

Fuzzy modeling of a composite agronomical feature using FisPro: the case of Vine Vigor

Cécile Coulon-Leroy, Brigitte Charnomordic, Marie Thiollet-Scholtus, S.
Guillaume

► **To cite this version:**

Cécile Coulon-Leroy, Brigitte Charnomordic, Marie Thiollet-Scholtus, S. Guillaume. Fuzzy modeling of a composite agronomical feature using FisPro: the case of Vine Vigor. 15. International Conference - IPMU 2014, Université Montpellier 2 (Sciences et Techniques) (UM2). Montpellier, FRA., Jul 2014, Montpellier, France. 564 p., 10.1007/978-3-319-08795-5 . hal-02743587

HAL Id: hal-02743587

<https://hal.inrae.fr/hal-02743587>

Submitted on 3 Jun 2020

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

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

Fuzzy Modeling of a Composite Agronomical Feature Using FisPro: The Case of Vine Vigor

Cécile Coulon-Leroy^{1,3}, Brigitte Charnomordic², Marie Thiollet-Scholtus³,
and Serge Guillaume⁴

¹ LUNAM Université, Groupe ESA, UPSP GRAPPE, Angers, France
c.coulon@groupe-esa.com

² INRA, SupAgro, UMR MISTEA, Montpellier, France
bch@supagro.inra.fr

³ INRA, SAD-UEVV-1117, Colmar, France
marie.scholtus@angers.inra.fr

⁴ Irstea, UMR ITAP, Montpellier, France
serge.guillaume@irstea.fr

Abstract. Fuzzy logic is a powerful interface between linguistic and numerical spaces. It allows the design of transparent models based upon linguistic rules. The *FisPro* open source software includes learning algorithms as well as a friendly java interface. In this paper, it is used to model a composite agronomical feature, the vine vigor. The system behavior is characterized by its numerical accuracy and analyzed according to the induced knowledge. Well known input output relationships are identified, but also some rules reflect local interactions.

Keywords: Fuzzy rules, Learning, Agronomy, Vine

1 Introduction

In many application fields, there is a need for interactive computerized system that gather data and knowledge from a wide range of sources, in order to help understanding a complex phenomenon and making a decision. In particular, the application fields of agronomy and environment are in demand for advanced modeling and decision support tools while presenting some specificities. Observational and modeling studies call for various spatio-temporal scales: plot, farm, catchment basin, designation of origin, and many influential parameters have to be considered. Data are acquired with very different means and resolutions, ranging from manual costly measurements such as soil analysis or plant hydric potential to high resolution data from embedded sensors and aerial images. In any case, the variability that characterizes life sciences is pervasive and results in data uncertainty and a lack of reproducibility. Therefore, human expertise is fundamental in interpreting data, and stakeholders in agronomy and environmental have always relied on expertise to interpret observations and to take decisions.

When expertise and data have to be integrated in a reasoning framework, fuzzy logic and Fuzzy Inference Systems (FIS) can play an original part in the

modeling and decision support process [1]. Let us take as an example the case of vine growing, that is a complex agricultural system involving several factors. The vegetative vine development is called vigor. It takes into account the rhythm and the intensity of the vine shoot growth. Empirically, in relative terms, vine vigor level of a plot is well known as being stable over the years. It is highly influenced by environmental factors, such as soil or climate, but can also be modified by agricultural practices (choice of rootstock, inter-row management, pruning type, among others). Vine vigor is a key parameter to control the balance between vegetative growth and productivity, which influences berry composition and then wine characteristics.

For a composite feature such as vine vigor, it is unrealistic to design formal mathematical models based on ecophysiological knowledge. An alternative approach consists in deriving empirical models from experiments. However, for perennial crops such as grapevine, full experimental designs to test a large number of factors in interaction are very difficult to implement. Furthermore the collected data are tainted with uncertainty; the features can suffer from imprecision, as many assessments are made by human beings. The learning process must be adapted to deal with partial and imperfect data, and to include valuable pieces of expert knowledge.

Various learning methods can be used to produce a model to study interactions between variables. They include artificial intelligence or statistical techniques. Both can deal with some kinds of data imperfection and both have been used in agri-environmental modeling. Common choices include classical linear models and decision trees [2] or, for more recent developments, Bayesian networks [3]. These statistical models are efficient in a wide range of situations, and often yield a confidence interval, since they are based on probability theory. However, they may be difficult to interpret or to use in cases where data imperfection and uncertainty is prevalent. Fuzzy modeling and FIS offer an interesting alternative in such a case, mainly because they provide an interface between the numerical and the linguistic spaces [4].

The objective of the present paper is to show the interest of FIS to study the interactions between environmental factors, agricultural practices and vine vigor. The approach attempts to make the best of domain expertise and of available field data, though they are incomplete, in order to design an interpretable model. The interpretability makes it possible to analyze the system behavior and to evaluate interactions between variables.


Software with a friendly interface is indispensable to allow interactive modeling and decision support. *FisPro*¹ is an open source software that has been recently used in several agronomic and environmental applications. These applications cover different topics: agricultural management using decision variables defined at catchment scale [5]; modeling interactions among sustainability components of an agro-ecosystem [6]; determining optimum rates of nitrogen for corn on the basis of field and crop features [7]; predicting vine vigor and

¹ <http://www.fispro.org/>, email: [fispro@supagro.inra.fr](mailto: fispro@supagro.inra.fr)

precocity [8]; characterizing management zones from viticulture data [9]; a soil occupation diagnostic for sugarcane harvest [10].

The paper is organized as follows. Section 2 presents some general functionalities for fuzzy modeling and learning with *FisPro*. Section 3 describes the application to vine vigor, and analyzes the results from a knowledge discovery point of view. Some conclusions are given in Section 4.

2 FisPro: Fuzzy Inference System Design and Optimization

FisPro  implements an applied research work in the field of interpretable FIS [11,4,1]. *FisPro* has been used to design FIS in different domains, including economics, hydraulics, robotics, agri-food industry, medicine, agronomy and environment. . . Among fuzzy software products, *FisPro* stands out because of the interpretability of fuzzy systems automatically learned from data. Interpretability is guaranteed at each step of the FIS design with *FisPro*: variable partitioning, rule induction, optimization.

2.1 A Framework for Fuzzy Modeling

A typical FIS is represented in Figure 1. It consists of three stages: 1) fuzzification to transform numerical values into membership degrees in the fuzzy sets associated to linguistic concepts, 2) fuzzy rule base, 3) defuzzification process, to infer a crisp value from the rule aggregation result.

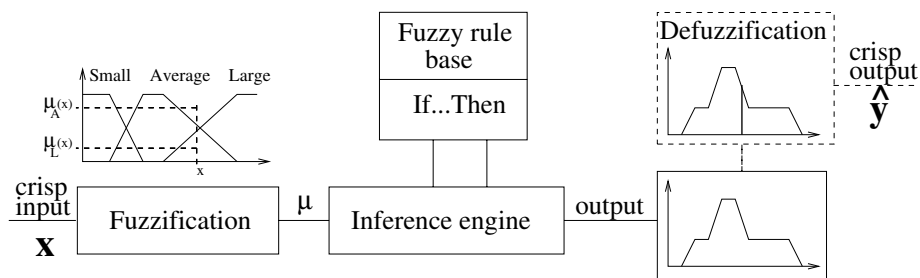


Fig. 1. A fuzzy inference system

In the present work, *FisPro* has been used to design FIS, according to the approach summarized in Figure 2.

That approach combines expertise and data. For instance, the number of fuzzy sets is chosen according to expert knowledge, but data make easier the lengthy task of linguistic modeling, by learning fuzzy set characteristic points. Automatic rule learning is done to highlight the interactions that arise from the

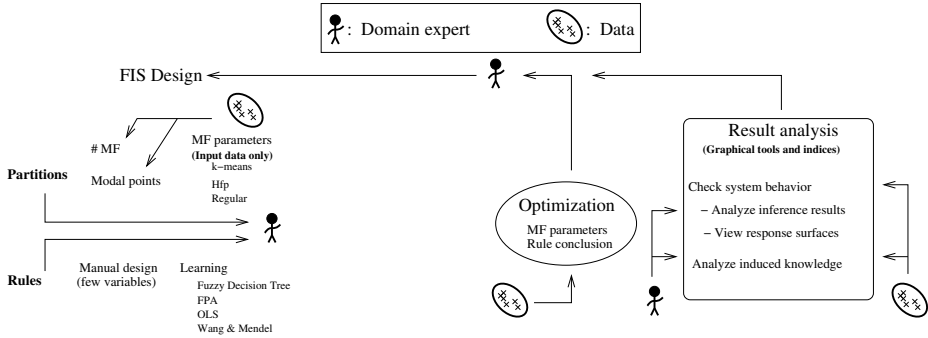


Fig. 2. FIS and linguistic modeling with *FisPro*

multidimensional characteristics. Optimization allows FIS parameter tuning. In a case study when data are scarce, as for the study of vine vigor, the FIS output is analyzed from a knowledge discovery point of view, instead of the classical validation step only based on numerical accuracy.

2.2 Linguistic Variable and Fuzzy Partitioning with *FisPro*

Working with the membership degrees in the different linguistic concepts, instead of the raw data values, reduces the system sensitivity to raw data variation. This is a convenient and meaningful way to tackle biological variability.

The readability of fuzzy partitioning is a pre-requisite condition to build an interpretable rule base. The necessary conditions for interpretable fuzzy partitions have been studied by several authors [12]: Distinguishable concepts, a justifiable number of fuzzy sets, coverage (each data point, x , should belong significantly, $\mu(x) > \epsilon$, at least to one fuzzy set), normal and significantly overlapping fuzzy sets. These requirements are all fulfilled by the strong fuzzy partitions, illustrated in Figure 3. For each point in the universe, the sum of the membership degrees in all the fuzzy sets is equal to one. Even if, in interactive design, other membership function (MF) shapes are available and fuzzy partitions can be freely adjusted, *FisPro* automatic procedures systematically generate strong fuzzy partitions with semi-trapezoidal shapes at the edges and either triangular or trapezoidal shaped MFs elsewhere.

The process of partitioning comes to choose the number of fuzzy sets and the corresponding characteristic points. When possible, the number of fuzzy sets is determined by expertise, in order to facilitate the interpretation. Variables can be continuous or discrete, under the condition that their values are ordered and have a progressive semantic meaning. Discrete variables are described by k ordered values. The characteristic points of MFs for continuous inputs are not easy to determine only by expertise so a learning procedure can be run (see Figure 2), for instance with the monodimensional k-means algorithm on the

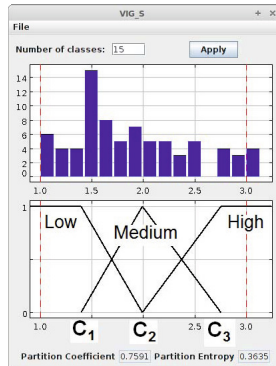


Fig. 3. A strong fuzzy partition with three linguistic labels

input data, independently for each variable. Then the cluster centers are chosen as characteristic points.

In *FisPro*, MF partitions can be visualized together with the data distribution, as shown in Figure 3. They can be edited and all modifications are dynamically passed on to other windows, including the inference window.

Several indices have been defined to characterize fuzzy partitions, and are available in *FisPro*. The partition coefficient (PC) and the partition entropy (PE), both proposed by Bezdek [13], are implemented in *FisPro*. Let N be the data set size, c the number of fuzzy sets and $\mu_i(k)$ the membership degree of the k th item in the i th group, the available indices are $PC = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^c \mu_i^2(k)$ and

$$PE = -\frac{1}{N} \left\{ \sum_{k=1}^N \sum_{i=1}^c [\mu_i(k) \log_a(\mu_i(k))] \right\}.$$

According to these criteria a good partition should minimize entropy and maximize the coefficient partition.

2.3 Rule Learning and Optimization with FisPro

The fuzzy rules are defined as:

$$IF X_1 \text{ is } A_1^r \text{ AND } X_2 \text{ is } A_2^r \dots \text{ AND } X_p \text{ is } A_p^r \text{ THEN } y \text{ is } C_r .$$

Rule Learning

It is difficult to define a *best method* for fuzzy rule learning. Several methods are available in *FisPro*, and are detailed in [4]. They all respect the interpretability of the fuzzy partitions. If the number p of input variables is high, it is advisable to select the most influential ones prior to learning, for instance by running fuzzy decision trees. If p is not too high, *WM* (Wang & Mendel) can be used to handle classification cases. For regression cases, when the number n of available data items is high, a statistical inspired method (*Fuzzy Orthogonal Least Squares*) is

interesting to detect outliers and select the most important sources of output variability. The *FPA* method (*Fast Prototyping Algorithm*), which is used in the present work, yields a data summary under the form of fuzzy rules.

FPA principles are recalled below. In a first step, the rules corresponding to the input combinations are generated, only if there are corresponding data in the data set. In a second step their conclusions are initialized according to the data

values: $C_r = \frac{\sum_{i \in E_r} W_r(x_i) * y_i}{\sum_{i \in E_r} W_r(x_i)}$, where $W_r(x_i)$ is the matching degree of the i th

example for the r th rule, defined as: $W_r(x) = \mu_{A_1^r}(x_1) \wedge \mu_{A_2^r}(x_2) \wedge \dots \wedge \mu_{A_p^r}(x_p)$ and E_r is a subset of examples chosen according to their matching degree to the rule. If there are not enough items that fire the r th rule with a degree higher than the user defined threshold, the rule is not kept. Thus, FPA yields a subset of all the possible rules. We set the threshold to a membership degree of 0.2, and the minimum cardinality of E_r to 1.

Optimization and Median FIS

Parameter optimization allows to optimize all parts of a FIS, using the Solis and Wets algorithm, see [14] for details. As partition parameters and rules have been generated separately, it is interesting to run an optimization procedure of the model as a whole. The optimization algorithm used in this work has been proposed in [1]. It is adapted from Glorennec [15] and based upon the work of Solis and Wets [16]. It allows optimizing all of the FIS parameters: input or output partitions and rule conclusions.

In our approach, the input variables were optimized each in turn, the order depending on the variable importance. To assess that importance, the variables were ranked according to a fuzzy decision tree. The data set was split into a learning set (70% of the vine plots) and a test set (30% of the vine plots). Ten pairs of learning and test sets were randomly created, taking into account the output distribution levels. The optimization procedure was guided by the root

mean square error $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{y}_i - y_i\|^2}$, where \hat{y}_i is the inferred value

for the i th item, y_i its observed value and N the number of items. The usual R-squared (R2) was also used to characterize the system accuracy.

The optimization process does not change the system structure; the number of MFs remains the same for all the variables as well as the rule premise structure. Only the MF parameters and the rule conclusions are modified. This allows the semantic properties of the initial model to be preserved while the model accuracy is improved. This also allows to define a *median FIS* from the ten-fold optimization sequence. The *median FIS* fuzzy set characteristic points are the median values of the corresponding values obtained for each fold, as well as the consequent parts of the rules.

FisPro provides some useful tools for exploratory analysis of the FIS behavior and representativeness, such as response surfaces and the summary of links

between rules and activated examples. Sampling procedures for cross-validation are integrated in the learning and optimization processes.

3 System Analysis

The case study is located in the middle Loire Valley, on the Saumur Protected Designation of Origin (PDO, in French Appellation d'Origine Contrôlée) area, in France. It includes 78 vine plots. All the vineyards considered in the study were planted with *Vitis vinifera* cultivar *Cabernet franc*, the main grape variety in the region, and with similar planting densities. In the studied area, according to expert knowledge, the vine vigor is influenced by soil factors and by two main agricultural practices: rootstock choice and inter-row management. The system was built using three input variables corresponding to the three main influential factors:

1. Vine vigor imparted by soil (VIG_S). This indicator is calculated using a fuzzy inference system [8], considering three input variables: the water holding capacity, the gravel percentage in the soil profile and the parent rock hardness. VIG_S is a numerical variable between 1 (Low imparted vigor) and 3 (High imparted vigor) .
2. Vigor conferred by rootstock (VIG_R). Vine is grafted on a rootstock to fight against the attack of an insect called *Phylloxera vastatrix*. The rootstock, at the interface between soil and vine variety, interacts with the variety to modify the development of the whole plant. For each rootstock, vigor level was determined from the literature. VIG_R is a discrete variable with five values (1 - Very low; 1.5 - Low; 2 - Medium; 2.5 - High and 3 - Very High).
3. The inter-row management constraint on vine vigor (VIG_C). A grass cover is introduced in the vineyard inter-rows to limit runoff and soil erosion. However it also limits vine vegetative development on account of competitions for soil water and nitrogen. VIG_C was defined as a discrete variable with 10 values (between 0 - No constraint and 3 - High constraint). Constraint values were obtained by crossing the constraint imparted by the cover crop variety, thanks to advisory services technical reports, and the cover crop area.

The system output is the vine vigor given by expert assessment, VIG_{OBS} . Vigor is linked to the shoot growth and leaf areas observed on vine plots. VIG_{OBS} can take one of the following labels: 1 - Very Low; 2 Low; 3 - High and 4 - Very High. The Medium label was not used on purpose to avoid *safe haven* assessments.

The number of fuzzy sets was determined by expertise, in order to have a number of concepts corresponding to the usual expert vocabulary used by domain experts and technicians. The discrete variable, VIG_R , was described by five ordered values. VIG_S and VIG_C were partitioned into three fuzzy sets corresponding to the usual terms Low, Medium and High. The initial fuzzy set characteristics points are given in Table 1 (left).

The rule base learnt by running the FPA procedure described in Section 2 is shown in Table 2 and some comments are given in the following paragraphs.

Table 1. The fuzzy partition parameters (left) and FIS accuracy (right)

Param.	VIG_S		VIG_C			FIS	RMSE	R^2
	Initial	Optimized	Initial	Optimized		Learning set	Initial	0.67
C1	1.4	1.4	1.00	1.02		Optimized	0.52	0.77
C2	2.0	2.1	1.65	1.50		Rel. gain	22%	20%
C3	2.8	2.1	2.25	2.18	Test set	Initial	0.67	0.60
						Optimized	0.54	0.73
						Rel. gain	19%	22%

Table 2. The fuzzy rule base

Rules	System inputs			Rule conclusions	
	VIG_S	VIG_R	VIG_C	Initial	Optimized
1	Medium	Low	High	2.6	2.1
2	Low	Medium	Medium	3.7	4.0
3	Low	Low	High	1.3	1.2
4	Low	Low	Medium	1.2	1.3
5	Medium	Low	Medium	2.5	2.4
6	Low	Medium	High	3.5	3.8
7	High	Low	High	4.0	4.0
8	Medium	Medium	Medium	2.7	1.4
9	Medium	Medium	High	2.7	1.1
10	Low	Medium	Low	2.5	2.2
11	High	Medium	Medium	3.0	2.9
12	Medium	Medium	Low	3.3	3.2
13	High	Medium	High	3.0	2.9
14	High	Low	Medium	4.0	4.0
15	Low	Low	Low	3.2	3.8
16	Low	High	Medium	4.0	3.9
17	High	Low	Low	4.0	3.9
18	Medium	Low	Low	2.7	2.5
19	High	Medium	Low	3.5	4.0

First of all, only 19 rules were generated because some combinations were absent from the learning data set. No vine plots were planted with a rootstock that confers either a Very Low or a Very High vine vigor level. After optimization, the fuzzy set parameters C_2 and C_3 of VIG_S were identical (Table 1), so that there was no smooth transition between a Medium level of VIG_S and a High level.

Consequents of rules 8 and 9 strongly decreased after optimization (-1.3 and -1.6 on a [1-4] scale) in contrast with the consequent of rule 2 that did not much change. For the rules corresponding to a Medium VIG_S , the rule conclusions systematically decreased after the optimization.

Table 1 (right) summarizes the results of optimization runs, comparing the average results of the initial and the median FIS over the learning and test samples. The median FIS significantly improved the accuracy over the test samples, with a relative gain of 19% for the RMSE and 22% for the R^2 . The median FIS-

based model has a relatively good accuracy, so its behavior can be interpreted to study the interactions pointed out by the rules. Some examples are given below.

In rules 11, 13 and 19, environmental factors imparting a High vigor are not associated to a rootstock that confers a Low vigor level. Goulet and Morlat [17] already noticed that the practices in the vineyard are sometimes unsuitable because they have not been well adapted to environmental factors. For example, the authors indicate that in the vineyard of the Sarthe in Loire Valley (France), 72% of the vine plots have a too vigorous rootstock since the environmental factors already induce a very strong vigor.

The effect of the VIG_C variable can also be discussed. When vine plots have no intercrop, i.e. no constraint on vine vigor, $VIG_C=Low$ (rules 10, 12, 15, 17, 18 and 19), the estimated vigor is always higher than 2, unlike vine plots with an intercrop. The impact of a grass cover as intercrop on vine is well known in the literature due to competition for water and nitrogen. In [18], the authors indicated that intercrop reduces vine growth, i.e. the vigor, of the present year but also of the next years by decreasing grapevine nitrogen reserves. These already known relationships, interpreted by expertise, confirm the ability of the method to extract knowledge from a database.

Let us now consider plots intercropped with a crop that involves a High constraint ($VIG_C=High$). When the soil imparts a Medium or a Low vigor, the estimated vigor is coherent with the empirical knowledge: a Low vigor rootstock leads to a lower vigor; the more the soil imparts a Low vigor, the greater the difference between rootstocks. On the contrary, when the soil imparts a High vigor level, and for Low vigor rootstock, the system estimates a High vigor level (rule 7). At first sight, this is unexpected. It might be explained by the ability of some Low vigor rootstocks to adapt to soil humidity.

4 Conclusion

The use of a fuzzy formalism for inference systems increases the model complexity by introducing more parameters, by having to choose fuzzy operators and so on. One must be careful that this rise in complexity is accompanied by some benefits. The work presented in this paper tried to show the interest of using fuzzy inference systems that integrate expertise and data, for modeling a complex phenomenon in agronomy, namely vine vigor.

From the agronomical point of view, our procedure allows to study the combinations of features, therefore complementing the expertise, which is often related to the effect of one feature, independently from the other ones. This work lays down the foundations of a decision support tool aiming to adapt the agricultural practices to the environment in order to get a given vigor target. A next step consists in testing the method in other vineyards, including rule analysis and system behavior assessment.

From the methodological point of view, some work remains to be done to deal with the uncertainty of input and output measurements or assessments, for instance to define accuracy indices taking into account a fuzzy target instead of a crisp one.

References

1. Guillaume, S., Charnomordic, B.: Fuzzy inference systems: an integrated modelling environment for collaboration between expert knowledge and data using fispro. *Expert Systems with Applications* 39, 8744–8755 (2012)
2. Thomopoulos, R., Destercke, S., Charnomordic, B., Iyan, J., Abécassis, J.: An iterative approach to build relevant ontology-aware data-driven models. *Information Sciences* 221, 452–472 (2013)
3. Sicard, M., Baudrit, C., Leclerc-Perlat, M.N., Wuillemin, P.H., Perrot, N.: Expert knowledge integration to model complex food processes. application on the camembert cheese ripening process. *Expert Syst. Appl.* 38, 11804–11812 (2011)
4. Guillaume, S., Charnomordic, B.: Learning interpretable fuzzy inference systems with fispro. *International Journal of Information Sciences* 181, 4409–4427 (2011)
5. Colin, F., Guillaume, S., Tisseyre, B.: Small catchment agricultural management using decision variables defined at catchment scale and a fuzzy rule-based system: a mediterranean vineyard case study. *Water Resources Management* 25, 2649–2668 (2011)
6. Rajaram, T., Das, A.: Modeling of interactions among sustainability components of an agro-ecosystem using local knowledge through cognitive mapping and fuzzy inference system. *Expert Systems with Applications* 37, 1734–1744 (2010)
7. Tremblay, N., Bouroubi, M., Panneton, B., Guillaume, S., Vigneault, P., Bélec, C.: Development and validation of fuzzy logic inference to determine optimum rates of n for corn on the basis of field and crop features. *Precision Agriculture* 11, 621–635 (2010)
8. Coulon-Leroy, C., Charnomordic, B., Rioux, D., Thiollet-Scholtus, M., Guillaume, S.: Prediction of vine vigor and precocity using data and knowledge-based fuzzy inference systems. *Journal International des Sciences de la Vigne et du vin* 46, 185–206 (2012)
9. Grelier, M., Guillaume, S., Tisseyre, B., Scholasch, T.: Precision viticulture data analysis using fuzzy inference systems. *Journal International des Sciences de la Vigne et du vin* 41, 19–31 (2007)
10. El Hajj, M., Bégué, A., Guillaume, S.: Integrating spot-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices - the case of sugarcane harvest on reunion island. *Remote Sensing of Environment* 113, 2052–2061 (2009)
11. Jones, H., Charnomordic, B., Dubois, D., Guillaume, S.: Practical inference with systems of gradual implicative rules. *IEEE Transactions on Fuzzy Systems* 17, 61–78 (2009)
12. de Oliveira, J.V.: Semantic constraints for membership functions optimization. *IEEE Transactions on Systems, Man and Cybernetics. Part A* 29, 128–138 (1999)
13. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Functions Algorithms*. Plenum Press, New York (1981)
14. Guillaume, S., Charnomordic, B.: Parameter optimization of a fuzzy inference system using the fispro open source software. In: CFP12FUZ-USB, I.C.N: IEEE International Conference on Fuzzy Systems, pp. 402–409. IEEE, Brisbane (2012)

15. Glorennec, P.Y.: Algorithmes d'apprentissage pour systèmes d'inférence floue. Editions Hermès, Paris (1999)
16. Solis, F.J., Wets, R.J.: Minimization by random search techniques. *Mathematics of Operation Research* 6, 19–30 (1981)
17. Goulet, E., Morlat, R.: The use of surveys among wine growers in vineyards of the middle-loire valley (france), in relation to terroir studies. *Land Use Policy* 28, 770–782 (2010)
18. Celette, F., Findeling, A., Gary, C.: Competition for nitrogen in an unfertilized intercropping system: the case of an association of grapevine and grass cover in a mediterranean climate. *European Journal of Agronomy* 30, 41–51 (2009)