Fuzzy Inference System as an Aggregation Operator - Application to the Design of a Soil Chemical Quality Index

Denys Yohana Mora-Herrera¹, Serge Guillaume², Didier Snoeck³, and Orlando Zúñiga Escobar¹

¹ Universidad del Valle, Cali, Colombia
² ITAP, Univ Montpellier, INRAE, Montpellier SupAgro, Montpellier, France
³ CIRAD, Montpellier, France

{denys.mora,orlando.zuniga}@correounivalle.edu.co
serge.guillaume@inrae.fr
didier.snoeck@cirad.fr

Abstract. Fuzzy logic is widely used in linguistic modeling. In this work, fuzzy logic is used in a multicriteria decision making framework in two different ways. First, fuzzy sets are used to model an expert preference relation for each of the individual information sources to turn raw data into satisfaction degrees. Second, fuzzy rules are used to model the interaction between sources to aggregate the individual degrees into a global score. The whole framework is implemented as an open source software called GeoFIS. The potential of the method is illustrated using an agronomic case study to design a soil chemical quality index from expert knowledge for cacao production systems. The data come from three municipalities of Tolima department in Colombia. The output inferred by the fuzzy inference system was used as a target to learn the weights of classical numerical aggregation operators. Only the Choquet Integral proved to have a similar modeling ability, but the weights would have been difficult to set from expert knowledge without learning.

Keywords: Fusion · Multicriteria · Preference · Decision

1 Introduction

Fuzzy logic is widely used as an interface between symbolic and numerical spaces, allowing the implementation of human reasoning in computers. Fuzzy inference systems are well known for their ability for linguistic modeling and approximate reasoning. They are used in this work in the framework of multicriteria decision making to model expert knowledge.

Complex systems, such as agricultural production systems, are characterized by several interrelated dimensions, for instance agronomic, social and economic. The production process includes different steps performed systematically, from the plant selection to the commercialization. Decisions are made at each step of
this process that may degrade or support the sustainability of the production system.

Decision making usually involve several, may be conflicting, attributes. Multicriteria decision analysis (MCDA) has been indicated as the appropriate set of tools to perform assessments of sustainability, by considering different sustainability spheres, perspectives, stakeholders, values, uncertainties and intra and inter-generational considerations [2]. In this article five MCDA methods are reviewed on the basis of ten criteria that they should satisfy to properly handle problems concerning sustainability. The methods are from three main families: i) utility-based theory: multi attribute utility theory (MAUT) and analytical hierarchy process (AHP), ii) outranking relation theory: elimination and choice expressing the reality (ELECTRE) and preference ranking organization method for enrichment of evaluations (PROMETHEE), iii) the sets of decision rules theory: dominance based rough set approach (DRSA). The latter uses crisp ‘if . . . then’ rules where the premise compares for each criterion its satisfaction degree to a threshold and the conclusion is a category or a set of categories.

In a review about MCDA applied to forest and other natural resource management [7] a special attention is paid to methods that deal with uncertainty. The possible causes of uncertainty are analyzed and the study reports how the methods are adapted to manage some dimension of uncertainty. A fuzzy multiple objective linear programming method is mentioned [8].

Agricultural production systems involve agronomic, social, cultural, institutional, economic and other natural elements that are interrelated. Cacao production has been studied for many years. It is grown in climatic, economic and social uncertain contexts, then more efforts by farmers (time, money or land) do not always produce more quality, quantity, profitability nor a better life quality for farmers. So, it is a dynamic and complex system characterized by non-linear relationships dependent of local contexts. Moreover data is not enough and knowledge is needed to turn data into valuable agronomic information, for instance to make a decision about fertilization from a soil content analysis.

The challenge is to design a tool that includes the available scientific knowledge to help farmers in decision making. This tool is designed as an indicator with three main components: agronomic, economic and social, and final product quality. For sustainable and competitive cacao production the fertility status of the soil is an important variable that is generally related to cacao agronomic yield. A subsystem of the agricultural quality part is analyzed in this work: the soil chemical quality index.

The implementation is achieved by the means of an open source software, called GeoFIS1, a platform for spatial data processing with a decision support perspective. The fusion module uses fuzzy logic in two different contexts: first to turn raw data into satisfaction degrees and second to model variable interaction using linguistic rules.

---

1 https://www.geofis.org.
Section 2 describes the data fusion framework, a subsystem of the soil chemical quality index is studied in Sect. 3. In Sect. 4 the FIS is compared with classical numerical operators. Finally the main conclusions are summarized in Sect. 5.

2 Data Fusion and Multicriteria Decision Making

The process of data fusion for decision making is driven by expert knowledge. Information fusion is done with a specific goal, for instance risk level evaluation or variable application rate in agriculture. The selection of the relevant and available information sources is done by the decision maker. Then the next step is to evaluate what could be the level of decision, e.g. risk or rate, according to each of the sources for a given entity defined by its spatial coordinates. The final step comes down to aggregate these partial levels, or degrees, to make the final decision. The aggregation function models the decision maker preferences: Are some attributes more important than others? How to combine conflicting information sources?

The whole framework can be illustrated as follows:

\[(a_1, \ldots, a_n), (b_1, \ldots, b_n) \xrightarrow{\text{A}} f(a), f(b)\]

\[\uparrow\hspace{2cm}\downarrow\]

\[(x_1, \ldots, x_n), (y_1, \ldots, y_n) \prec (a, b)\]

There are two steps to formalize expert knowledge and preferences for the decision process. The first one deals with each individual variable, or information source. The second one addresses the interaction between sources.

The first step aims to turn, for each individual variable, \(x_i\) or \(y_j\), raw data into satisfaction degrees, \(a_i\) or \(b_j\). This is done by defining a preference relation through a fuzzy set. The scale is the unit interval, \([0, 1]\) with zero meaning the criterion is totally not satisfied and that it is fully satisfied with one. Then the degrees are aggregated using the \(f\) function to compute a global score, \(f(a)\) or \(f(b)\).

2.1 Numerical Operators

The most popular techniques to aggregate commensurable degrees are numerical operators. The main families of such operators, with suitable properties, are the Weighted Arithmetic Mean (WAM), the Ordered Weighted Average (OWA) and the Choquet Integral (CI).

Let \(X\) be the set of sources to aggregate: \(X = \{a_1, \ldots, a_n\}\).

The WAM aggregation is recalled in Eq. 1.

\[WAM(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_i\] (1)
with \( w_i \in [0,1] \) and \( \sum_{i=1}^{n} w_i = 1 \). The weights are assigned to the sources of information.

The OWA is computed as shown in Eq. 2.

\[
OWA(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{(i)}
\]  

(2)

where \((.)\) is a permutation such as \( a_{(1)} \leq \cdots \leq a_{(n)} \).

In this case, the degrees are ordered and the weights are assigned to the locations in the distribution, from the minimum to the maximum, whatever the information sources.

These operators are easy to use, the number of parameters is the number of information sources to aggregate\(^2\), but their modeling ability is limited. The Choquet Integral \([1]\) proved to be useful in multicriteria decision making \([3]\). It is computed according to Eq. 3.

\[
CI(a_1, \ldots, a_n) = \sum_{i=1}^{n} (a_{(i)} - a_{(i-1)}) w(A_{(i)})
\]  

(3)

where \((.)\) is a the permutation previously defined with \( a_{(0)} = 0 \) and \( A_{(i)} = \{(i), \ldots, (n)\} \), meaning the set of sources with a degree \( a \geq a_{(i)} \).

The weights must fulfill the two following conditions:

1. Normalization: \( w(\emptyset) = 0 \), \( w(X) = 1 \)
2. Monotonicity condition: \( \forall A \subset B \subset X, w(A) \leq w(B) \).

The weights are not only defined for each of the information sources but for all their possible combinations. Specific configurations include WAM and OWA modeling. The Choquet Integral is equivalent to a WAM when the sum of the weights assigned to the individual sources is one and when the weight of any coalition is the sum of the weights of its individual components. In this case the measure is additive. It is equivalent to an OWA when the weight of a coalition only depends on its size: for instance all the subsets with two elements have the same weight. This kind of measure is said to be symmetric. In the general case, the aggregation of \( n \) information sources requires \( 2^n - 2 \) coefficients. These are usually set by learning algorithms \([9]\).

### 2.2 Fuzzy Inference Systems

A fuzzy inference system usually requires more parameters than the former numerical aggregators but, in this particular case of data fusion, the design can be simplified as all the input variables are satisfaction degrees that share the same scale, the unit interval, and the same meaning. Strong fuzzy partitions

\(^2\) Only \((n - 1)\) parameters have to be defined for \( n \) sources as their sum is 1.
with regular grids are used to ensure semantic integrity. The inference system is also automatically defined: the membership degrees are aggregated using the minimum operator to compute the matching degree of each rule, the rule conclusions are aggregated using the Sugeno operator in the crisp case or using the centroid operator if fuzzy. More details are available in the FisPro documentation³.

The unique parameter left to the user is the number of linguistic terms for each input variable. In this work it was set at 2, Low and High, for all of them. With two linguistic labels by variable, the number of rules is $2^n$, i.e. the number of coefficients required by the Choquet Integral. A rule describes a local context the expert domain, the decision maker, is able to understand. In this way, the rule conclusions are easier to define than the Choquet Integral coefficients.

### 2.3 Implementation

The fusion module is implemented as an open source software in the GeoFIS program. The data must be co-located, i.e. a record includes the spatial coordinates of the cell, from a pixel to a zone, and the corresponding attributes.

The available functions to turn raw data into degrees are of the following shapes:

- **SemiTrapInf**: low values are preferred
- **SemiTrapSup**: high values are preferred
- **Trapezoidal**: about an interval
- **Triangular**: about a value

Three aggregation operators are currently available: WAM, OWA and a fuzzy inference system (FIS) including linguistic rules.

For WAM and OWA the weights can be learned provided a co-located target is available. Rule conclusions can also be learned using the FisPro software [4].

Rule conclusion can be either a linguistic term, fuzzy output, or a real value, crisp output. Using a fuzzy output, it would be necessary to define as many labels as different suitable rule conclusions. As a crisp conclusion may take any value in the output range, it allows for more versatility.

The output should also range in the unit interval. This constraint ensures the output can feed a further step of the process as shown in Fig. 1.

![Fig. 1. A hierarchical structure.](https://www.fispro.org)
In this way, the intermediate systems can be kept small, making their design and interpretation easier.

The *GeoFIS* program includes a distance function based on a fuzzy partition that allows for integrating expert knowledge into distance calculations [5] as well other functionalities, such as a zone delineation algorithm [10]. An illustration of its potential use in Precision Agriculture can be found in [6].

**3 Case Study**

Different indicators had been designed as guidance for characterizing and improving agricultural system sustainability, one of them is *Soil Fertility*. It is defined by the Soil Science Society of America as⁴: “The quality of a soil that enables it to provide nutrients in adequate amounts and in proper balance for the growth of specified plants or crops.”

The chemical index structure is shown in Fig. 2. The *Soil Nutritional Balance* is analyzed in this work. This subsystem include three input variables:

- **Soil pH** depends of the nature of the soil and controls the chemical processes that take place in the soil; specifically, the availability of nutrients. Moreover, it is a parameter that is very easy to measure in the field. The preference relation is defined as: $\sqrt{5, 5.5, 6.5, 7.5}$ [11].
- **Organic matter** brings Carbon that is an important macronutrient, its presence is very regular and usual in cacao plantations and, usually, it is not a limiting criterion. The preference relation is defined as: $\sqrt{3, 5}$ [11].
- **Macronutrients**: this variable is calculated from a combination of soil chemical data as shown in Fig. 2.

It is worth mentioning that the knowledge included in the index is generic: it is not related to a specific location but results from a worldwide analysis.

The study area is located in three municipalities of Tolima department in Colombia.

Soils were sampled at depth 0–15 cm (3 points per farm and 3 repetitions in each one). An agronomist surveyed the crops and filled out field register

---

⁴ [https://www.soils.org/publications/soils-glossary](https://www.soils.org/publications/soils-glossary)
about nutritional, productivity and plant health stage. An economist applied a structured interview to farmers to collect information about their agronomic and post-harvest practices and socioeconomic conditions. The fieldwork was achieved between December 2018 and January 2019.

Table 1. The Soil Nutritional Balance rule base.

| Organic matter | Soil pH | Macronutrients | Conclusion |
|----------------|---------|----------------|------------|
| 1 Low          | Low     | Low            | 0.0        |
| 2 Low          | Low     | High           | 0.2        |
| 3 Low          | High    | Low            | 0.4        |
| 4 Low          | High    | High           | 0.6        |
| 5 High         | Low     | Low            | 0.3        |
| 6 High         | Low     | High           | 0.5        |
| 7 High         | High    | Low            | 0.7        |
| 8 High         | High    | High           | 1.0        |

Even if the main interactions are known, there are various ways of modeling them. Rule conclusions were valued taking account hierarchical importance of each input variable and their contribution to aggregated variable when it is satisfied individually and when it is satisfied in combination with others. Always considering the premise that all is more than the sum its parts. The rule base for the Soil Nutritional Balance is given in Table 1. The most important input variable is Soil pH because it controls the chemical processes that take place in the soil, so when only this variable is satisfied, High, then the conclusion is 0.4, rule number 3. If only Organic matter or Macronutrients is High then the respective rule conclusions are 0.3 and 0.2. For any combination of two variables with High, the rule conclusion is set at the sum of conclusions when only one of them is High. For instance, rule number 4 involves the High label for Soil pH and Macronutrients and the conclusion is $0.6 = 0.4 + 0.2$. However, when the three variables are satisfied to a High level the conclusion is higher than the sum of the individual contributions, 0.9, to highlight their positive interaction on the crop agricultural quality. The corresponding rule conclusion, number 8, is set at 1.

4 Comparison with Numerical Aggregation Operators

The output of the fuzzy system for the 10 farms from the Chaparral municipality is used to learn the weights of the WAM, OWA and Choquet Integral. The weights of the WAM and the OWA were learned using a least square minimization procedure under two constraints for the weights: they must be positive and their sum should be 1. The `pnnls` function from the `lsei R` package was used. In a preprocessing step, the degrees for each farm were sorted in an increasing order to learn the OWA weights.
The data are given in Table 2.

**Table 2.** Comparison of the aggregation operators using the 10 farms from Chaparral.

|   | OM  | pH   | MacN | FIS  | WAM  | OWA  | CI  |
|---|-----|------|------|------|------|------|-----|
| 1 | 1   | 0.5  | 0.696| 0.655| 0.731| 0.730| 0.655|
| 2 | 0.87| 0.76 | 0.414| 0.586| 0.666| 0.660| 0.607|
| 3 | 0   | 1    | 0.337| 0.467| 0.437| 0.447| 0.463|
| 4 | 0   | 1    | 0.669| 0.534| 0.561| 0.520| 0.533|
| 5 | 0.525| 1   | 0.470| 0.675| 0.653| 0.680| 0.687|
| 6 | 0.86| 1    | 0.484| 0.770| 0.764| 0.759| 0.749|
| 7 | 0.355| 1   | 0.559| 0.652| 0.632| 0.641| 0.651|
| 8 | 0   | 1    | 0.773| 0.555| 0.599| 0.543| 0.555|
| 9 | 0   | 0.6  | 0.380| 0.309| 0.328| 0.307| 0.315|
| 10| 0   | 1    | 0.614| 0.523| 0.540| 0.508| 0.522|
| R²| -   |      |      | 0.912| 0.936| 0.993|     |

The degrees to aggregate are in the first three columns for Organic matter (OM), Soil pH (pH) and Macronutrients (MacN), followed by the output inferred by the fuzzy system (FIS). The remaining columns give the score for the Soil Nutritional Balance for the three aggregation operators tested. The determination coefficient between the operators and the FIS target are in the last row.

As expected, the WAM yields the poorest result. The weights for OM, pH and MacN are: 0.317, 0.312 and 0.371. The score for the first farm is similar to the one for farm #6 and higher than the one for farm #5. This result is not expected as soil of farm #1 has a less suitable Soil pH than farm #6. Farm #2 is also assigned a high score, higher than the ones of farms #5 and #7. This is not expected as the degrees of Soil pH and, to a lower extent, of Macronutrients are better for farms #5 and #7. The exception of Organic matter should not have been sufficient to get a higher score.

The weights for the OWA, from the minimum to the maximum, are: 0.407, 0.220 and 0.373. The results are slightly improved compared to the WAM, but the same comments about the score of farm #1 can be made. Farm #2 also gets an higher score than farm #7.

The Choquet Integral gives the best results and the farms are ranked in the same order. This is not surprising as the model requires more coefficients, $2^n - 2$ as the empty coalition is assigned a zero value and the whole set a one. The weights yielded by the HLMS algorithm [9] are shown in Table 3.

This table shows that the coefficients of the optimal Choquet Integral are really different from the fuzzy rule conclusions reported in Table 1. Two weights, OM and MacN, are set at zero. The first one has been optimized by the algorithm but this is not the case for the weight assigned to MacN. This is explained
Table 3. The weights of the *Choquet Integral* learned by the HLMS algorithm using the 10 farms from Chaparral.

|        | $OM$ | $pH$ | $Ma$ | $OM-pH$ | $OM-Ma$ | $pH-Ma$ |
|--------|------|------|------|---------|---------|---------|
| weight | 0.000| 0.392| 0.000| 0.558   | 0.794   | 0.603   |

by the fact that this degree is never the highest in the training set. This is an identified drawback of the algorithm, some values are not handled by the algorithm, depending on the data: they are called untouched coefficients in [9]. Any value lower or equal to 0.603, the minimum value for a coalition that includes $Ma$, would be acceptable.

A zero value for $OM$ does not mean this variable is not used: the weight was put on coalitions that include $OM$. In the two cases, the weight of the set is higher than the sum of the weights of its elements. For instance, the $OM-pH$ set is given a 0.558 weight, higher than the sum 0 + 0.392.

Even if the *Choquet Integral* proved to have an important modeling ability, its tuning remains difficult without learning algorithms.

5 Conclusion

In multicriteria decision making various kinds of operators can be used. Some are easy to use but have a limited modeling ability, such as the Weighted Arithmetic Mean. Others are more efficient but require a more important number of parameters whose setting may be difficult. This is the case for the *Choquet Integral*. This work shows that fuzzy logic can be used in two key steps of the aggregation process. First, fuzzy membership functions are used to model individual preferences and to turn raw data into satisfaction degrees for each of the information sources. Second, fuzzy inference systems, that implement linguistic reasoning, are suitable to model variable interaction and collective behavior in local contexts. Linguistic rules are easy to design for domain experts as they naturally use linguistic reasoning.

In the general case fuzzy inference systems require a lot of parameters to define the input partitions and the inference operators. In the particular case of data aggregation all the input variable are satisfaction degrees with common scale and common meaning. This leads to an automatic setting of inputs using a strong fuzzy partition with two linguistic terms, *Low* and *High*. As a consequence, only the rule conclusions have to be specifically defined by the user. This is the way expert knowledge about variable interaction is modeled.

This framework is implemented as an open source software called *GeoFIS* (See footnote 1). This is a strong asset as software support availability is a key factor for a method to be adopted.

The proposal was used to design a soil chemical quality index for cacao crop. It has a hierarchical structure with intermediate outputs easy to analyze. The membership functions were defined according to the available scientific knowledge. Even if the main interactions are known, there are several ways of modeling them. The results were easy to analyze and consistent with the field observations.
The output inferred by the fuzzy system was used as a target to learn the weights of alternative numerical aggregation operators. The most simple ones, WAM and OWA, yielded poor results. Only the Choquet Integral proved able to fit the target. The weights defined by the learning algorithm proved that the expert tuning of the Choquet Integral would have been difficult.

Fuzzy inference systems thanks to their proximity with natural language and expert reasoning are a good alternative framework for modeling preferences and multicriteria decision making.

References

1. Choquet, G.: Theory of capacities. Ann. de l’Institut Fourier 5, 131–295 (1954). https://doi.org/10.5802/aif.53
2. Cinelli, M., Coles, S.R., Kirwan, K.: Analysis of the potentials of multi criterion decision analysis methods to conduct sustainability assessment. Ecol. Ind. 46, 138–148 (2014). https://doi.org/10.1016/j.ecolind.2014.06.011. http://www.sciencedirect.com/science/article/pii/S1470160X14002647
3. Grabisch, M., Labreuche, C.: A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. 4OR 6(1), 1–44 (2008). https://doi.org/10.1007/s10288-007-0064-2
4. Guillaume, S., Charnomordic, B.: Learning interpretable fuzzy inference systems with FisPro. Inf. Sci. 181, 4409–4427 (2011). https://doi.org/10.1016/j.ins.2011.03.025
5. Guillaume, S., Charnomordic, B., Loisel, P.: Fuzzy partitions: a way to integrate expert knowledge into distance calculations. Inf. Sci. 245, 76–95 (2013). https://doi.org/10.1016/j.ins.2012.07.045
6. Leroux, C., et al.: GeoFIS: an open source, decision-support tool for precision agriculture data. Agriculture 8(6), 73 (2018). https://doi.org/10.3390/agriculture8060073
7. Mendoza, G., Martins, H.: Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. For. Ecol. Manag. 230(1), 1–22 (2006). https://doi.org/10.1016/j.foreco.2006.03.023. http://www.sciencedirect.com/science/article/pii/S0378112706002258
8. Mendoza, G.A., Bare, B.B., Zhou, Z.: A fuzzy multiple objective linear programming approach to forest planning under uncertainty. Agric. Syst. 41(3), 257–274 (1993). https://doi.org/10.1016/0308-521X(93)90003-K. http://www.sciencedirect.com/science/article/pii/0308521X9390003K. Multiple Criteria Analysis in Agricultural Systems
9. Murillo, J., Guillaume, S., Tapia, E., Bulacio, P.: Revised HLMS: a useful algorithm for fuzzy measure identification. Inf. Fusion 14(4), 532–540 (2013). https://doi.org/10.1016/j.infus.2013.01.002
10. Pedroso, M., Taylor, J., Tisseyre, B., Charnomordic, B., Guillaume, S.: A segmentation algorithm for the delineation of management zones. Comput. Electron. Agric. 70(1), 199–208 (2010). https://doi.org/10.1016/j.compag.2009.10.007
11. Rojas, F., Sacristán Sánchez, E.J.: Guía ambiental para el cultivo del cacao. Ministerio de Agricultura y Desarrollo Rural y Federación Nacional de Cacaoteros, Colombia (2013)