of the causal structure of associations between the underlying variables and an objective interpretation of the maturation. With this aim in mind, we focus on the use of the concept of dynamic Bayesian networks (DBNs) (Murphy, 2002) for an agri-food problem that provides a practical mathematical formalism that makes it possible to describe complex stochastic dynamical systems. They are an extension of Bayesian networks (BNs) (Jensen and Nielsen, 2010; Pearl, 1988) that rely on the probabilistic graphical models where the network structure provides an intuitively appealing interface by which humans can model highly-interacting sets of variables and provides a qualitative representation of knowledge. Uncertainty pertaining to the system is taken into account by quantifying dependence between variables in the form of conditional probabilities. The concept of DBNs makes it possible to combine different sources of expert knowledge with experimental data at different levels and scales of knowledge. This approach has been investigated recently in others domains (Baudrit et al., 2013).

This paper aims to elaborate a probabilistic modeling of the principle coupled dynamics of maturity indices (sugar, anthocyanin and total acidity concentrations) influenced by environmental climatic conditions by means of a dynamic Bayesian network. After a description of the material and methods, the next section is successively dedicated to (1) the results of maturation kinetics for three vintages in different locations of Loire Valley; and (2) the elaboration of mathematical model where predictive simulations are compared to experimental data. To finish, we conclude and introduce the possible extension of the predictive model for further works.