# PREDICTION OF VINE VIGOR AND PRECOCITY USING DATA AND KNOWLEDGE-BASED FUZZY INFERENCE SYSTEMS

Cécile COULON-LEROY[1]*, Brigitte CHARNOMORDIC[2], Dominique RIOUX[3], Marie THIOLLET-SCHOLTUS[1], and Serge GUILLAUME[4]

1:INRA UE1117 Vigne et Vin, UMT Vinitera, 49071 Beaucouzé, France
2:INRA, Supagro, UMR MISTEA, 34060 Montpellier, France
3:Cellule Terroirs Viticoles, UMT Vinitera, 49071 Beaucouzé, France
4: Cemagref, UMR ITAP, F-34196 Montpellier, France

## Abstract

**Aims**: The evolution of the economic and environmental context (low-input management practices, increase of energy cost and climate change) requires adaptation and/or optimization of winegrower's practices in order to elaborate competitive and yet still qualitative wines. To adapt and sustain their practices at the plot scale (*e.g.*, rootstock selection or plantation density), winegrowers and viticultural consultants need indicators to predict vine development based on permanent environmental factors (soil, parent rock and landscape). As of today, such indicators are either non-existent or too basic. The aim of this work is to develop operational and useful indicators based on strong scientific evidence.

**Methods and results**: This paper proposes a new approach based on a computer model composed of a cascade of fuzzy expert systems to estimate the two variables that best characterize vine development: vigor (VIG) and precocity (PRE). This model combines pedological expertise and data analysis. Based on scientific literature, and in particular on a previous expert system using analytical equations (Morlat *et al.*, 2001), the new approach allows a continuous estimation of VIG and PRE imparted by soil, parent rock and landscape. Further, it avoids the drawbacks of the previous expert system, due to the use of traditional crisp partitions for continuous input variables. Another novel aspect is the parameter setting, which efficiently combines expert knowledge and data mining. Finally, the method is tuned and validated against two different databases.

**Conclusion**: VIG and PRE imparted by environmental factors can now be evaluated more efficiently than with the former methods. The new method eliminates the need for post-evaluation correction by experts, which saves time. It also allows a continuous estimation of these variables. Each step can be controlled and analyzed during the design. Finally, the method is generic in the sense that the reasoning used to represent the relations between variables is not restricted to a given area. It can easily be customized and adapted to new areas by adjusting the parameters using local pedological knowledge and data.

**Significance and impact of the study**: This work answers the significant problem of VIG and PRE assessment according to environmental factors, which is a prerequisite in order to best adapt long-term cultural practices.

**Key words**: composite indicators, fuzzy inference system, soil, landscape, mesoclimate, pedoclimate, crop, imprecision, expert knowledge, k-means

## Résumé

**Objectifs** : L'évolution du contexte économique et environnemental (réduction des intrants, augmentation des coûts énergétiques et changement climatique) nécessite une adaptation et/ou une optimisation des pratiques des viticulteurs pour continuer à élaborer des vins de qualité tout en restant compétitif. Pour adapter leurs pratiques et plus particulièrement celles qui sont pérennes à l'échelle de la parcelle (par exemple, le choix du porte-greffe ou la densité de plantation), les viticulteurs et les conseillers viticoles ont besoin d'indicateurs pour prédire le développement de la vigne en fonction des facteurs environnementaux permanents (sol, roche-mère et environnement paysager). De tels indicateurs sont actuellement absents ou trop simplistes dans leur mode de construction. L'objectif de ce travail est de développer des indicateurs opérationnels et utiles avec une justification scientifique bien construite.

**Méthodes et résultats** : Cet article propose une nouvelle approche basée sur un modèle informatique composé de plusieurs systèmes experts flous. Il combine l'expertise pédologique et l'analyse de données pour estimer les deux variables qui caractérisent le mieux le développement de la vigne : la vigueur et la précocité. Grâce à une analyse bibliographique et en particulier à partir d'un précédent système expert basé sur une équation analytique (Morlat *et al.*, 2010), la nouvelle méthode permet d'obtenir une estimation continue de la vigueur et de la précocité conférées par les facteurs du sol, de la roche-mère et du paysage. Elle élimine les inconvénients du précédent système expert dus à un découpage traditionnel en classes des variables continues. Un autre aspect nouveau concerne l'obtention des paramètres du modèle en combinant efficacement la connaissance experte et l'analyse de données. La méthode est mise au point et validée à partir de deux bases de données.

**Conclusion** : La vigueur et la précocité conférées par les facteurs environnementaux peuvent désormais être évaluées par deux indicateurs qui fonctionnent mieux que ceux construits précédemment, évitant une correction *a posteriori* des valeurs des indicateurs par les experts. La nouvelle méthode permet une estimation continue de ces variables. Chaque étape peut être contrôlée et analysée lors de l'agrégation des variables. La méthode est générique car le raisonnement lié aux relations entre les variables n'est pas spécifique à une région. Elle peut facilement être transférée à de nouvelles zones en adaptant les valeurs des paramètres à partir de la connaissance experte et des données du nouveau terrain d'étude.

**Signification et impact de l'étude** : Ce travail apporte une réponse au problème de l'estimation de la vigueur et de la précocité conférées par les facteurs environnementaux permanents, qui est un pré-requis pour pouvoir mieux adapter les pratiques culturales pérennes.

**Mots clés** : indicateurs composites, système d'inférence floue, sol, paysage, mésoclimat, pédoclimat, culture, imprécision, connaissance experte, k-means

*manuscript received 7th November 2011 - revised manuscript received 4th June 2012*

*Corresponding author: c.coulon@groupe-esa.com



*J. Int. Sci. Vigne Vin*, 2012, **46**, n°3, 185-205

©Vigne et Vin Publications Internationales (Bordeaux, France)

## INTRODUCTION

Winegrowers must adapt their cultural practices to changing wine quality objectives and environmental factors in order to evolve towards sustainable viticulture, to remain competitive and to improve both production methods and wine quality. Indeed berry composition and wine characteristics depend on the joint effects of environmental factors and cultural practices. The key elements among these effects are (Carbonneau *et al.*, 2007): the level of photosynthetic source (estimated, for instance, by the "Exposed Leaf Area" model); the sinks corresponding to yield or quantity of grapes (particularly during ripening); the competitive sink represented by shoot vigor (particularly in summer with the growth of laterals); the general equilibrium between these elements as it determines the carbon balance of the plant (in that respect, it is remarkable to note that a moderate water limitation optimizes the carbon flux towards berries, and to some extent the level of secondary metabolisms); the general earliness of the phenological stages; and the duration of the growth cycle, which mostly depends on temperature, as it determines the time of the previous physiological functions on a time scale. Although all of these effects could be used to develop our approach, we chose to focus on vine vigor and the precocity of the vine vegetative cycle because they are of major interest in cool climate viticulture (Jackson and Lombard, 1993; Morlat, 2010). Let us first define those terms:

First, vine vigor (VIG) corresponds to the growth rate and intensity of the shoots (Carbonneau *et al.*, 2007). At the plant level, the 'vine vigor' term is commonly used, both in practice and in literature, to characterize the total vegetative biomass of the plant ('vegetative expression'), knowing that there is no direct correlation with shoot vigor: a vigorous vine, for instance, can have either weak or vigorous shoot growth depending on the number of shoots or buds produced (Carbonneau *et al.*, 2007). Vineyard management to control vine vigor determines the balance between vegetative growth and fruit production (Kliewer and Dokoozlian, 2005). This balance can be achieved through a rational management system involving vine shape (Reynolds and Heuvel, 2009) and interaction choices in the plantation, such as rootstock selection, planting density and canopy height (Dry and Loveys, 1998; Rives, 2000). The management strategy also depends on the environmental factors.

Second, precocity (PRE), also called earliness, depends on variables affecting vine growth. These are partly linked to water availability, which plays a major role in the early stages of the vegetative cycle (Tesic *et al.*, 2002; Carbonneau *et al.*, 2007). Early vine development means an earlier onset of the main phenological stages (budbreak, flowering and veraison). Depending on the degree of earliness, the optimal maturity of the grapes for a given type of wine will be reached earlier or later in the growing season. For example, in the northern vineyards of France, a late variety may not achieve optimal maturity, so harvest would have to start to avoid excessive rot of grapes.

Let us now describe the different ways of measuring or estimating VIG and PRE. This is not an easy task as they both depend on permanent factors such as environmental (soil, parent rock and landscape) and planting (*e.g.*, density, row orientation, etc) characteristics as well as non-permanent factors such as climate and annual practices (*e.g.*, ploughing, pruning, etc).

VIG and PRE can be estimated during the growth cycle by various measurements. VIG can be evaluated directly or indirectly by destructive or non destructive methods (Tregoat *et al.*, 2001): in precision viticulture, remote-sensing is a widely used technique (Drissi *et al*., 2009; Goutouly *et al.*, 2006; Homayouni *et al.*, 2008), along with leaf area or pruning weight measurements (Carbonneau *et al.*, 2007) and expert evaluations (Bodin and Morlat, 2003; Carey *et al.*, 2007). PRE can be directly observed in the vineyard and estimated according to indices combining the dates of flowering and veraison (Barbeau *et al.*, 1998), with some models allowing the prediction of phenological dates according to air temperature (Garcia de Cortazar Atauri *et al.*, 2009; Parker *et al.*, 2011). Concerning budbreak specifically, Pouget (1963) demonstrated that the time of budbreak (Baggiolini's stage B) is dependent on air temperature (and thus bud temperature) whatever the temperature of the other vine parts might be, particularly roots. But in the Loire Valley, Morlat and Hardy (1987) and Morlat (2010) found that the time of budbreak better correlates with soil temperature than with air temperature, leading to the conclusion that the temperature of the roots is responsible for budbreak. However, in these two studies, the authors have ignored the fact that buds, just like the soil, require some time to warm up when the air is on a warming trend. In this particular vineyard, these warming lag phases were similar, which explains the better correlation


©Vigne et Vin Publications Internationales (Bordeaux, France)

between budbreak and soil temperature rather than air temperature. Besides, these authors noted budbreak using Baggiolini's stage C, which may correspond more to the beginning of growth (for which root activity is determining) than to budbreak. Nevertheless, soil temperature is more important for the beginning of growth than for budbreak. Similarly, other authors showed the impact of root-zone temperature on budbreak (Woodham and Alexander, 1966; Kliewer, 1975; Zelleke and Kliewer, 1980; Tagliavini and Marangoni, 1992). Even though some measurements are available, they do not provide an *a priori* evaluation to ease the choice of long-term practices at the time of plantation, because they are only 'available' during the growth cycle.

To address the lack of *a priori* evaluations and to propose operational tools to best adapt cultural practices, models can be designed to estimate VIG and PRE. VIG and PRE variables can be useful for reasoning under realistic operating conditions. For example, if the objective is medium VIG, winegrowers can compensate a high VIG imparted by environmental factors by selecting a low vigor rootstock (*e.g.*, Riparia). Therefore, to build efficient decision support tools to help winegrowers and viticultural consultants to best adapt long-term cultural practices, it is important to design models capable of predicting VIG and PRE variables at the plot level by using a limited number of easy-to-assess input variables.

Meynard (2008) noticed the necessity to think in a systemic way by considering the interactions among practices themselves and with the environment. A systemic approach in viticulture can benefit from modeling. VIG and PRE modeling according to permanent factors such as soil, parent-rock and landscape characteristics contribute to the development of this systemic approach. The aim of our modeling approach is to provide operational tools to better evaluate the potentialities of plots according to permanent factors in order to optimize the cultural practices. It is not about proposing an exhaustive representation of the system involving all the physiological knowledge of vine development such as that proposed by Dai *et al.* (2010). We used a limited number of input variables and we aggregated these inputs using expert systems to predict the VIG and the PRE levels imparted by permanent environmental factors.

Numerous studies on the characterization and mapping of viticultural terroirs have been carried out during the last 20 years (Vaudour, 2003). VIG and PRE imparted by soil, parent rock and landscape are not directly measured but can be estimated through various mathematical algorithms that aggregate input variables obtained from viticultural terroir mapping (Morlat *et al.*, 2001). These VIG and PRE aggregated input variables are called 'composite variables' to distinguish them from the original input variables called 'raw variables'. We based our research work on these existing algorithms.

The algorithms by Morlat *et al.* (2001) use analytical equations that require continuous variables to be partitioned into crisp classes. In many cases, the estimations need to be re-evaluated by expert pedologists because of the sharp transition between classes. To avoid this problem, we propose a new strategy to characterize VIG and PRE imparted by environmental factors in which the analytical model of Morlat *et al.* (2001) has been replaced by a computer model involving data and knowledge-based fuzzy expert systems. As our method aims to improve the well documented mathematical method of Morlat *et al.* (2001), a comparison of the results obtained with the two different methods is included.

## MATERIALS AND METHODS

Relations between environmental factors and VIG and PRE (Bodin and Morlat, 2003; Morlat *et al.*, 2001) settle the expert knowledge on which the present method is based. First, we detail the basis and input values of the analytical algorithms described by Morlat *et al.* (2001). Then we briefly recall the principles of fuzzy logic and fuzzy inference systems (FIS). Next, we present the design of the FIS that replace the analytical algorithms. Finally, we describe the different validation steps.

### 1. Input variables in the former algorithms

Here we present in detail the input variables used to estimate VIG and PRE in the former algorithms (Morlat *et al.*, 2001).

VIG is estimated on the basis of three input variables: water holding capacity (WHC), gravel percentage in the soil profile (GOP) and parent rock hardness (PRH). The main variable that influences VIG is WHC: it impacts VIG twice as much as GOP and PRH. When characterizing a vineyard, WHC must be estimated at a large scale (vineyard or region) with sufficient precision at the plot scale to allow a good adaptation of the viticultural


©Vigne et Vin Publications Internationales (Bordeaux, France)

practices. There are many methods for measuring WHC, however, most of them are expensive and tend to be used at only a few sampling points (Acevedo-Opazo *et al.*, 2008). The alternative approach used to evaluate WHC is based on the Baize's equation (Baize, 2000) (Equation 1). It takes into account the pedological data gathered during viticultural terroir mapping. The Baize's equation uses the following input variables: number of horizons (h), soil horizon depth (HD), percentage of fine soil (FS), field capacity humidity (FCH), wilting point humidity (WPH) and bulk density (BD). FCH, WPH and BD are determined at each sampling point according to the texture of the soil horizons and the characteristics of the parent rock, as proposed by Goulet *et al.* (2004).

$$WHC = \sum_{h=1}^{h} [(FCH - WPH) \times FS \times HD \times BD]$$

**Equation 1 - Estimation of water holding capacity (WHC) according to Baize (2000). h: number of horizons, FCH: field capacity humidity (%), WPH : wilting point humidity (%), FS: fine soil (%), HD : horizon depth (dm) and BD: bulk density.**

HD and h are determined by the pedologist. HD of the last horizon can easily be determined when PRH is soft to hard. When PRH is crumbly and the soil is very deep, WHC is calculated for the whole rooting depth, which can be directly observed in pedological trenches. However, it must be noted that in some cases, even if the parent rock is hard, some roots can grow through the cracks of the rock and significantly contribute to the water supply. GOP is also directly observed in pedological trenches: it is estimated for each horizon, and the average GOP is calculated according to the depth of each horizon. PRH is determined according to the parent rock and deep horizon types and is classified into three classes: 3 for hard, 2 for soft and 1 for crumbly.

As mentioned by Piedallu *et al.* (2011), the horizon texture, HD and GOP are visually quantified, involving subjectivity in their estimation. Since more than one observer is often involved in the data survey, observers must agree on the same frame of reference. The assessments of FCH, WPH, FS and BD are performed in a laboratory; a source of uncertainty can be introduced by the disaggregation of the soil samples.

PRE is influenced by the following environmental factors: VIG, root profile, pedoclimate (PED) and mesoclimate (MES). As for the VIG estimation algorithm, the PRE imparted by environmental factors is estimated on the basis of nine input variables: GOP, HD, PRH, FCH, natural drainage (NAD), color of soil surface (CSS), maximum rooting density (MRD), altitude (ALT), landscape opening (LOP) and exposure (EXP). NAD is estimated according to Jamagne's scale (Jamagne, 1967) and CSS according to Munsell's Color Charts (Munsell Color Inc, 2000). Root profile and consequently MRD are directly determined according to the most influential soil type rather than the rootstock genotype (Smart *et al.*, 2006). The highest root density is generally found within one soil depth corresponding to the second and third soil horizons. However, some soils with sand in the first horizon and clay in the deep horizons present two depths of high root density (*i.e.*, superficial and deep): the deepest roots exploit the clayey sands, which have a higher water content (Morlat, 1993), and are related to the increase of the vine cycle duration, while the superficial roots established in sands induce early budbreak. As the algorithm estimates the precocity of budbreak, and since the precocity of the next phenological stages is very dependent on air temperature, we chose to attribute a superficial rooting depth to this type of soil (Barbeau *et al.*, 1998). LOP is determined according to the method of Jacquet and Morlat (1997), and EXP is determined according to the gradient and orientation of slopes.

All input variables are partitioned into crisp classes and their values are aggregated using analytical equations, which can be found in Morlat *et al.* (2001). In the following, this method will be referred to as 'MORLAT2001'.

In building the new strategy, we stand by the relations between environmental factors and VIG and PRE established by Morlat *et al.* (2001). These relations were validated by comparing viticultural terroir mapping and winegrower's surveys (Bodin and Morlat, 2006; Morlat, 2010; Morlat and Lebon, 1992).

## 2. New strategy

VIG is estimated according to WHC, GOP and PRH as with the previous algorithm.

To estimate PRE, we choose a different approach for the aggregation of the input variables. Firstly, we estimate PED and MES. Secondly, PRE is estimated according to four input variables: VIG, PED, MES and the maximum rooting density (MRD). This breakdown in the estimation process is suitable because MES can also be useful to adapt


©Vigne et Vin Publications Internationales (Bordeaux, France)

the pruning date, since cold mesoclimate situations have a higher risk of frost. The impact of ALT on temperature, and consequently on PRE, was noticed by several authors (Falcao *et al.*, 2010; Guyot, 1999; Lebon, 1993; Morlat, 2010): higher ALT is associated with lower air temperature and, consequently, decreased PRE. The opposite can also occur since atmospheric inversion may happen along hillsides. The down slope case is taken into consideration with the 'topography' (TOP) variable.

The data used to tune the new method come from the soil and landscape characteristics of 2353 sampling points in the Chinon vineyard of the middle Loire Valley region (France) and the observations of nine representative pedological trenches of soil diversity.

The proposed architecture (Figure 1) is a hierarchical structure with several linked modules, composed of four FIS. Firstly, VIG, MES and PED are evaluated according to three or four input variables and secondly, PRE is evaluated according to VIG, MES, PED and MRD. The four FIS are implemented using a computer model. In the following, the new method will be referred to as 'CFS' (Cascading Fuzzy Systems).

## 3. Introduction to fuzzy sets and fuzzy logic

Our 'CFS' method is based on fuzzy logic and FIS (Bouchon-Meunier and Marsala, 2003; Nguyen and Walker, 1996; Zadeh, 1965). The goal of this section is to provide the reader with some theoretical background on FIS and a brief description of fuzzy linguistic modeling.

FIS are one of the most famous applications of fuzzy logic and fuzzy set theory (Guillaume and Charnomordic, 2011 and 2012). They are able to handle linguistic concepts (*e.g.*, High or Low) implemented using fuzzy sets. Fuzzy logic is used as an interface between the linguistic space - the one of human reasoning - and the space of numerical computation.

A fuzzy set is defined by its membership function (MF). A point in the universe, x, belongs to a fuzzy set with a membership degree, where $0 \leq \mu_{L(x)} \leq 1$. If L is a set of Low soil water content,

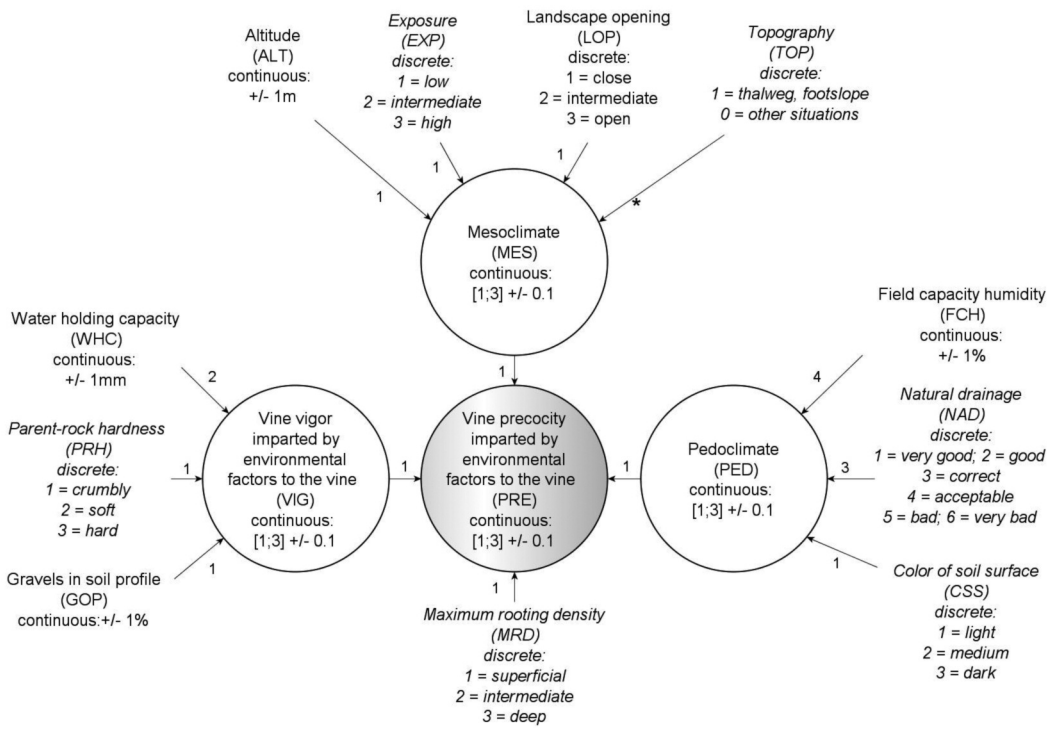



**Figure 1 - Architecture of the system with input variables to estimate the four output variables: vine vigor, thermal mesoclimate, thermal pedoclimate and vine precocity. The continuous or discrete characteristics of the input variables are indicated. Input variables are weighted (numbers next to arrows) and then aggregated by means of fuzzy inference systems. *if Topography is equal to 1 then MES is equal to 1 whatever the values of the other input variables, if Topography is equal to 0 then MES depends on the values of the other input variables.**


©Vigne et Vin Publications Internationales (Bordeaux, France)

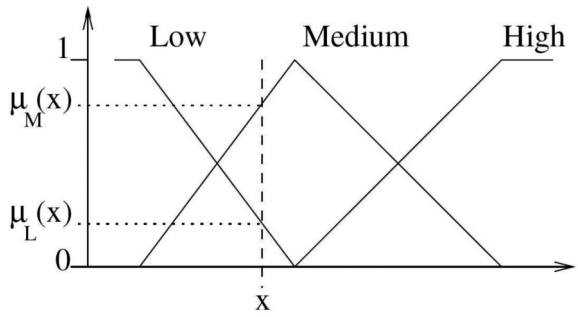



**Figure 2 - Example of three fuzzy sets defined on the same universe. 'x': a point of the universe, $\mu_L(x)$: the membership degree in the 'Low' fuzzy set, $\mu_M(x)$: the membership degree in the 'Medium' fuzzy set, and $\mu_H(x)$: the membership degree in the 'High' fuzzy set.**

the membership degree $\mu_{L(x)}$ of a given soil water content value x can be interpreted as the level up to which the x water content should be considered as Low. Several fuzzy sets (*e.g.*, Low, Medium and High) can be defined on the same universe, as illustrated in Figure 2.

As fuzzy sets usually overlap, a data point is likely to belong to more than one fuzzy set. In the partition shown in Figure 2, the value x belongs to both the Low and the Medium fuzzy set, with the corresponding membership degree $\mu_{L(x)}$ and $\mu_{M(x)}$. This allows to manage a smooth transition between concepts.

With such partitions, called strong fuzzy partitions, a given point may belong, with a non null degree, to at most two fuzzy sets. Moreover, for each point in the universe, the sum of the membership degrees of all the fuzzy sets of the partition is normalized to one. These partitions have been shown to have good semantic properties (Valente de Oliveira, 1999).

Fuzzy sets are used in FIS to build linguistic rules, for instance "*If WCH is High then ...*". In that single-input case, for a numerical value x of WHC, the rule matching degree, denoted by w, which means how true the rule is for the example, is given by the membership degree of x in the fuzzy set High, $\mu_{H(x)}$. Usually several input variables are involved in the rule premises. The *j*th rule in the FIS rule base is written as follows: *If* $x^1$ *is* $A^1_j$ *and* $x^2$ *is* $A^2_j$ *... and* $x^p$ *is* $A^p_j$ *then y is* $C^r$, where $A^k_j$ is the fuzzy set of the *k*th input variable used within the *j*th rule and $C^r$ is the rule conclusion. The rule conclusion being a scalar, this formulation corresponds to a *Sugeno* FIS (Takagi and Sugeno, 1985).

In the multi-input case, the rule matching degree is obtained by the conjunction of the premise elements, as follows: $W^j = \mu A_j^1(x^1) \wedge \mu A_j^2(x^2) \wedge \ldots \wedge \mu A_j^p(x^p)$, where $\mu A^k_j(x^k)$ is the membership degree of $x^k$ in the fuzzy set $A^k_j$, $\wedge$ being the 'minimum' conjunction operator. Because of the fuzzy set overlap, a given input is likely to fire several rules simultaneously. Consequently, all the fired rules will be involved in the system inference and the rule conclusions will be aggregated to give the final output. For a *Sugeno* FIS with m rules, the aggregation is performed using a weighted sum of the rule conclusions, the weights being the respective rule matching degrees (Equation 2).

$$y = \frac{\sum_{j=1}^{m} W^j C^j}{\sum_{j=1}^{m} W^j}$$

**Equation 2 - Calculation of a final output value y. m: number of rules, $W^j$: *j*th rule matching degree and $C^j$: *j*th rule conclusion.**

Fuzzy modeling allows the representation of continuous input information (*e.g.*, a WHC expressed in millimeters) in a lexical form (*e.g.*, 'low' or 'high' WHC) that can be integrated into an expert system. With this mathematical method, imprecision is taken into account through a progressive transition between input variable classes.

## 4. Fuzzy inference system design

The first step of FIS design deals with variable partitioning. In the present model, the number of linguistic concepts is set to two for each of the input variables, corresponding to the linguistic terms Low and High. This is the lowest possible number. It limits complexity but is still good in terms of expressiveness, as the fuzzy sets provide a smooth transition between concepts.

Two kinds of variables are considered: continuous and discrete variables. Discrete variable partitioning using fuzzy sets is possible because the discrete values, in the considered case, are ordered and have a progressive semantic meaning. The characteristic points of discrete inputs are not too difficult to set by experts given the limited number of possible values. The characteristic points of continuous inputs, $C_1$ and $C_2$, are not so easy to determine because of the infinite number of possible choices (Figure 3). So the one-dimensional *k-means* algorithm (MacQueen, 1967) was run on


©Vigne et Vin Publications Internationales (Bordeaux, France)

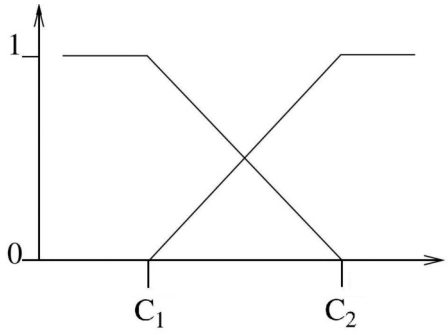


**Figure 3 - Fuzzy parameters: $C_1$ and $C_2$.**

the input data, independently for each variable. The two cluster centers, $C_1$ and $C_2$, were chosen as characteristic points.

The second step of FIS design is the choice of the rule base. For each composite variable, the rules are easily extracted from the former algorithms (Morlat *et al.*, 2001) and transposed to the fuzzy formalism. The set of decision rules cover all the situations that may occur. The rule conclusion is a scalar ranging from 1 to 3; it takes into account the weights of the various input variables.

FisPro 3.4 software was used to build the FIS (FisPro User Guide, 2010).

## 5. System behavior evaluation

The accurate transcription of the influence of environmental factors on VIG and PRE estimations needed to be confirmed. As expert evaluation by pedologists of some environmental factors raises the level of uncertainty, the effects of these factors also have to be assessed to evaluate the acceptable level of uncertainty. To answer these two points, it is necessary to study the system behavior. We chose to conduct a parameter sensitivity analysis and to analyze the output response to input variations.

The study of the sensitivity to the partition parameters ($C_1$ and $C_2$, Figure 3) of the continuous variables aims to relate the output variations and the uncertainties that can exist on data values. A dataset of input values was built considering all combinations of the following items: ten distinct values of WHC and GOP, and three distinct input values of PRH. For WHC as well as for GOP, the ten distinct values included a value lower than $C_1$, a value higher than $C_2$, $C_1$, $C_2$ (membership degree in the corresponding MF equal to 1) and six values equally distributed between $C_1$ and $C_2$ (membership degree in both MFs less than 1). To test the sensitivity to parameters, inferred output values, for each composite variable, obtained for the original value of $C_1$ or $C_2$, were compared with those inferred for parameters values, modified by more or less 20%.

The output variation in response to input variable changes was studied to check the system behavior. Output value variation was analyzed with respect to a given input, while the others were set to a fixed value. Ten values of the studied input (the same as those for the parameter sensitivity test) and three given values of the fixed inputs were considered: the value allowing to predict the lower value of the composite variable, the value allowing to predict the higher value of the composite variable, and the value corresponding to the intersection of the two MFs ($(C_1 + C_2) / 2$).

Sensitivity to parameters and output response analysis constitute a system behavior evaluation.

## 6. Validation

The VIG and PRE indicators that are used in this study are well documented and validated by Morlat *et al.* (2001) and Goulet and Morlat (2011). These composite variables cannot be directly validated by comparison with measured data on plots from previous years because they only evaluate the impact of permanent factors such as soil, parent rock and landscape characteristics. VIG, PED and PRE also depend on cultural practices and annual factors. For instance, observations have shown that the PRE variable is also dependent on annual climatic factors and pruning date (Martin and Dunn, 2000). Similarly, VIG is also dependent on the rootstock (Ollat *et al.*, 2003). Since we changed the mathematical method to calculate VIG and PRE, we needed to validate the indicators again. VIG and PRE can not be validated directly by comparison with observed data on the vineyard due to the influence of the climate and of the annual practices. Therefore, instead of a complete direct validation, we conducted a three-step validation:

- In a first step, we thoroughly investigated the system behavior presented in the previous Section 5. To that end, we first had to consolidate the design of the new method to make sure the relations between variables are respected.

- In a second step, we compared the model outputs with winegrower's evaluations on plots with similar practices but different soil, parent rock and landscape factors. This step was conducted using a dataset collected in the Chinon area of the


©Vigne et Vin Publications Internationales (Bordeaux, France)

middle Loire Valley (France). This dataset corresponds to the characterization of 2353 sampling points and 756 vine plots. Winegrowers were surveyed to provide an evaluation of the VIG and precocity of budbreak levels, the soil temperature and the frost risk of each plot. The evaluation can be correlated to the four outputs of our model. In the Chinon area, the majority of the vine plots are intercropped with a grass cover. This situation generates a competition for water and nitrogen resources and consequently impacts vine development (Celette *et al.*, 2005). The rootstock planted on plots confer different levels of VIG and PRE (Institut Français de la Vigne et du Vin, 2007). To compare model outputs with winegrowers' evaluations, we selected non-intercropped plots planted with a rootstock conferring medium VIG and PRE levels and pruned on the same month (January). The selection led to the extraction of 19 plots from the dataset. These plots are planted on 15 contrasted Terroir Basic Units (TBU); a TBU is defined by three associated components: a geological component, a pedological component and a landscape component (Morlat *et al.*, 2001). The environmental factor variability appears to be sufficient to test our model. We evaluated the model predictions using the root mean square error (RMSE) (Equation 3), by comparing predicted model values with observed values by winegrowers.

$$\mathrm{RMSE} = \sqrt{\frac{1}{A}\sum_{i=1}^{A}(\hat{y}_i - y_i)^2}$$

**Equation 3 - Root mean square error (RMSE). A: number of plots, $\hat{y}_i$: predicted value and $y_i$: observed value.**

- In a third step, the new CFS-based method was tested on a dataset from a completely different area to assess its generalization capability and what adaptations needed to be considered. The dataset originates from a vineyard located in the Douro region (Portugal). The numerical quality was assessed by comparing the predicted outputs with the MORLAT2001 outputs. The transferability of the method to end-users was also examined, as well as its user-friendly characteristics.

## RESULTS AND DISCUSSION

The CFS computer model design and tuning using the French Chinon dataset presented in Materials and Methods is detailed here. It includes the fuzzy partitioning of the input variables and the rule base composition. Then the system behavior is studied and the model outputs compared with expert evaluations. Finally, the transferability to a new dataset and to new users is tested by comparing the MORLAT2001 and CFS methods using a dataset from the Douro region (Portugal).

### 1. Fuzzy partitioning of the input variables

As stated in Materials and Methods (Section 3), the number of linguistic concepts for each of the input variables is set to two, corresponding to the linguistic terms Low and High. In practice, discrete variables may of course have more than two modalities. For instance PRH, EXP, LOP, CSS and MRD can take three different values, and the middle one will be considered as both Low and High with a 0.5 membership degree in the corresponding fuzzy sets. The NAD variable has a whole range of modalities between 1 and 6. The corresponding fuzzy set partition is designed considering that NAD is completely Low for values lower than 2 and completely High for values higher than 5 (Jamagne scale; Jamagne, 1967).

These partitions were built according to general expert knowledge and may not be suited to the specific data of the area under study. For instance, medium is the most important modality (1933 sampling points) for the CSS variable, with only 130 and 290 sampling points having a high and low value, respectively. This unbalanced distribution reflects the homogeneity of the area under study. Other areas might have a more balanced distribution of CSS.

The two limits $C_1$ and $C_2$ (see Figure 3) for the four continuous variables (WHC, ALT, GOP and FCH) are found by the *k-means* algorithm. As they are computed from a specific dataset, one can wonder about their meaning. For example, Figure 4 (top left) shows that WHC is completely low for values lower than 134 mm. Zufferey and Murisier (2006) observed that plots with a WHC below 100 mm are sensitive to water stress, so a high membership degree in the first fuzzy set means that, according to the local climatic conditions, a water stress can be observed.

The data used for calibrating the model have a wide range of variation. This is illustrated by the histograms of the distributions of each input variable, with the class intervals determined according to Sturges' method (Sturges, 1926). Histograms are shown in Figure 4, together with the corresponding fuzzy partitions.


©Vigne et Vin Publications Internationales (Bordeaux, France)

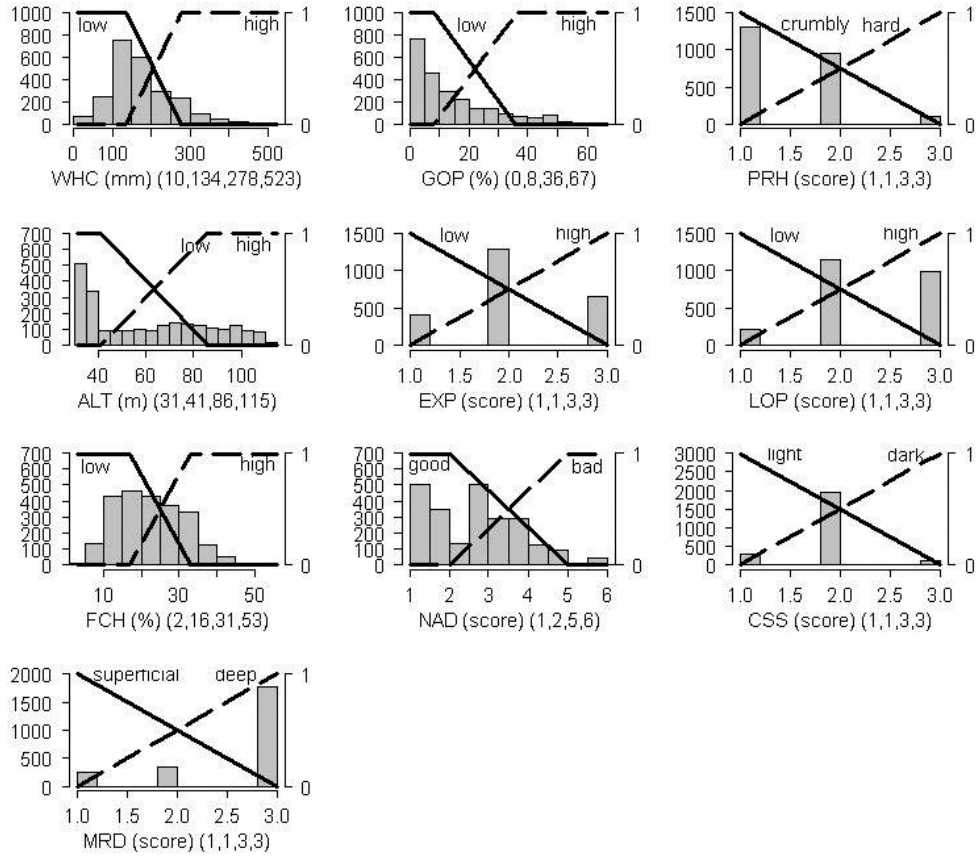



**Figure 4 - Histograms showing the distributions of input variables and their fuzzy partitioning for the Chinon (France) dataset. The four values in parentheses correspond to (minimum, first parameter $C_1$, second parameter $C_2$, maximum). The class intervals are determined according to Sturges' method (Sturges, 1926). WHC: water holding capacity, GOP: gravel in the soil profile, PRH: parent rock hardness, ALT: altitude, EXP: exposure, LOP: landscape opening, FCH: field capacity humidity, NAD: natural drainage, CSS: color of soil surface and MRD: maximum rooting density.**

The partitioning characteristics of our approach are summarized in the following remarks: i. Discrete variables are partitioned by expert knowledge. Depending on local environmental characteristics, some classes are not well represented. ii. Conversely, continuous variables need to be interpreted and validated by expertise as they are automatically partitioned using data mining.

## 2. Generation of decision rules

The expert rules for the composite variable assessment (VIG, PED, MES and PRE) are shown in Tables 1 to 4. As expected, the WHC input variable impacts VIG estimation twice as much as the other inputs (Table 1). These combinations lead to five distinct rule conclusions for VIG.

For PED estimation, the three input variables have distinct weights that lead to seven distinct rule conclusions (Table 2). FCH is the most influential variable. Combinations with high FCH and bad NAD (bad drainage) lead to two low rule conclusions (1 and 1.3), while those with low FCH and good NAD lead to two high rule conclusions (2.8 and 3).

The particular part played by TOP in MES estimation appears in the rules given in Table 3. If TOP is a thalweg or a foot slope (TOP 1), rule 1's conclusion is 1, whatever the value of the other input variables. If TOP is neither a thalweg nor a foot slope (TOP 2, rules 2 to 9), the three other input variables (ALT, EXP and LOP) have the same weight, leading to four distinct rule conclusions for MES.

As shown in Table 4, the four input variables of PRE have the same weight, leading to five distinct rule conclusions equally distributed between 1 and 3.


©Vigne et Vin Publications Internationales (Bordeaux, France)

**Table 1 - Decision rules for vine vigor. WHC: water holding capacity, GOP: gravel percentage in the soil profile, PRH: parent rock hardness and VIG: vine vigor evaluation.**

| Rules | if WHC | and GOP | and PRH | then VIG |
|---|---|---|---|---|
| 1 | low | high | hard | 1 |
| 2 | low | low | hard | 1,5 |
| 3 | low | high | crumbly | 1,5 |
| 4 | low | low | crumbly | 2 |
| 5 | high | high | hard | 2 |
| 6 | high | low | hard | 2,5 |
| 7 | high | high | crumbly | 2,5 |
| 8 | high | low | crumbly | 3 |

**Table 2 - Decision rules for pedoclimate. FCH: field capacity humidity, NAD: natural drainage, CSS: color of soil surface and PED: pedoclimate evaluation.**

| Rules | if FCH | and NAD | and CSS | then PED |
|---|---|---|---|---|
| 1 | high | bad | light | 1 |
| 2 | high | bad | dark | 1,3 |
| 3 | high | good | light | 1,8 |
| 4 | high | good | dark | 2 |
| 5 | low | bad | light | 2 |
| 6 | low | bad | dark | 2,3 |
| 7 | low | good | light | 2,8 |
| 8 | low | good | dark | 3 |

**Table 3 - Decision rules for mesoclimate. TOP: topography, ALT: altitude, EXP: exposure, LOP: landscape opening, MES: mesoclimate evaluation, Top 1: thalweg or foot slope and Top 2: other topography.**

| | | | | | |
|---|---|---|---|---|---|
| 1 | Top 1 | high or low | high or low | open or close | 1 |
| 2 | Top 2 | high | low | close | 1 |
| 3 | Top 2 | high | high | close | 1,7 |
| 4 | Top 2 | low | low | close | 1,7 |
| 5 | Top 2 | high | low | open | 1,7 |
| 6 | Top 2 | low | low | open | 2,3 |
| 7 | Top 2 | high | high | open | 2,3 |
| 8 | Top 2 | low | high | close | 2,3 |
| 9 | Top 2 | low | high | open | 3 |


©Vigne et Vin Publications Internationales (Bordeaux, France)

Reasoning and linguistic rules are generic. For example, a low WHC imparts a low VIG whatever the studied area. The expert knowledge is expressed in natural language and can thus be transcribed accurately by these rules as they are also expressed in natural language. The weights given to the various variables are reflected in the rule conclusions. The rule base is complete, meaning that the decision rules cover all the possible situations.

## 3. Fuzzy system behavior

The computer model is based upon expert rules. The parameters used to partition input variables are determined by expertise and data analysis. As previously underlined, an important step in the design of the computer model is to check its representativeness and evaluate its behavior when confronted with real-world data. The dataset from Chinon (France) is used for that purpose.

The output range of each composite variable (1 to 3) is well represented in the vineyard except for PRE with 1.4 as minimum and 2.6 as maximum. All the output range of PRE was predicted on other Loire vineyards. The four output distributions are presented in Figure 5. The different VIG values are found in similar proportions. Few sampling plots have a low MES value (*i.e.*, a cold mesoclimate), corresponding to the specific topography of thalweg or foot slope. There is a higher proportion of sampling points with a high PED value (*i.e.*, a hot pedoclimate), characterized by a low FCH (between 2 and 16%) and a good drainage (1 or 2) corresponding to a hard chalky parent rock of the Turonian geological stage from the cretaceous period, frequently observed in the considered area. The majority of the sampling points yields low to medium PRE estimations and the distribution of PRE is quasi-Gaussian. The range of VIG, MES, PED and PRE observed over the same area highlights the necessity to adapt some viticultural practices, like planting a low vigor rootstock (*e.g.*, Riparia) while targeting medium VIG to compensate for the high VIG imparted by the environmental factors (VIG value close to 3).

**Table 4 - Decision rules for vine precocity. VIG: vine vigor, PED: pedoclimate, MES: mesoclimate, MRD: maximum rooting depth and PRE: vine precocity evaluation.**

| Rules | if VIG | and PED | and MES | and MRD | then PRE |
|---|---|---|---|---|---|
| 1 | high | cold | cold | deep | 1 |
| 2 | low | cold | cold | deep | 1.5 |
| 3 | high | cold | cold | superficial | 1.5 |
| 4 | high | cold | hot | deep | 1.5 |
| 5 | high | hot | cold | deep | 1.5 |
| 6 | low | cold | cold | superficial | 2 |
| 7 | low | cold | hot | deep | 2 |
| 8 | low | hot | cold | deep | 2 |
| 9 | high | cold | hot | superficial | 2 |
| 10 | high | hot | cold | superficial | 2 |
| 11 | high | hot | hot | deep | 2 |
| 12 | low | cold | hot | superficial | 2.5 |
| 13 | low | hot | cold | superficial | 2.5 |
| 14 | low | hot | hot | deep | 2.5 |
| 15 | high | hot | hot | superficial | 2.5 |
| 16 | low | hot | hot | superficial | 3 |


©Vigne et Vin Publications Internationales (Bordeaux, France)

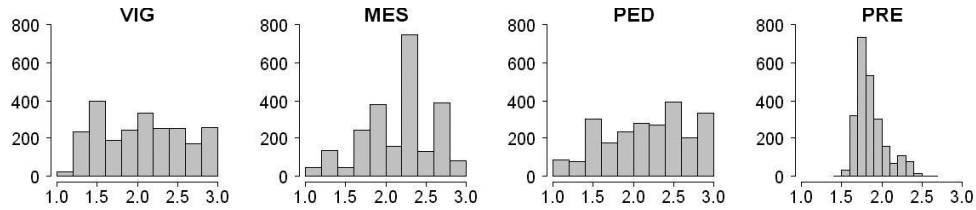



**Figure 5 - Histograms showing the distribution of output variables.**
**VIG: vine vigor, MES: mesoclimate, PED: pedoclimate and PRE: vine precocity.**
**The class intervals are determined according to Sturges' method (Sturges, 1926).**

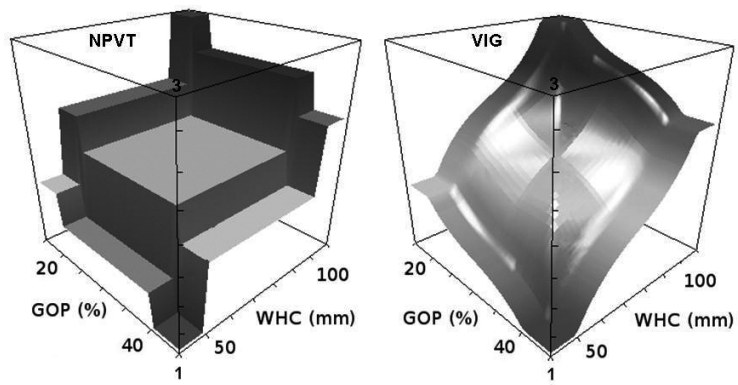



**Figure 6 - Comparison of vine vigor estimation according to the former MORLAT2001 method, variable 'NPVT' (left) and the new CFS method (right), variable 'VIG'. VIG values are generated for combinations of WHC: water holding capacity and GOP: gravel percentage in the soil profile and for intermediate parent rock hardness.**

Thanks to fuzzy logic modeling, the new CFS method allows a smooth response function *(*Figure 6). VIG is plotted as a function of WHC and GOP when PRH is set to the intermediate value. For an easier comparison, $C_1$ and $C_2$ are the ones found in MORLAT2001: 50 and 100 mm for WHC and 20 and 40% for GOP. The threshold effect leads to a wide plateau and sharp transitions for MORLAT2001 (left part of the figure). These drawbacks are avoided in CFS (right part of the figure).

Let us consider two plots (Plot 1 and Plot 2) characterized by the same PRH type (intermediate hardness) and the same GOP percentage (10%), with different WHCs (Plot 1: 59 mm and Plot 2: 99 mm). If we use the MORLAT2001 method to estimate their VIG value, the two plots would have the same VIG value (2.25, only 9 values being possible between 1 and 3 with a precision of +/- 0.25), despite their different WHCs. They would need to be reclassified together by expertise, with the value for Plot 2 reset to 2.75, to be more realistic. With the CFS model, the inferred VIG values for Plots 1 and 2 are respectively 2.0 and 2.8, which is more satisfactory.

This illustrates the fact that the new method avoids the expert reclassification of plots having values very close to the class limits. All estimation procedures of VIG and PRE imparted by the vineyard site can now be completely automated.

## 4. Sensitivity to parameter analysis

As previously explained in Materials and Methods (Section 5), the sensitivity to MF parameters was studied for the four continuous variables and each of the three FIS. We used a synthetic dataset of input values, including all combinations of values that cover all the WHC and GOP domain, combined with the three distinct PRH input values.

For each continuous input variable, the $C_1$ and $C_2$ parameters vary one after the other by +/- 20% (Table 5).

Table 6 shows that a 20% variation of input variable parameters leads to an important variation of output values, especially for the WHC parameters (0.4 on a scale ranging from 1 to 3). When one of the two parameters decreases by 20%, VIG decreases by 0.4, corresponding to 20% of its


©Vigne et Vin Publications Internationales (Bordeaux, France)

**Table 5 - Values of the fuzzy parameters for the continuous input variables used in the sensitivity analysis (WHC: water holding capacity, GOP: gravel percentage in the soil profile, FCH: field capacity humidity and ALT: altitude).**

| Input variables | C1 | | C2 | |
|---|---|---|---|---|
| | -20% | +20% | -20% | +20% |
| WHC (mm) | 74 | 112 | 150 | 224 |
| GOP (%) | 7 | 11 | 35 | 53 |
| FCH (%) | 14 | 20 | 26 | 40 |
| ALT (m) | 42 | 62 | 63 | 95 |

**Table 6 - Output value variation of VIG: vine vigor, PED: pedoclimate and MES: mesoclimate estimation when the fuzzy parameters $C_1$ and $C_2$ of the continuous input variables (WHC: water holding capacity, GOP: gravel percentage in the soil profile, FCH: field capacity humidity and ALT: altitude) vary by approximately 20%.**

| Output composite variables | Input variables | C1 | | C2 | |
|---|---|---|---|---|---|
| | | -20% | +20% | -20% | +20% |
| VIG (1-3 scale) | WHC | 0,4 | -0,4 | 0,4 | -0,4 |
| | GOP | -0,2 | 0,2 | -0,2 | 0,2 |
| PED (1-3 scale) | FCH | -0,3 | 0,2 | -0,2 | 0,3 |
| MES (1-3 scale) | ALT | -0,2 | 0,1 | -0,1 | 0,2 |

domain of variation. When GOP parameters undergo the same variation, VIG varies by approximately 10%. Variation of FCH parameters leads to PED variation of approximately 15%. MES varies by approximately 10% when ALT parameters vary.

Figure 7 illustrates, for each output variable, the individual impact of each input variable when the others are set to a given value. Let us take an example and detail the impact of GOP on VIG when the other two inputs, WHC and PRH, are given fixed values (Figure 7: top left plot, light grey curves). Three given combinations of input values are considered for WHC and PRH: i. Values set to predict the lowest VIG value, *i.e.*, WHC < 134 mm and PRH = '3-hard', ii. Values set to predict the highest VIG value, *i.e.*, WHC > 278 mm and PRH = '1-crumbly' and iii. Values set to predict intermediate VIG values that correspond to the intersections of the two fuzzy sets for each input, *i.e.*, WHC = 206 mm and PRH = '2-soft'. The impact of GOP on the estimated VIG output is plotted for each of these three combinations. Relations between the inputs and the output are monotonic but not linear, as shown by the intermediate curve. This is an important characteristic of fuzzy set modeling.

If we now make a similar study to assess the impact of PRH on VIG, we observe that the response curves are identical, this being due to GOP and PRH having the same weight in the rule base. However, this is no longer true for the response curve of VIG to WHC (Figure 7, top left plot, black curves). Indeed, the more influence the input variable has on the output estimation, the more important the amplitude of the curve is: for example, the amplitude of WHC is 1 while that of GOP or PRH is 0.5.

This representation also allows to visually check the relation between variables: for example, VIG increases when WHC increases, and it decreases when GOP increases. Similar interpretations can be made about the impact of FCH, NAD and CSS on PED estimation, ALT, EXP and LOP on MES estimation, and VIG estimation, PED, MES and MRD on PRE estimation. All of the corresponding plots are shown in Figure 7.


©Vigne et Vin Publications Internationales (Bordeaux, France)

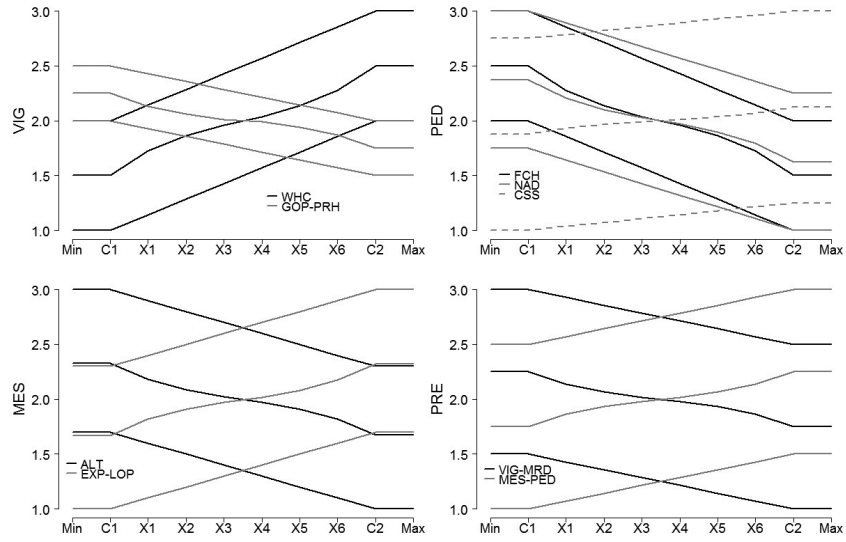



**Figure 7 - Individual impact of each input variable (WHC: water holding capacity, GOP: gravel percentage in the soil profile, PRH: parent rock hardness, FCH: field capacity humidity, NAD: natural drainage, CSS: color of soil surface, ALT: altitude, EXP: exposure, LOP: landscape opening and MRD: maximum rooting depth) on output values (VIG: vine vigor, MES: mesoclimate, PED: pedoclimate and PRE: vine precocity). min: minimum and $C_1$ and $C_2$: fuzzy parameters. The three curves correspond to three given combinations of given input values: value set to predict the lower output value (the lower curve), the higher one (the higher curve) and values that correspond to the intersections of the two membership functions.**

The computer model response and its representation therefore allow the specialist to assess the individual effect of each input variable.

## 5. Comparison of output values with expert evaluations

We compared model outputs and winegrower evaluations on 19 plots with similar practices. Four variables (VIG level, PRE level, frost risk and soil temperature) were evaluated on a five-point scale: 1 'very low', 1.5 'low', 2 'medium', 2.5 'high' and 3 'very high'. They were respectively compared to the predictions given by the VIG, PRE, MES and PED model outputs. The 19 plots correspond to 15 BTU. Diversity of soil, parent rock and landscape factors is well represented except the 'very bad drainage' modality (Table 7).

Scatter plots revealing the correlations between the predicted and observed values of the four outputs are shown in Figure 8. RMSE values vary between 0.4 and 0.6 (Table 8), while the output range is [1-3]. These RMSE values quantify the goodness-of-fit and should be discussed while keeping in mind the uncertainty existing on expert evaluation, which we estimate to be approximately 0.5. VIG is the most accurately predicted variable among the four considered ones, with a determination coefficient equal to 0.53. Winegrowers can easily assess VIG, as this variable is familiar to them, and they naturally use a wide range of variation, which may be the explanation for the relatively good prediction. In the absence of an absolute referential, experts used the whole range of assessments, ranging from 1 to 3, but on a limited set of conditions that form a reduced subset of the possible ones. This explains why the model outputs are limited to the [1.6-2.3]. The range of variation of PRE is not entirely represented as it was not possible to validate the prediction of extreme values (close to 1 or 3). All that leads to a very low determination coefficient (0.05). Concerning MES output, some plots (numbers 8, 9 and 10) are characterized by a medium to hot MES, however, winegrowers indicated a risk of frost that can be explained by local cold airstreams.

The low correlation between the composite variables and the winegrowers' estimations suggests that some influential factors, such as the previously mentioned impact of specific local climatic conditions or the fertilisation practice, may not have been taken into account. Winegrowers adjust the amount of fertilizers to avoid a deficiency but in some cases it cannot be optimized with regard to soil nutrient content. In our study, we have selected plots with homogeneous practices. All the plots were known to be pruned in the same

**Table 7 - Characteristics of the soil, parent rock and landscape factors of the 19 vine plots used to evaluate model predictions. WHC: water holding capacity (mm), GOP: gravel in the soil profile (%), PRH: parent rock hardness (score between 1 and 3), ALT: altitude (m), EXP: exposure (score between 1 and 3), LOP: landscape opening (score between 1 and 3), FCH: field capacity humidity (%), NAD: natural drainage (score between 1 and 6), CSS: color of soil surface (score between 1 and 3) and MRD: maximum rooting depth (score between 1 and 3).**

| Plots | Basic Terroir Unit | WHC | GOP | PRH | ALT | EXP | LOP | FCH | NAD | CSS | MRD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Stony old alluvium | 84 | 42 | 2 | 38 | 2 | 2 | 9 | 2 | 2 | 3 |
| 2 | Weakly weathered rock | 110 | 10 | 3 | 72 | 1 | 3 | 12 | 1 | 2 | 2 |
| 3a | Hydromorphic young alluvium | 146 | 50 | 1 | 33 | 2 | 1 | 16 | 4 | 2 | 3 |
| 3b | Hydromorphic young alluvium | 155 | 3 | 1 | 33 | 2 | 3 | 15 | 3 | 2 | 3 |
| 4 | Sandy and silty old alluvium | 180 | 10 | 1 | 77 | 3 | 3 | 16 | 2 | 1 | 3 |
| 5a | Sandy old alluvium | 154 | 1 | 1 | 33 | 2 | 2 | 15 | 3 | 2 | 3 |
| 5b | Sandy old alluvium | 125 | 12 | 1 | 40 | 2 | 3 | 13 | 2 | 2 | 3 |
| 5c | Sandy old alluvium | 125 | 12 | 1 | 40 | 2 | 3 | 13 | 2 | 2 | 3 |
| 6 | Calcareous colluvium | 160 | 15 | 2 | 60 | 3 | 2 | 22 | 2,5 | 2 | 3 |
| 7 | Chalky and sandstony weakly weathered rock | 91 | 14 | 2 | 56 | 2 | 3 | 19 | 1,5 | 2 | 1 |
| 8 | Chalky and silty weakly weathered rock | 144 | 2 | 2 | 52 | 3 | 2 | 35 | 1 | 2 | 2 |
| 9 | Chalky and silty moderately weathered rock | 201 | 10 | 1 | 39 | 2 | 3 | 26 | 2 | 3 | 3 |
| 10 | Sandy moderately weathered rock | 237 | 9 | 2 | 103 | 2 | 3 | 17 | 2,5 | 1 | 3 |
| 11 | Clayey old alluvium | 278 | 2 | 1 | 73 | 3 | 3 | 26 | 3,5 | 2 | 3 |
| 12a | Moderately weathered rock | 273 | 2 | 1 | 69 | 2 | 3 | 26 | 2 | 2 | 3 |
| 12b | Moderately weathered rock | 132 | 15 | 1 | 69 | 2 | 3 | 18 | 1 | 2 | 3 |
| 13 | Sandy and clayey old alluvium | 269 | 3 | 1 | 74 | 2 | 3 | 31 | 3,5 | 2 | 3 |
| 14 | Sandy and clayey strongly weathered rock | 214 | 4 | 1 | 92 | 1 | 1 | 24 | 4,5 | 2 | 3 |
| 15 | Clayey and greensandy moderately weathered rock | 191 | 14 | 2 | 35 | 2 | 2 | 34 | 4 | 2 | 3 |



*J. Int. Sci. Vigne Vin*, 2012, **46**, n°3, 185-205

©Vigne et Vin Publications Internationales (Bordeaux, France)

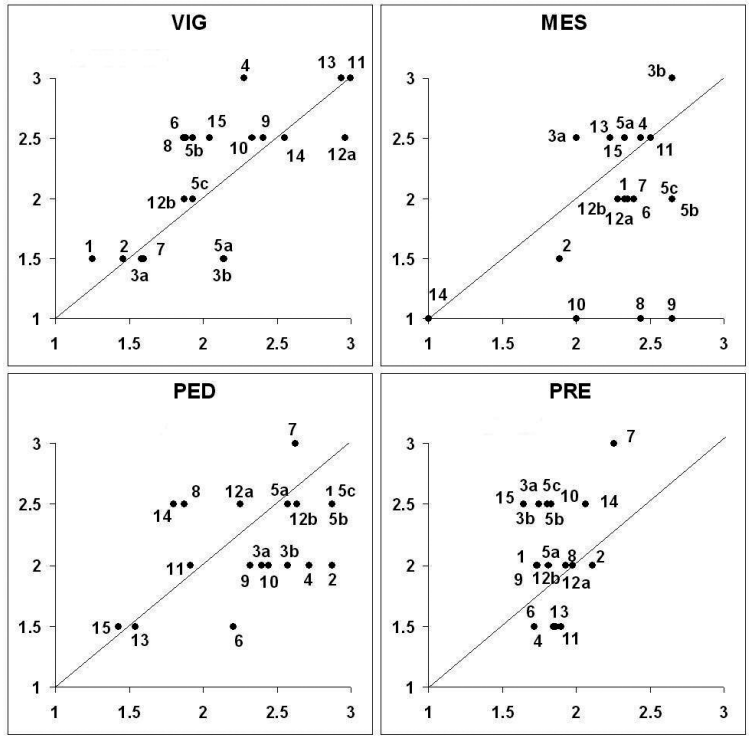



**Figure 8 - Comparison of values predicted by the model (in abscissa) and observed by winegrowers (in ordinate) for vine vigor (VIG) level, mesoclimate (MES), PED (pedoclimate) and precocity (PRE) level, for 19 plots corresponding to 15 Basic Terroir Units.**

month and a more precise pruning period was not recorded in the database. Martin and Dunn (2000) observed that the budbreak of a vine pruned six weeks later than another one is delayed by 5.3 days and this delay persists at anthesis (5.0 days) and veraison (4.1 days). In comparison, the highest delay according to PED mentioned by Morlat and Hardy is about 9 days. The difference of PRE due to variation of pruning date could be detected by winegrowers and might explain the poor correlation for the composite variable that estimate PRE. In the future, the pruning date should more precisely recorded.

We used the observed values given by the experts (*i.e.*, five possible values) as references to compare predicted and observed values. Future work should also consider the uncertainty on the reference expert values in order to define a more suitable index than the classical RMSE and $R^2$ values.

## 6. Comparison between the former and the new methods using a new database

The transferability to new datasets and to end-users was tested comparing the two method: MORLAT2001 and the new CFS computer model.

The data used to compare the methods and evaluate the transferability of CFS come from a new dataset of 267 sampling points and 17 augers located in the Douro vineyard (Portugal). As previously mentioned, the reasoning and linguistic rules given in Tables 1 to 4 are generic, so the rule base can be reused as it is. The input variables are the same ones and the weights of all input variables are preserved as well. An adaptation was made regarding EXP: according to local experts, the west exposure is considered as an intermediate EXP in the Chinon area and as a high EXP in the Douro area. The fuzzy set $C_1$ and $C_2$ parameters of the four continuous variables WHC, GOP, FCH and ALT were adapted as well (Figure 9). The crisp boundaries of MORLAT2001 were determined using expert knowledge. The results show that crisp boundaries and fuzzy parameters are very different. In MORLAT2001, crisp boundaries were suited to local characteristics. In the vineyard of Anjou in the middle Loire Valley (France) where Morlat *et al.* (2001) developed their method, the crisp boundaries for WHC were 50 and 100 mm, while in the vineyard of Chinon and Douro they were 100 and 150 mm. Using *k-means* analysis, fuzzy set parameters are higher for WHC (144 and 290 mm),

**Table 8 - Root mean square error (RMSE) and coefficient of determination ($R^2$) obtained from the model validation.**

| | VIG | MES | PED | PRE |
|---|---|---|---|---|
| RMSE | 0.4 | 0.6 | 0.5 | 0.5 |
| $R^2$ | 0.53 | 0.16 | 0.24 | 0.05 |

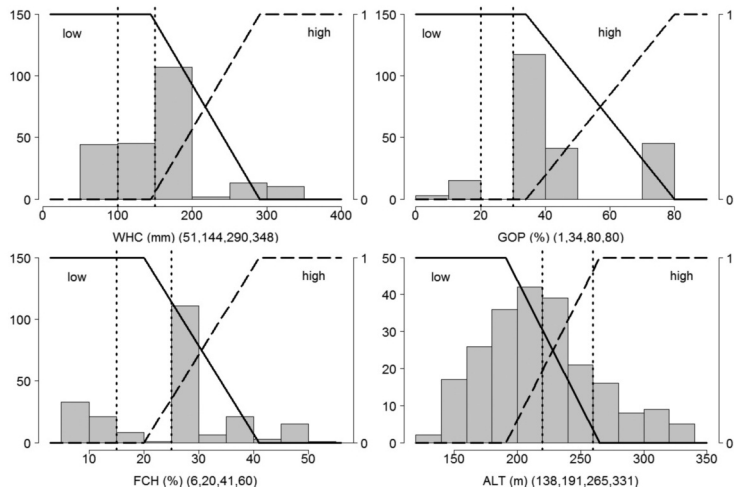



**Figure 9 - Histograms showing the distributions of input variables and their fuzzy partitioning on the Douro (Portugal) dataset. Values in parentheses correspond to (minimum, first parameter $C_1$, second parameter $C_2$, maximum). The class intervals are determined according to Sturges' method (Sturges, 1926). WHC: water holding capacity, GOP: gravel in the soil profile, FCH: field capacity humidity and ALT: altitude. Dotted vertical lines correspond to crisp boundaries of the former method.**

because of high WHC values for some sampling points.

End-users highlighted the advantage of an automated data mining technique to determine specific parameters from data but confirmed that they must be validated by pedologists and agronomists.

The results of three estimation methods for PRE are plotted as maps in Figure 10. Map (a) is computed using MORLAT2001, map (b) is map (a) corrected by pedologists and winegrowers' expertise from the 'Cellule Terroirs Viticoles' and map (c) is computed using CFS.

The lowest estimated value of MORLAT2001 (20) corresponds to 1 in CFS and the highest (60) to 3. Four classes related to four linguistic terms are considered: low, low to normal, normal to high and high.

At first glance, MORLAT2001 was not able to estimate Low PRE. Let us discuss two different situations, labeled Situation 1 and 2 on map (c), where low PRE is predicted with CFS.

Situation 1 is characterized by pediments with deep and silty soil inducing a high VIG level (~ 2.5), a cold PED (~ 1.2) and an intermediate MES (2.3). The entire area was assigned a "low to normal" or "normal to high" PRE level by MORLAT2001. Pedologists found the new map more appropriate. Situation 2 evaluation should also be assigned a low PRE, according to pedologists, given the specific BTU of this area (altered schist with high clay content and an important deep runoff before budbreak) that leads to cold MES and PED. MORLAT2001 did not correctly estimate PRE in this kind of situation.

To summarize, the computer model parameters may be easily adapted according to the local environment. Statistical analysis may help to propose local parameters; the effort of pedologists and researchers is thus reduced to the validation of the new values. But it could also make sense to compare areas or regions using the same set of


©Vigne et Vin Publications Internationales (Bordeaux, France)

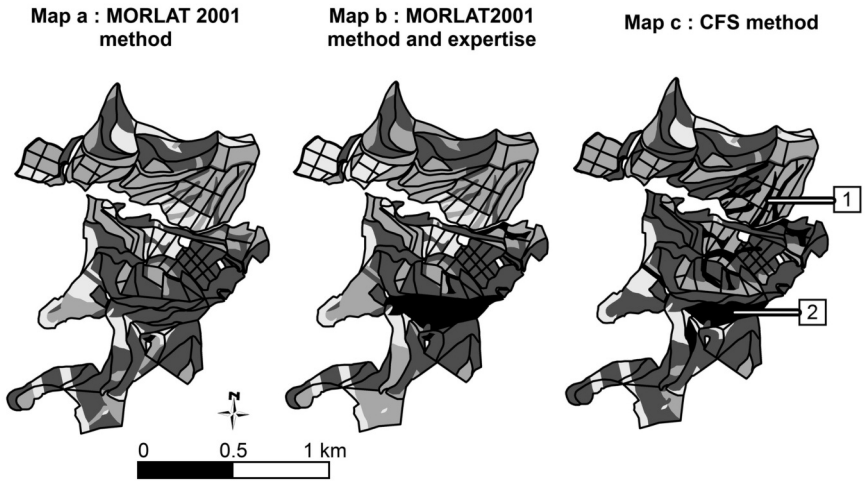



**Figure 10 - Precocity estimation for a vineyard located in the Douro area (Portugal). Map a: built using the former MORLAT2001 method (Morlat *et al.*, 2001), Map b: built by the 'Cellule Terroirs Viticoles' based on MORLAT2001 method plus pedologist and winegrower's expertise and Map c: built using the new CFS method without any expert modification. The output values of MORLAT2001 vary between 20 and 60; the same amplitude intervals are used to compare the two methods. Vine plots are represented in black. Situation 1: pediments with deep and silty soil and Situation 2: altered schist with high clay content and an important deep water movement before budbreak. Situations 1 and 2 are low precocity area undetected by MORLAT2001 method.**

parameters. The method is fast to implement and proved satisfactory when used by end-users.

## 7. Discussion on model validation

The composite variables built in this study can be given an indicator status. Gras *et al.* (1989) defined an indicator as "a variable which supplies information on other variables which are difficult to access (...) and can be used as bench marker to take a decision". Mitchell *et al.* (1995) considered indicators as ''alternative measures (. . .) that enable us to gain an understanding of a complex system (...) so that effective management decisions can be taken that lead towards initial objectives". VIG and PRE of a plot may constitute useful indicators to select grape variety, rootstock and cultural practices. MES may help to adapt the pruning date; a cold plot can be pruned later to minimize frost risk.

Bockstaller and Girardin (2003) highlight that indicators are used to assess complex processes that often do not have quantitative equivalents. They mention that many indicators cannot be validated in the same way as simulated models by comparison with measured data. They propose a three-stage approach: design validation, output validation and end-use validation. Cloquell-Ballester *et al.* (2006) consider a similar approach to validate environmental and social indicators: a self-validation stage carried out by the 'working team' itself to ensure the correct performance and to guarantee the proper documentation of indicators; a scientific validation stage integrating independent experts' judgments and a social validation stage integrating stakeholder's opinion.

We improved the method of Morlat *et al.* (2001) by using cascading fuzzy expert systems. Figure 6 and Table 8 show that output variables are now more precisely estimated. A partial validation of our indicators predicting VIG and PRE levels according to permanent environmental factors has been carried out using plots located on the same area and with similar practices. According to Bockstaller and Girardin (2003) and Cloquell-Ballester *et al.* (2006) recommendations, we concentrated our efforts on a deep analysis of the system behavior.

The parameters of the new system were validated on a real-world dataset. Our new method using fuzzy logic allows taking into account the imprecision on input variables, since a progressive transition between input classes is possible. Nevertheless, sharp parameters need to be determined. The model behavior, studied according to the parameter values, is satisfactory and confirms the efficiency of the new design. The relations between variables are well respected and could be correlated to winegrowers' observations. However,

for input variables that are directly observed by specialists, it is suggested to use a common frame of reference in order to limit the heterogeneity of observations among experts for a given studied area.

## CONCLUSION

This paper proposes a FIS-based computer model as a new approach to estimate VIG, PED, MES and PRE imparted by soil, parent rock and landscape. Once VIG, PED, MES and PRE are evaluated, they constitute useful indicators to better adapt viticultural practices.

Fuzzy logic allows transparency, and each step of the construction of the model can be controlled and analyzed. It allows a continuous estimation of the four output variables in only one step. It avoids a second step of reclassification through expertise to correct output values due to the problem of the sharp transition between input variable classes. This is achieved because the model handles linguistic terms implemented by using fuzzy sets that manage the progressiveness of the class change and because it efficiently combines expert knowledge and data mining.

Input variables were determined according to measurements or expert evaluations. The latter data acquisition involves subjectivity. Even if FIS allow to better manage uncertainty on input variables, this highlights the necessity to check the homogeneity among expert's evaluations and within the studied area before using them as input variables.

The method proposed in this paper is generic because reasoning and linguistic rules are not area-specific and accurately transcribe expert knowledge concerning relations between raw variables and VIG and PRE.

Using the new implementation, sharp parameters have been determined. These parameters are easily adapted to local characteristics combining data mining and expert validation.

The method has been tested in different areas and has been validated by end-users. Comparison of output values with measured data would seem ideal but data representing a wide range of situations were not available. A partial validation of the predicted composite variables has been carried out by comparing the model estimations with winegrower's evaluations and the model limits have been discussed.

FIS allow to consider uncertainty on data sources thanks to a progressive transition between input variable classes. However, the comparison of the system output with an expert or sensory assessment would necessitate more methodological developments to deal with the output uncertainty, and this is an attractive perspective.

Another interesting perspective consists in studying the aggregation of these indicators with viticultural practices (*e.g.*, VIG and PRE imparted by rootstock). It will permit to optimize the choice of long-term practices in relation to environmental factors and targeted vine development. It is likely to limit corrective practices, reducing production costs while promoting the production of targeted quality products and maximizing the potential value of a terroir.

**Acknowledgments**: The data used to build the method presented in this paper were collected by the 'Cellule Terroirs Viticoles', supported by ONIVINS-VINIFLHOR (now called FranceAgriMer), Pays de la Loire Region and Interloire (France). C. Coulon-Leroy received a fellowship from the 'INRA-SAD' department and Pays de la Loire Region (France). We are also grateful to Sogrape Vinhos S.A. for providing access to their data and to G. Barbeau, C. Bockstaller, V. Courtin, R. Morlat and S. Sicard for valuable comments on this work.

## REFERENCES

Acevedo-Opazo C., Tisseyre B., Ojeda H., Ortega-Farias S. and Guillaume S., 2008. Is it possible to assess the spatial variability of vine water status? *J. Int. Sci. Vigne Vin*, **42**(4): 203-219.

Baize D., 2000. *Guide des analyses en pédologie*. Éd. INRA, Paris.

Barbeau G., Asselin C. and Morlat R., 1998. Estimation du potential viticole des terroirs en Val de Loire selon un indice de précocité du cycle de la vigne. *Bull. OIV*, **71**(805/806): 247-262.

Bockstaller C. and Girardin P., 2003. How to validate environmental indicators. *Agric. Systems*, **76**(2): 639-653.

Bodin F. and Morlat R., 2003. Characterizing a vine terroir by combining a pedological field model and a survey of the vine growers in the Anjou Region (France). *J. Int. Sci. Vigne Vin*, **37**(4): 199-211.

Bodin F. and Morlat R., 2006. Characterization of viticultural terroirs using a simple field model based on soil depth I. Validation of the water supply regime, phenology and vine vigour, in the Anjou vineyard (France). *Plant Soil*, **281**(1/2): 37-54.

Bouchon-Meunier B. and Marsala C., 2003. *Logique floue, principes, aide à la décision.* d. Hermes Sciences Publications, Paris.


©Vigne et Vin Publications Internationales (Bordeaux, France)

Carbonneau A., Deloire A. and Jaillard B., 2007. *La vigne. Physiologie, terroir, culture*. Éd. Dunod, Paris.

Carey V., Archer E., Barbeau G. and Saayman D., 2007. The use of local knowledge relating to vineyard performance to identify viticultural terroirs in Stellenbosch and surrounds. *In: Proceedings of the International Workshop on Advances in Grapevine and Wine Research*, Venosa (Italy), 2005. *Acta Horticulturae (ISHS)*, 754: 385-392.

Celette F., Wery J., Chantelot E., Celette J. and Gary C., 2005. Belowground interactions in a vine (*Vitis vinifera* L.)-tall fescue (*Festuca arundinacea* Shreb.) intercropping system: water relations and growth. *Plant and Soil*, **276**(1): 205-217.

Cloquell-Ballester V.-A., Cloquell-Ballester V.-A., Monterde-Díaz R. and Santamarina-Siurana M.-C., 2006. Indicators validation for the improvement of environmental and social impact quantitative assessment. *Environmental Impact Assessment Review*, **26**(1): 79-105.

Dai Z.W., Vivin P., Barrieu F., Ollat N. and Delrot S., 2010. Physiological and modelling approaches to understand water and carbon fluxes during grape berry growth and quality development: a review. *Aust. J. Grape Wine Res.*, **16**(S1): 70-85.

Drissi R., Goutouly J.-P., Forget D. and Gaudillère J.-P., 2009. Nondestructive measurement of grapevine leaf area by ground normalized difference vegetation index. *Agronomy J.*, **101**(1): 226-231.

Dry P.R. and B.R. Loveys. 1998. Factors influencing grapevine vigour and the potential for control with partial rootzone drying. *Aust. J. Grape Wine Res.,* **4**:140-148.

Falcao L.D., Burin V.M., Sidinei Chaves E., Vieira H.J., Brighenti E., Rosier J.-P. and Bordignon-Luiz M.T., 2010. Vineyard altitude and mesoclimate influences on the phenology and maturation of Cabernet-Sauvignon grapes from Santa Catarina State. *J. Int. Sci. Vigne Vin*, **44**(3): 135-150.

FisPro User Guide, 2010. FisPro: An open source portable software for fuzzy inference systems. http://www.inra.fr/mia/M/fispro/FisPro_EN_doc.htm.

Garcia de Cortazar Atauri I., Brisson N. and Gaudillère J.P., 2009. Performance of several models for predicting budburst date of grapevine (*Vitis vinifera* L.). *Int. J. Biometeorology*, **53**(4): 317-326.

Goulet E., Morlat R., Rioux D. and Cesbron S., 2004. A calculation method of available soil water content: application to viticultural terroirs mapping of the Loire Valley. *J. Int. Sci. Vigne Vin*, **38**(4): 231-235.

Goulet E. and Morlat R., 2011. The use of surveys among wine growers in vineyards of the middle-Loire Valley (France), in relation to terroir studies. *Land Use Policy*, **28**(4): 770-782.

Goutouly J.-P., Drissi R., Forget D. and Gaudillère J.-P., 2006. Caractérisation de la vigueur de la vigne par l'indice NDVI mesuré au sol. *In*: *VI*[e] *Congrès Int. des Terroirs Viticoles*, Bordeaux-Montpellier (France).

Gras R., Benoit M., Deffontaines J.-P. Duru M., Lafarge M., Langlet A. and Osty P.-L., 1989. *Le fait technique en agronomie*. Éd. L'Harmattan, INRA, Paris.

Guillaume S. and Charnomordic B., 2011. Learning interpretable fuzzy inference systems with FisPro. *Inform. Sci.*, **181**(20): 4409-4427.

Guillaume S. and Charnomordic B., 2012. Fuzzy inference systems: An integrated modeling environment for collaboration between expert knowledge and data using FisPro. *Expert Systems with Applications*, **39**(10): 8744-8755.

Guyot G., 1999. *Climatologie de l'environnement*. Ed. Dunod, Paris.

Homayouni S., Germain C., Lavialle O., Grenier G., Goutouly J.P., Van Leeuwen C. and Da Costa J.P., 2008. Abundance weighting for improved vegetation mapping in row crops: application to vineyard vigour monitoring. *Can. J. Remote Sensing*, **34**(S2): S228-S239.

Institut Français de la Vigne et du Vin, 2007. *Catalogue des variétés et clones de vigne cultivés en France*, 2[e] éd, Ed. IFV.

Jackson D.I. and Lombard P.B., 1993. Environmental and management practices affecting grape composition and wine quality - A review. *Am. J. Enol. Vitic.*, **44**(4): 409-430.

Jacquet A. and Morlat R., 1997. Caractérisation de la variabilité climatique des terroirs viticoles en val de Loire. Influence du paysage et des facteurs physiques du milieu. *Agronomie*, **17**(9/10): 465-480.

Jamagne M., 1967. *Bases et techniques d'une cartographie des sols*. Annales Agronomiques, Vol. 18. Éd. INRA.

Kliewer W.M., 1975. Effect of root temperature on budbreak, shoot growth, and fruit-set of Cabernet Sauvignon grapevines. *Am. J. Enol. Vitic.*, **26**(2): 82-89.

Kliewer W.M. and N.K. Dokoozlian, 2005. Leaf area/crop weight ratios of grapevines: influence on fruit composition and wine quality. *Am. J. Enol. Vitic.,* **56**:170-181.

Lebon E., 1993. De l'influence des facteurs pédo et mésoclimatiques sur le comportement de la vigne et les caractéristiques du raisin. *Thèse de Doctorat*, Université de Dijon (France).

MacQueen J., 1967. Some methods for classification and analysis of multivariate observations. *In: Proceedings of the 5th Berkeley Symp. on Mathematical Statistics and Probability*, University of California Press, Berkley, pp 281-297.

Martin S.R. and Dunn G.M., 2000. Effect of pruning time and hydrogen cyanamide on budburst and subsequent phenology of *Vitis vinifera* L. variety Cabernet Sauvignon in central Victoria. *Aust. J. Grape Wine Res.*, **6**(1): 31-39.


©Vigne et Vin Publications Internationales (Bordeaux, France)

Meynard J.M., 2008. *Produire autrement : réinventer les systèmes de culture.* Éd. Educagri, Dijon.

Mitchell G., May A. and McDonald A., 1995. PICABUE: a methodological framework for the development of indicators of sustainable development. *International J. Sustainable Development World Ecology*, **2**(2): 104-123.

Morlat R. and Hardy P., 1987. Résultats concernant les variations de précocité de la vigne dans le Val de Loire. Importance du pédoclimat thermique. *In: Actes du 3e Symp. Int. sur la physiologie de la vigne*, Bordeaux (France).

Morlat R. and Lebon E., 1992. Une expérience des essais multilocaux au travers de l'étude des terroirs viticoles. *Progrès Agric. Vitic.*, **109**(3): 55-58.

Morlat R., 1993. The soil effects on the grapevine root system in several vineyards of the Loire Valley (France). *Vitis*, **32**(1): 35-42.

Morlat R., Guilbault P., Rioux D., Asselin C. and Barbeau G., 2001. *Terroirs viticoles : étude et valorisation*. Éd. Œnoplurimedia, Avenir Œnologie, Paris.

Morlat R., 2010. *Traité de viticulture de terroir.* Éd. Tec et Doc Lavoisier, Paris.

Munsell Color Inc, 2000. *Munsell soil color charts*, GretagMacbeth, New Windsor, NY, USA.

Nguyen H.T. and Walker. E.A., 1996. *A first course in fuzzy logic*. Ed. CRC-Press.

Ollat N., Tandonnet J.P., Bordenave L., Decroocq S., Gény L., Gaudillère J.P., Fouquet R., Barrieu F. and Hamdi S., 2003. La vigueur conférée par le porte-greffe : hypothèses et pistes de recherches. *Bull. OIV*, **76**(869/870): 581-595.

Parker A.K., De Cortazar-Atauri I.G., Van Leeuwen C. and Chuine I., 2011. General phenological model to characterise the timing of flowering and veraison of *Vitis vinifera* L. *Aust. J. Grape Wine Res.*, **17**(2): 206-216.

Piedallu C., Gegout J.-C., Bruand A. and Seynave I., 2011. Mapping soil water holding capacity over large areas to predict potential production of forest stands. *Geoderma*, **160**(3-4): 355-366.

Pouget R., 1963. Recherches physiologiques sur le repos végétatif de la vigne (*Vitis vinifera* L.) : la dormance des bourgeons et le mécanisme de sa disparition. *Thèse de Doctorat*, Université de Bordeaux (France).

Reynolds, A.G. and J.E.V. Heuvel. 2009. Influence of grapevine training systems on vine growth and fruit composition: A Review. *Am. J. Enol.Vitic.,* **60**:251-268.

Rives M., 2000. Vigour, pruning, cropping in the grapevine (*Vitis vinifera* L.). I. A literature review. *Agronomie,* **20**:79-91.

Smart D.R., Schwass E., Lakso A. and Morano L., 2006. Grapevine rooting patterns: A comprehensive analysis and a review. *Am. J. Enol. Vitic.*, **57**(1): 89-104.

Sturges H.A., 1926. The choice of a class interval. *J. Am. Statistical Assoc.*, **21**(153): 65-66.

Tagliavini M. and Marangoni B., 1992. Effects of high root-zone temperature on growth, water relations and mineral uptake of Trebiano grapevine. *In: Proceedings of the 4th Int. Symp. on Grapevine Physiology*, Turin (Italy).

Takagi T. and Sugeno M., 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, **15**(1): 116-132.

Tesic D., D.J. Woolley, E.W. Hewett, and D.J. Martin. 2002. Environmental effects on cv Cabernet Sauvignon (*Vitis vinifera* L.) grown in Hawke's Bay, New Zealand. 1. Phenology and characterisation of viticultural environments. *Aust. J. Grape Wine Research*, **8**:15-26.

Trégoat O., Ollat N., Grenier G. and Leeuwen C.V., 2001. Étude comparative de la précision et de la rapidité de mise en œuvre de différentes méthodes d'estimation de la surface foliaire de la vigne. *J. Int. Sci. Vigne Vin*, **35**(1): 31-39.

Valente de Oliveira J., 1999. Semantic constraints for membership function optimization. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, **29**(1): 128-138.

Vaudour E., 2003. *Les terroirs viticoles. Définitions, caractérisation et protection.* Éd. Dunod-Lavigne, Paris.

Woodham R.C. and Alexander D.M., 1966. The effect of root temperature on development of small fruiting Sultana vines. *Vitis*, **5**: 345-250.

Zadeh L.A., 1965. Fuzzy sets. *Information and Control*, **8**(3): 338-353.

Zelleke A. and Kliewer W.M., 1980. Effect of root temperature, rootstock and fertilization on bud-break, shoot growth and composition of 'Cabernet-Sauvignon' grapevines. *Scientia Horticulturae*, **13**(4): 339-347.

Zufferey V. and Murisier F., 2006. Terroirs viticoles vaudois et alimentation hydrique de la vigne. *Rev. Suisse Vitic., Arboric. et Hortic.*, **38**(5): 283-287.


©Vigne et Vin Publications Internationales (Bordeaux, France)