A fuzzy logic based soil chemical quality index for cacao

Denys Yohana Mora-Herrera, Serge Guillaume, Didier Snoeck, Orlando Zúñiga Escobar

A fuzzy logic based soil chemical quality index for cacao

Denys Yohana Mora-Herrera[a], Serge Guillaume[b,*], Didier Snoeck[c], Orlando Zúñiga Escobar[a]

[a] Universidad del Valle, Cali, Colombia
[b] ITAP, Univ Montpellier, INRAE, Montpellier SupAgro, Montpellier, France
[c] CIRAD, Tree Crop Based Systems Research Unit, Montpellier, France

Abstract

A soil chemical quality index is designed for cacao production systems. It is based on worldwide scientific knowledge and built using data from three municipalities of Tolima department in Colombia. 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*. A part of the data was used for calibration, then the remaining data served as a validation set. The results were easy to analyze and in agreement with field observations. 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 its optimal tuning would have been difficult without learning.

Keywords: Data fusion, aggregation, fuzzy inference system

1. Introduction

Complex systems, such as agricultural production systems, are characterized by several agronomic, social and economic interrelated dimensions. The production process includes different steps performed systematically, from the plant selection to the commercialization. Decisions are made at each step of

*Corresponding author
Email addresses: denys.mora@correounivalle.edu.co (Denys Yohana Mora-Herrera), serge.guillaume@inrae.fr (Serge Guillaume), didier.snoeck@cirad.fr (Didier Snoeck), orlando.zuniga@correounivalle.edu.co (Orlando Zúñiga Escobar)

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 (Cinelli et al., 2014). In this work five MCDA methods were analyzed 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) 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 (Mendoza & Martins, 2006) a special attention was paid to methods dealing with uncertainty. The possible causes of uncertainty were analyzed and the study reported how the methods were adapted to manage some dimension of uncertainty. A fuzzy multiple objective linear programming method is mentioned (Mendoza et al., 1993).

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 quality life for farmers. So, it is a dynamic and complex system characterized by nonlinear 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 has to be designed as an indicator that includes the three main components, agronomic, economic and social, and that enables to assess the quality of the final product. This work aims to develop a subsystem of the index for agricultural quality: the soil chemical quality index.

The soil must be acknowledged as a complex living, dynamic and natural system composed of a myriad of interacting chemical, physical, and biological processes (Kelting et al., 1999). Physical properties are related to soil structure and are quite stable in the mid term, they are not taken into account. Biological characteristics are likely to impact the chemical processes that are essentials for the functioning of terrestrial ecosystems, such as production of food, therefore, soil conservation and the maintenance of its long-term capability of food production are fundamental to the sustainability goals and ensure food security (Vogel et al., 2019). Hence, for the study of the cacao processing systems, the index is restricted to the chemical properties.

When soils are used for food production, farmers' management-induced changes affect the soil functions, for example, fertilization practices are required but in adequate conditions according to crop, deficient or excessive doses disturb the balance of soil and limit the crop yield. For sustainable and competitive cacao production the soil condition is an important parameter (Arthur et al., 2017).

The soil chemical subsystem was selected because it is most directly associated with agricultural management practices. Some soil properties are likely to be altered in the short and medium term (Liu et al., 2006). The predominant chemical indicators for soil quality are soil pH, electrical conductivity, adsorption capacity, available cations, organic matter and the nutrients balance (Thakur & Sharma, 2019). Additionally for cacao cultivation it is necessary to consider cadmium content because it is a strong restriction for Latin America countries.

Information about how some soil chemical properties, independent of each other, affect the cacao crops is available in several literature sources. However it is necessary to condense this knowledge into an index that considers the interactions between properties and processes to help farmers in management decision making. Agriculture professionals design recommendations about conservation and sustainable use of soil for this purpose. This is even more relevant when taking account that most of the cacao farmers have a low education level. So sometimes they have soil data but they do not known how to use them ; hence the proposed index can also improve their accessibility to existing information.

Several attempts have been made to design soil quality indices, see (Mukherjee & Lal, 2014; Vogel et al., 2019) for recent surveys. Some of them are dedicated to specific areas such as semi-arid tropics (Aravindh et al., 2019), Chinese blacks soils (Duan et al., 2009, 2011; Gu et al., 2018) or forests (Kelting et al., 1999). Others are designed with a specific purpose such as evaluation of soil degradation (Nosrati & Collins, 2019) or Life Cycle Assessment (De Laurentiis et al., 2019). Ferraro (2009) used a fuzzy rule framework designed from knowledge elicitation for Argentinean cropping systems.

Soil quality indices for crop cultivation (pollution issues excluded) are usually used to ensure that soil farming practices are compatible with the respect of environmental quality. In most cases (Thoumazeau et al., 2019), the calculated index is computed as the sum of several parameters on soil structure, nutrient availability and biological activity. The values can be scaled (in % of the minimum/maximum thresholds) or weighted (individual score between 0 and 1) or sorted (only the most effective are kept).

The proposed approach is original in that the parameters and thresholds used are specific to the cacao cultivation and based on worldwide expert knowledge. A preliminary version of this work was proposed in (Mora-Herrera et al., 2020).

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. Data combination is fully driven by expert knowledge and fuzzy logic is used at two different steps of the process. First, fuzzy sets are used to model an expert preference relation for each of the individual information sources according to the decision to be made. Second, fuzzy rules are used to model the interactions between sources to aggregate the individual degrees into a global score.

The implementation is achieved by the means of an open source software, called *GeoFIS*[1], a platform for spatial data processing with a decision support perspective. The design of the FIS is mainly automated. Only the rule conclusions, either real values in the unit interval or linguistic labels, are left to the user. This framework was used to manage another agronomic case study dealing the design of a nitrogen fertilization map within a vineyard (Guillaume et al., 2020).

The basics of linguistic modeling are recalled in Section **??**. Section 2 describes the data fusion framework. The soil chemical quality index based on fuzzy logic is studied in Section 3 and a comparison with classical aggregation operators is proposed in Section 4. Finally the main conclusions are summarized in Section 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 proposed approach computes a global score for each item, i.e, a site location. This is achieved using an aggregation *function* that handles values in the same range and with the same meaning: this kind of data is said to be commensurable. But except in the case of sensor fusion, the sources to be aggregated have not the same units and are not in the same scale. As a consequence an intermediate step is usually needed

[1] https://www.geofis.org

to turn raw data into satisfaction degrees. This is done, for each individual variable, by defining a preference relation for the considered attribute: What are the preferred values for the decision for this variable? Have some of these values of a similar meaning for the decision to be made? This is the first way to formalize expert knowledge. The design of the aggregation function allows to model interactions between variables: this is the second way to include another king of expert knowledge. These two steps are detailed below.

2.1. Dealing with individual variables

The preference relation defined for a given individual variable depends on the decision to be made. To assess the sustainability of a cropping system, the preferred value for any chemical input are the lowest. The preference relation can be modeled, or implemented, using a fuzzy set as shown in Figure 1 with an output range between 0 and 1, with 0 meaning the criterion is not at all satisfied and 1 that it is fully satisfied.

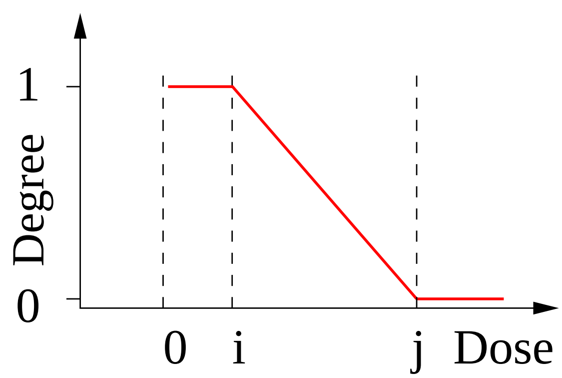

Figure 1: The chemical dose applied to the field and the corresponding satisfaction degree.

The satisfaction degree is 1 for any $x \leq i$ and 0 for any $x \geq j$. It is linearly decreasing with increasing values of x between i and j.

The transformation function includes the part of the expert knowledge related to each individual variable, without consider its interaction with the other variables.

2.2. Numerical operators

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

Let X be the set of sources to aggregate: $X = \{a_1, \dots, a_n\}$.

The WAM aggregation is recalled in Equation 1.

$$WAM(a_1, \dots, a_n) = \sum_{i=1}^{n} w_i a_i \tag{1}$$

with $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$. The weights are assigned to the sources of information. Unfortunately, WAM cannot model compromise as shown in this example. Let's consider three items described by two attributes with the following satisfaction degrees:

	a_1	a_2
It 1	0.7	0.7
It 2	0.4	1
It 3	1	0.4

Item 1 is preferred to the other two ones. This leads to the two conditions for the weights to fulfill:

- $Score(It1) > score(It2) \quad \Longrightarrow \quad w_1 > w_2$
- $Score(It1) > score(It3) \quad \Longrightarrow \quad w_2 > w_1$

These two conditions are contradictory, and there is no combination (w_1, w_2) that can model the decision maker preference.

The OWA (Yager, 1988) is computed as shown in Equation 2.

$$OWA(a_1, \dots, a_n) = \sum_{i=1}^{n} w_i a_{(i)} \tag{2}$$

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. *WAM* and *OWA* can be combined in the Weighted *OWA* operator, *WOWA* (Torra, 1996). This toy example can be modeled using an OWA with all the weight on the minimum.

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* (Choquet, 1954) proved to be useful in multicriteria decision making (Grabisch & Labreuche, 2008). It is computed according to Equation 3.

$$C(a_1, \dots, a_n) = \sum_{i=1}^{n} (a_{(i)} - a_{(i-1)}) w \left(A_{(i)} \right) \tag{3}$$

where (.) is a the permutation previously defined with $a_{(0)} = 0$ and $A_i = \{(i), \dots, (n)\}$, 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 *Choquet Integral* overcomes the limitations of the two previous operators by adding a weight on the combination. The weights are thus 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

[2] Only $(n-1)$ parameters have to be defined for n sources as their sum is 1.

of n information sources requires $2^n - 2$ coefficients. These are usually set by learning algorithms (Murillo et al., 2013).

2.3. Fuzzy inference systems

A fuzzy inference system, illustrated in Figure 2, usually requires more parameters than the former numerical aggregators.

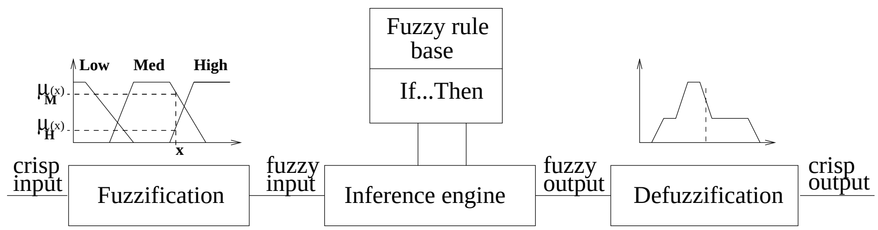

Figure 2: A fuzzy inference system.

To simplify the design the following choices have been made: the rule conclusions are aggregated using the Sugeno operator in the crisp case or using the centroid operator if fuzzy and the minimum is used to combine the individual membership degrees within a rule. The reader may refer to the *FisPro* documentation[3] for further details. Another important characteristics of this framework is that all the input variables are satisfaction degrees that share the same scale. These degrees are the output of the previous step, the abscissa of this figure is the ordinate of Figure 1. Regular grids with fully overlapping were used: the user can only chose the number of linguistic labels (Guillaume & Charnomordic, 2012). In this study, it was set at 2, *Low* and *High*, for all the input variables. An illustration of the automatically generated input partition is shown in Figure 3.

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 that the expert domain, the decision maker, is able to understand.

[3] `https://www.fispro.org`

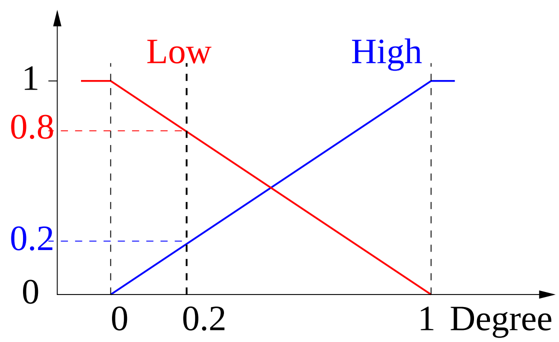

Figure 3: A regular fuzzy partition with two linguistic labels.

This way, the rule conclusions are easier to define than the *Choquet Integral* coefficients.

2.4. 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 target is available. Rule conclusions can also be learned using the *FisPro* software (Guillaume & Charnomordic, 2011).

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 Figure 4.

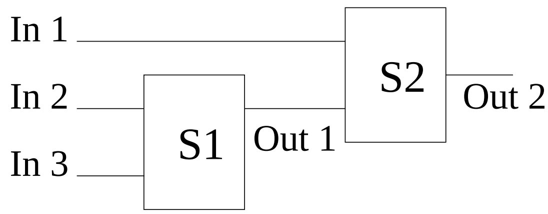

Figure 4: A hierarchical structure.

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 (Guillaume et al., 2013) as well other functionalities, such as a zone delineation algorithm (Pedroso et al., 2010). An illustration of its potential use in Precision Agriculture can be found in Leroux et al. (2018).

3. The soil chemical quality index

Multicriteria or multi-attribute analysis in agriculture shows a greater interest since 2005. The researches focused on sustainability assessment of production models, through sustainability comprehensive indicators that result of a hierarchy of partial indicators. They are basic inputs to design explicit and formal models to support decision-making about reorientation of farm activities. The main question to answer that involves multicriteria or multi-attributes decision making is: What are the best practices to maximize farm's productive potential? Soil plays a key role as it is fundamental natural resource to develop any agriculture activity. Therefore, 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 as[4]: “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.”

3.1. Architecture

The chemical index structure is shown in Figure 5. The *Soil pH* ranks first in the soil nutrient balance as it depends of the nature of the soil and it controls the chemical processes that take place in the soil; specifically, the availability of nutrients.

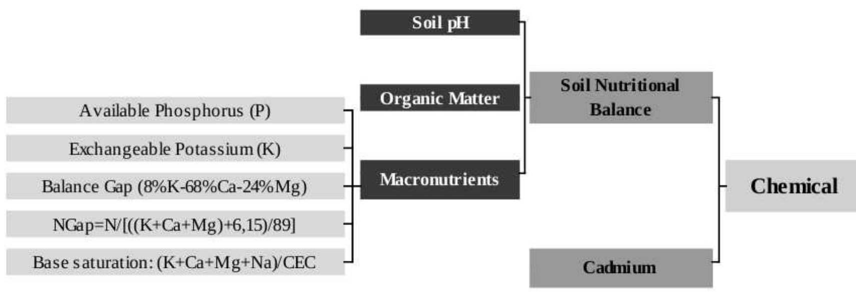

Figure 5: The chemical index architecture.

The *Macronutrients* variable is calculated from direct assessments, K and P; as well as a combination of soil chemical data as suggested by (Snoeck et al., 2016):

- K was put in first place because its availability in the soil is crucial given its role in cacao nutrition.
- P was placed in second position for the same reasons.
- The availability of the Ca and Mg cations is also very important. But it is more their balance with respect to K which is determinant; so, we put the balance for the calculation.

[4] `https://www.soils.org/publications/soils-glossary`

- N is less important for cacao because it was demonstrated that it enhances vegetative growth to the detriment of fruit development; hence, it is only recommended when N becomes a limiting factor.
- *Base saturation* was placed last. It helps to know the amount of fertilizer that can be brought.

Organic Matter has been set apart from other macronutrients because, although it brings Carbon that is an important macronutrient, its presence is very regular and usual in cacao plantations.

Finally, *Cadmium* is combined with *Soil Nutritional Balance* to yield the *Chemical* index. *Cadmium* was considered like an input variable in the last phase in the Chemical system because it is an accumulative heavy metal that nowadays is of important concern to cacao producing countries in Latin America for two main reasons. First, *Cadmium* in soil can be absorbed by cacao beans to become a source of contamination for consumers. Second, from an agronomic point of view, *Cadmium* is toxic to cacao plants as stated by Yadav (2010):

> When plants are exposed to high levels of *Cadmium* this causes reduction in photosynthesis, water uptake, and nutrient uptake. Plants grown in soil containing high levels of *Cadmium* show visible symptoms of injury reflected in terms of chlorosis, growth inhibition, browning of root tips, and finally death.

3.2. The experimental data

The study area is located in three municipalities of Tolima department in Colombia, these are, Chaparral (4^o $55'$ N, 75^o $07'$ E), Ortega (4^o $07'$N, 75^o $26'$E) in south and Mariquita (5^o $12'$N, 74^o $54'$E) in north of Tolima as shown in Figure 6.

Soils were sampled at depth $0 - 15$ cm (3 points per farm and 3 replicates in each one). The fieldwork was achieved between December 2018 and January 2019.

The data are shown in Table A.7.

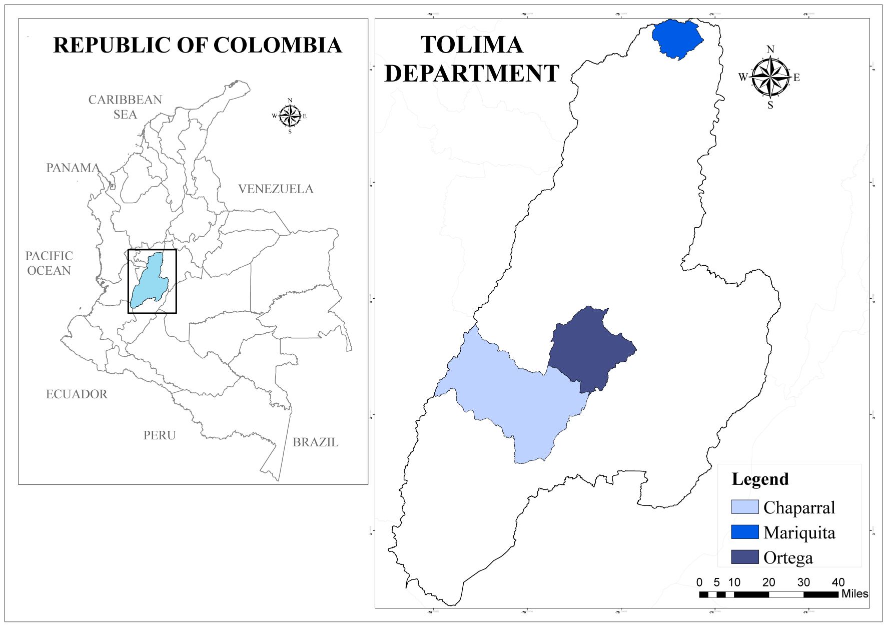

Figure 6: The study area: Three municipalities of the Tolima department, Colombia.

3.3. The index parameters

For each of the selected input variables, a transformation function to model the preference relation was defined. The shape of the membership function as well as its parameters were set according to the available literature. These data are summarized in Table 1.

This step in the modeling process comes to model expert knowledge about the behavior of each individual variable without taking into account the possible interactions. The semantics attached to the membership degree now reflects the level of satisfaction of the criterion, meaning the attribute together with the preference relation.

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.

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

Table 1: The input variables of the chemical index.

Input	Param	Ref.
Exchangeable Potassium (K)		(Rojas & Sacristán Sánchez, 2013)
(meq/100 g)	$0.2, 0.6$	(Arvelo Sánchez et al., 2017; Snoeck et al., 2016)
Available Phosphorus (P)		
(ppm)	$5, 15$	(Arvelo Sánchez et al., 2017; Snoeck et al., 2016)
(%K+%Ca+%Mg) BalanceGap		
(%)	$0, 0.5$	(Snoeck et al., 2016)
NGap		
(N/N_{target})	$0.5, 1, 1.5$	(Snoeck et al., 2016)
Base Saturation		
(%)	$0.4, 0.6$	(Snoeck et al., 2016)
Organic Matter		
(%)	$3, 5$	(Rojas & Sacristán Sánchez, 2013)
pH Soil		
(^{o}pH)	$5, 5.5, 6.5, 7.5$	(Rojas & Sacristán Sánchez, 2013)
Cadmium in Soil		
(ppm)	$0, 0.43$	(U.S. Environmental Protection Agency, 1996)

Table 2: The *Soil Nutritional Balance* rule base.

	pH	*OM*	*MacN*	Conclusion
1	Low	Low	Low	0.0
2	Low	Low	High	0.2
3	Low	High	Low	0.3
4	Low	High	High	0.5
5	High	Low	Low	0.4
6	High	Low	High	0.6
7	High	High	Low	0.7
8	High	High	High	1.0

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 2. The most important input variable is *Soil pH* (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 #5. If only *Organic Matter* (OM) or *Macronutrients* ($MacN$) 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 #4 involves the *High* label for OM and $MacN$ and the conclusion is $0.5 = 0.3 + 0.2$. 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, #8, is set at 1.

3.4. Results and discussion

The rule base was first tested and adjusted using the data from Chaparral (C). The results were checked by the agronomist who visited the farms. Then, the same rule base was used to compute the chemical index from the other locations, Ortega, O, and Mariquita, M. The results are given in Table 3.

For each location, the lowest score is printed in red and the highest one in blue ink. The first comment to be made is about the ranges: each location exhibits an important variability.This can be explained by the biological variability in soil composition but also by the impact of different agricultural practices, for instance fertilization. In Mariquita, all the farmers work with a commercial perspective, even in small areas, and thus fertilization tends to be more homogeneous. This leads to a smaller variation in the index. Most of the variability is explained by farm #7: in this particular case the poor score comes from a low content of *Organic Matter*, lower than the minimum required for Cacao production, and also a deficit in phosphorus and potassium.

The smallest index is assigned to $O10$: it is explained by the high level of *Cadmium* (0.647 mg/kg), a very acidic *Soil pH* (4.98) and a poor *Organic Mat-*

Table 3: The chemical index for the 30 farms.

	C	O	M
1	0.459	0.436	0.480
2	0.410	0.328	0.598
3	0.467	0.288	0.433
4	0.374	0.468	0.482
5	0.773	0.603	0.492
6	0.700	0.819	0.681
7	0.651	0.529	0.297
8	0.547	0.280	0.621
9	0.421	0.302	0.524
10	0.666	0.050	0.600
Mean	0.545	0.410	0.521
σ	0.14	0.21	0.11

ter content (2.33%). These three variables being the most important ones for the chemical quality of soil: plants can die by high level of Cadmium (Yadav, 2010), plant and microbial activity can be reduced when cacao plantations *Soil pH* is lower than 5 or higher than 7.5 and *Organic Matter* lower that 3% limits soil fertility and stability while increasing susceptibility to erosion. This is is especially important for this farm that located on a hill slope: the elevation is increased by 50% between the lowest and the highest sampling points (Bautista Cruz et al., 2004; García & Ramírez, 2012; Gutiérrez D. et al., 2018). These factors limit this crop's productivity potential because chemical properties are essential to soil and plant relationship, then, the index states that in this farm the soil is marginally suitable for cacao.

The farms with the highest index in the three locations, $C5$, $O6$ and $M6$, get a high value because they do not have any problem with these three key variables. The degrees are all higher than 0.7 except *Organic Matter* for $C5$ and *Soil pH* for $M6$: the two of them are 0.52. However, the $C5$ and $M6$ degrees for

Macronutrients are 0.47 and 0.38, respectively. They could be clearly improved using an appropriate fertilization plan.

These results were analyzed jointly with the producers. The basic *pH* for $C1$ (see Table A.7) is due to Cal application. Cal is the input most frequently freely provided by state agencies to all farmers. But, instead of improving the cacao productivity this input limits it even more. The most acidic *pH* in the municipality was found at $C9$. This is explained by the farm location: the access conditions (distance, type of road and time) restrict the supply of external inputs (donated or purchased). In Mariquita, the macronutrient balance is low for farms that most use fertilizers. This is possibly explained by a common practice: the doses are not based on equilibrium relationship between elements, but rather on standard thresholds established for each element independently, which modifies natural balance of the soil for cacao negatively. Cadmium (Cd) content confirmed what was expressed by the farmers prior to conducting the survey: Chaparral zone is the most threatened of the three studied. However, $O10$ drew also attention by its high level. It was found that recently, on the edge of the cacao parcel, the road was paved. This use of petroleum derivatives may be a possible contamination source as suggested by (Mite et al., 2010).

High scores in soil chemical properties do not ensure a good yield because soil fertility also depends on biological and physical properties, which are more restrictive in the short term.

This short analysis highlights the transparency of the whole process: any intermediate score makes sense and the final result is easy to analyze. The low scores in the previous steps come from criteria that can be improved.

4. Comparison with other aggregation operators

The *Soil Nutritional Balance* component of the *Chemical* index is used in this section to investigate if an equivalent aggregation operator to the fuzzy system with the rule base shown in Table 2 can be designed.

The output of the fuzzy system for the 10 farms from Chaparral is used

Table 4: Comparison of the aggregation operators using the 10 farms from Chaparral.

	pH	OM	$MacN$	FIS	WAM	OWA	CI_E	CI_L
1	0.50	1	0.70	0.655	0.731	0.730	0.689	0.655
2	0.76	0.87	0.41	0.586	0.666	0.660	0.689	0.607
3	1	0	0.34	0.467	0.437	0.447	0.467	0.463
4	1	0	0.67	0.534	0.561	0.520	0.534	0.533
5	1	0.53	0.47	0.675	0.653	0.680	0.699	0.687
6	1	0.86	0.48	0.770	0.764	0.759	0.803	0.749
7	1	0.36	0.56	0.652	0.632	0.641	0.654	0.651
8	1	0	0.77	0.555	0.599	0.543	0.555	0.555
9	0.60	0	0.38	0.309	0.328	0.307	0.316	0.315
10	1	0	0.61	0.523	0.540	0.508	0.523	0.522
R^2				-	0.912	0.936	0.953	0.993

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 4.

The degrees to aggregate are in the first three columns OM, *pH* and $MacN$), followed by the output inferred by the fuzzy system (FIS). The remaining columns of Table 4 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* yielded the poorest result. The weights for OM, *pH* and $MacN$ were: 0.317, 0.312 and 0.371. The score for the first farm was similar to the one for farm #6 and higher than the one for farm #5. This result was not expected as soil of farm #1 had a less suitable *Soil pH* than farm #6.

Farm #2 was also assigned a high score, higher than the ones of farms #5 and #7. This was not expected as the degrees of *Soil pH* and, to a lower extent, of *Macronutrients* were 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, were: 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 got an higher score than farm #7.

The reasoning used to design the FIS rule base is now applied to define the coefficients of the *Choquet Integral*. They are given in Table 5.

Table 5: The expert weights of the *Choquet Integral*.

pH	OM	$MacN$	pH-OM	pH-$MacN$	OM-$MacN$
0.4	0.3	0.2	0.7	0.6	0.5

The index computed using this fuzzy measure is reported in column CI_E of Table 4. The determination coefficient is improved. 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 results were quite similar to the ones yielded by the FIS except for farm #2 that was still given a higher score than farm #7.

This was not expected as the values for pH and $MacN$ were higher for farm #7. Farms #1 and #2 were given the same output value. Even if the first one had a slightly less optimal pH, its OM and $MacN$ were better, this combination should yield a higher score for Farm #1.

The best results are given by the *Choquet Integral* when the weights are learned using the FIS output as a target. They are reported in column CI_L of Table 4. The weights that minimize the least squares of errors between the target and the output, yielded by the HLMS (Heuristic Least Mean Squares) algorithm (Murillo et al., 2013), are really different from the ones defined using expert reasoning. They are shown in Table 6.

Table 6: The weights of the *Choquet Integral* learned by the HLMS algorithm using the 10 farms from Chaparral.

pH	OM	$MacN$	pH-OM	pH-$MacN$	OM-$MacN$
0.392	0.000	0.000	0.558	0.603	0.795

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 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 Murillo et al. (2013). Any value lower or equal to 0.603, the minimum value for a coalition that includes $MacN$, 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 that 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 optimal tuning remains difficult without training.

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 a 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* available at: `https://www.geofis.org`. 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.

A part of the available data, 10 farms, was used to tune and calibrate the system, the other 20 ones were used as a validation sample. 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.

Appendix A.

For the sake of completeness, the data, Table A.7, as well as the aggregation operators that are not studied in this work are given in this appendix. They are the *Macronututrients* rule base, Table A.8, and the Chemical final step subsystem WAM aggregator.

The Chemical subsystem operator is a WAM with the following weights:

- Soil Nutritional Balance: 0.7
- Cadmium: 0.3

References

Aravindh, S., Chinnadurai, C., & Balachandar, D. (2019). Development of soil biological quality index for soils of semi-arid tropics. *SOIL Discussions*, *2019*, 1–24. URL: `https://soil.copernicus.org/preprints/soil-2019-60/`. doi:`10.5194/soil-2019-60`.

Arthur, A., Afrifa, A. A., & Dogbatse, J. A. (2017). Assessment of soil fertility status of cocoa farms around the ankasa national park in the jomoro district of the western region of ghana. In *International Symposium on Cocoa Research (ISCR)* (pp. 1–29). URL: `https://www.icco.org/about-us/international-cocoa-agreements/cat_view/68-icco-workshops-a`

Arvelo Sánchez, M., González León, D., Delgado López, T., Maroto Arce, S., & Montoya Rodríguez, P. (2017). Estado actual sobre la producción, el comercio y cultivo del cacao en américa. Instituto Interamericano de Cooperación para la Agricultura. URL: `repositorio.iica.int/bitstream/11324/6422/1/BVE18019631e.pdf`.

Bautista Cruz, A., Etchevers Barra, J., Del Castillo, R. F., & Gutierrez, C. (2004). La calidad del suelo y sus indicadores. *Ecosistemas*, *13*, 90 – 97.

Choquet, G. (1954). Theory of capacities. *Annales de l'Institut Fourier*, *5*, 131–295. doi:`10.5802/aif.53`.

Table A.7: The dataset

	Cd	pH	OM	P	K	Ca	Mg	N	Na	CEC	MacN
C1	1.19	7.0	6.4	77.9	0.49	21.7	4.3	0.31	0.21	27.8	0.70
C2	1.08	6.7	4.7	2.8	0.35	15.5	4.2	0.23	0.23	20.1	0.41
C3	0.23	5.6	2.8	2.7	0.25	8.4	2.2	0.14	0.20	11.0	0.34
C4	1.22	5.5	2.7	7.6	0.66	6.1	1.1	0.13	0.16	8.6	0.67
C5	0	6.0	4.0	5.0	0.49	10.2	4.7	0.20	0.21	15.6	0.47
C6	0.20	6.0	4.7	4.8	0.49	11.5	5.1	0.23	0.17	17.6	0.48
C7	0.15	5.7	3.7	2.3	0.67	14.0	6.8	0.18	0.21	21.8	0.56
C8	0.20	6.0	2.5	25.8	0.66	12.7	3.0	0.12	0.24	18.1	0.77
C9	0.14	5.3	3.0	3.9	0.28	9.5	3.1	0.14	0.18	12.6	0.38
C10	0	5.6	2.9	2.8	1.01	8.2	3.3	0.14	0.21	12.8	0.61
O1	0.05	5.2	2.4	5.0	0.20	9.5	3.0	0.12	0.24	13.2	0.27
O2	0.08	4.6	2.9	8.2	0.31	4.4	1.4	0.14	0.23	7.6	0.60
O3	0.18	5.0	2.7	6.1	0.38	5.3	2.5	0.13	0.24	8.8	0.48
O4	0.25	5.6	2.7	12.6	0.29	18.7	11.7	0.13	0.26	30.9	0.46
O5	0.05	5.7	2.4	5.8	0.36	7.9	3.9	0.12	0.24	14.9	0.43
O6	0.12	6.0	6.2	28.8	0.24	5.8	1.8	0.30	0.22	8.5	0.55
O7	0.04	5.2	3.3	100	0.32	7.3	2.0	0.16	0.24	9.7	0.68
O8	0.09	4.4	1.1	6.2	0.17	2.3	0.7	0.05	0.22	3.9	0.30
O9	0.08	4.5	2.4	1.8	0.26	2.5	0.8	0.12	0.24	8.0	0.42
O10	0.65	5.0	2.3	7.9	0.19	5.9	2.9	0.11	0.23	9.5	0.36
M1	0.11	4.9	12.9	2.5	0.27	2.1	0.8	0.63	0.25	3.9	0.34
M2	0.17	5.2	15.4	10.6	0.72	9.3	1.6	0.75	0.28	11.3	0.64
M3	0.25	4.9	6.2	9.8	0.73	5.6	2.0	0.30	0.35	8.9	0.71
M4	0.10	4.6	14.2	2.9	0.22	1.1	0.4	0.68	0.33	3.1	0.28
M5	0.15	4.6	6.6	7.2	0.88	4.4	1.2	0.32	0.24	7.9	0.62
M6	0.07	5.3	16.3	2.2	0.33	2.4	0.8	0.79	0.24	4.2	0.38
M7	0.12	5.0	2.9	3.6	0.15	4.4	3.0	0.14	0.25	7.9	0.26
M8	0.12	5.2	18.8	4.4	0.36	6.8	1.0	0.91	0.29	8.4	0.32
M9	0.26	5.5	3.5	8.2	0.66	6.9	2.8	0.17	0.26	10.1	0.75
M10	0.05	5.2	4.7	3.2	0.19	5.7	1.4	0.22	0.28	7.9	0.22

Table A.8: The Macronutrients rule base

Balance Gap	K	P	N Balance	Base Saturation	Conclusion
Low	Low	Low	Low	Low	0
Low	Low	Low	Low	High	0.1
Low	Low	Low	High	Low	0.15
Low	Low	Low	High	High	0.2
Low	Low	High	Low	Low	0.25
Low	Low	High	Low	High	0.35
Low	Low	High	High	Low	0.4
Low	Low	High	High	High	0.45
Low	High	Low	Low	Low	0.3
Low	High	Low	Low	High	0.4
Low	High	Low	High	Low	0.45
Low	High	Low	High	High	0.5
Low	High	High	Low	Low	0.55
Low	High	High	Low	High	0.65
Low	High	High	High	Low	0.7
Low	High	High	High	High	0.75
High	Low	Low	Low	Low	0.4
High	Low	Low	Low	High	0.3
High	Low	Low	High	Low	0.4
High	Low	Low	High	High	0.45
High	Low	High	Low	Low	0.5
High	Low	High	Low	High	0.55
High	Low	High	High	Low	0.65
High	Low	High	High	High	0.7
High	High	Low	Low	Low	0.55
High	High	Low	Low	High	0.6
High	High	Low	High	Low	0.7
High	High	Low	High	High	0.75
High	High	High	Low	Low	0.8
High	High	High	Low	High	0.85
High	High	High	High	Low	0.9
High	High	High	High	High	1

Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators*, *46*, 138 – 148. doi:10.1016/j.ecolind.2014.06.011.

De Laurentiis, V., Secchi, M., Bos, U., Horn, R., Laurent, A., & Sala, S. (2019). Soil quality index: Exploring options for a comprehensive assessment of land use impacts in lca. *Journal of Cleaner Production*, *215*, 63 – 74. URL: http://www.sciencedirect.com/science/article/pii/S095965261833960X. doi:https://doi.org/10.1016/j.jclepro.2018.12.238.

Duan, X., Xie, Y., Ou, T., & Lu, H. (2011). Effects of soil erosion on long-term soil productivity in the black soil region of northeastern china. *CATENA*, *87*, 268 – 275. URL: http://www.sciencedirect.com/science/article/pii/S0341816211001329. doi:https://doi.org/10.1016/j.catena.2011.06.012.

Duan, X.-w., Xie, Y., Feng, Y.-j., & Yin, S.-q. (2009). Study on the method of soil productivity assessment in black soil region of northeast china. *Agricultural Sciences in China*, *8*, 472 – 481. URL: http://www.sciencedirect.com/science/article/pii/S1671292708602345. doi:https://doi.org/10.1016/S1671-2927(08)60234-5.

Ferraro, D. O. (2009). Fuzzy knowledge-based model for soil condition assessment in argentinean cropping systems. *Environmental Modelling and Software*, *24*, 359 – 370. URL: http://www.sciencedirect.com/science/article/pii/S1364815208001357. doi:https://doi.org/10.1016/j.envsoft.2008.07.006.

García, Y., & Ramírez, W. (2012). Soil quality indicators: A new way to evaluate this resource. *Pastos y Forrajes*, *35*, 125 – 138.

Grabisch, M., & Labreuche, C. (2008). A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *4OR*, *6*, 1–44. doi:10.1007/s10288-007-0064-2.

Gu, Z., Xie, Y., Gao, Y., Ren, X., Cheng, C., & Wang, S. (2018). Quantitative assessment of soil productivity and predicted impacts of water erosion in the black soil region of northeastern china. *Science of The Total Environment*, *637-638*, 706 – 716. URL: `http://www.sciencedirect.com/science/article/pii/S0048969718316966`. doi:`https://doi.org/10.1016/j.scitotenv.2018.05.061`.

Guillaume, S., Bates, T., Lablée, J.-L., Betts, T., & Taylor, J. (2020). Combining spatial data layers using fuzzy inference systems: Application to an agronomic case study. In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2020)* (pp. 62–71). Prague, Czech Republic: SCITEPRESS.

Guillaume, S., & Charnomordic, B. (2011). Learning interpretable fuzzy inference systems with fispro. *Information Sciences*, *181*, 4409–4427. doi:`10.1016/j.ins.2011.03.025`.

Guillaume, S., & Charnomordic, B. (2012). Fuzzy inference systems: an integrated modelling environment for collaboration between expert knowledge and data using fispro. *Expert Systems with Applications*, *39*, 8744–8755. doi:`10.1016/j.eswa.2012.01.206`.

Guillaume, S., Charnomordic, B., & Loisel, P. (2013). Fuzzy partitions: a way to integrate expert knowledge into distance calculations. *Information Sciences*, *245*, 76–95. doi:`10.1016/j.ins.2012.07.045`.

Gutiérrez D., J., Cardona, W., & Monsalve C., O. (2018). Potencial en el uso de las propiedades químicas como indicadores de calidad de suelo. una revisión. *Revista Colombiana de Ciencias Hortícolas*, *11*, 450–458. doi:`10.17584/rcch.2017v11i2.5719`.

Kelting, D. L., Burger, J. A., Patterson, S. C., Aust, W., Miwa, M., & Trettin, C. C. (1999). Soil quality assessment in domesticated forests – a southern pine example. *Forest Ecology and Management*, *122*, 167 – 185. URL:

http://www.sciencedirect.com/science/article/pii/S0378112799000407. doi:https://doi.org/10.1016/S0378-1127(99)00040-7.

Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., Naud, O., Crestey, T., Lablee, J.-L., & Tisseyre, B. (2018). Geofis: An open source, decision-support tool for precision agriculture data. *Agriculture*, *8*, 73. doi:10.3390/agriculture8060073.

Liu, X., Herbert, S. J., Hashemi, A. M., Zhang, X., & Ding, G. (2006). Effects of agricultural management on soil organic matter and carbon transformation - a review. *Plant, Soil and Environment*, *52*, 531–543. doi:https://doi.org/10.17221/3544-pse.

Mendoza, G., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, *230*, 1 – 22. doi:10.1016/j.foreco.2006.03.023.

Mendoza, G. A., Bare, B. B., & Zhou, Z. (1993). A fuzzy multiple objective linear programming approach to forest planning under uncertainty. *Agricultural Systems*, *41*, 257 – 274. doi:10.1016/0308-521X(93)90003-K. Multiple Criteria Analysis in Agricultural Systems.

Mite, F., Carrillo, M., & Durango, W. (2010). Avances del monitoreo de presencia de cadmio en almendras de cacao, suelos y aguas en ecuador. In *Proceedings of XII Congreso ecuatoriano de la ciencia del suelo* (pp. 1–21). URL: http://www.secsuelo.org/wp-content/uploads/2015/06/6.-Francisco-Mite.-Cadmio.-INIAP.pdf.

Mora-Herrera, D. Y., Guillaume, S., Snoeck, D., & Escobar, O. Z. (2020). Fuzzy inference system as an aggregation operator - application to the design of a soil chemical quality index. In M.-J. L. et al. (Ed.), *Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2020, CCIS 1238* (pp. 1–10). Switzerland: Springer Nature. doi:10.1007/978-3-030-50143-3_35.

Mukherjee, A., & Lal, R. (2014). Comparison of soil quality index using three methods. *PLoS ONE*, *9*, e105981. doi:https://doi.org/10.1371/journal.pone.0105981.

Murillo, J., Guillaume, S., Tapia, E., & Bulacio, P. (2013). Revised hlms: A useful algorithm for fuzzy measure identification. *Information Fusion*, *14*, 532–540. doi:10.1016/j.inffus.2013.01.002.

Nosrati, K., & Collins, A. L. (2019). A soil quality index for evaluation of degradation under land use and soil erosion categories in a small mountainous catchment, iran. *Journal of Mountain Science*, *16*, 2577–2590. doi:10.1007/s11629-019-5567-8.

Pedroso, M., Taylor, J., Tisseyre, B., Charnomordic, B., & Guillaume, S. (2010). A segmentation algorithm for the delineation of management zones. *Computer and Electronics in Agriculture*, *70*, 199–208. doi:10.1016/j.compag.2009.10.007.

Rojas, F., & Sacristán Sánchez, E. J. (2013). Guía ambiantal para el cultivo del cacao. Ministerio de Agricultura y Desarrollo Rural y Federación Nacional de Cacaoteros, Colombia. URL: https://www.fedecacao.com.co/portal/images/recourses/pub_doctecnicos/fedecacao-pub-doc_05

Snoeck, D., Koko, L., Joffre, J., Bastide, P., & Jagoret, P. (2016). Cacao nutrition and fertilization. In E. Lichtfouse (Ed.), *Sustainable Agriculture Reviews: Volume 19* (pp. 155–202). Cham: Springer International Publishing. doi:10.1007/978-3-319-26777-7_4.

Thakur, N., & Sharma, R. (2019). Soil quality. *International Journal of Current Microbiology and Applied Sciences*, *8*, 2920–2930. doi:10.20546/ijcmas.2019.806.351.

Thoumazeau, A., Bessou, C., Renevier, M.-S., Trap, J., Marichal, R., Mareschal, L., Decaëns, T., Bottinelli, N., Jaillard, B., Chevallier, T., Suvannang, N., Sajjaphan, K., Thaler, P., Gay, F., & Brauman, A. (2019). Biofunctool®: a new framework to assess the impact of land management on soil quality. part a: concept and validation of the set of indicators. *Ecological Indicators*, *97*, 100–110. doi:https://doi.org/10.1016/j.ecolind.2018.09.023.

Torra, V. (1996). Weighted owa operators for synthesis of information. In *Proceedings of IEEE 5th International Fuzzy Systems* (pp. 966–971). volume 2. doi:10.1109/FUZZY.1996.552309.

U.S. Environmental Protection Agency (1996). Soil screening guidance: User's guide. URL: https://www.nrc.gov/docs/ML0824/ML082480252.pdf.

Vogel, H.-J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., & Wollschläger, U. (2019). Quantitative evaluation of soil functions: Potential and state. *Frontiers in Environmental Science*, *7*, 164. URL: https://www.frontiersin.org/article/10.3389/fenvs.2019.00164. doi:10.3389/fenvs.2019.00164.

Yadav, S. (2010). Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. *South African Journal of Botany*, *76*, 167 – 179. doi:10.1016/j.sajb.2009.10.007.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*, 183–190. doi:10.1109/21.87068.

Highlights

The index is used to help cacao farmers to make decisions

It is designed from scientific knowledge and tested using data from Colombia

Fuzzy inference system is used as an aggregation operator

The whole process is implemented in an open source software, GeoFIS

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.