TITLE OF THE SURVEY:

Integrating Agricultural Sustainability into policy planning: a geo-referenced framework based on Rough Set theory

AUTHORS:

Eugenio DEMARTINI[a], Anna GAVIGLIO[a], Danilo BERTONI[b]

[a]Department of Health, Animal Science and Food Safety – VESPA

University of Milan – Via Celoria, 10 – 20133 – Milano [MI] – Italy

[b]Department of Economics, Management and Quantitative Methods – DEMM

University of Milan – Via Celoria, 2 – 20133 – Milano [MI] – Italy

E-MAIL, TELEPHONE AND FAX NUMBER:

eugenio.demartini@unimi.it - tel. 0039-0250316462 - fax 0039-0250316486

anna.gaviglio@unimi.it - tel. 0039-0250316493 - fax 0039-0250316486

danilo.bertoni@unimi.it - tel. 0039-0250316465 - fax 0039-0250316486

MAIN CORRESPONDING AUTHOR:

Eugenio Demartini: eugenio.demartini@unimi.it

SUMMARY:

We propose a geo-referenced framework for agricultural sustainability assessment aimed at supporting policy planning. The framework is based on Rough Set theory and (i) integrates the three pillars of sustainability; (ii) proposes an easy measurement of agricultural systems' ability to resist over time (agricultural resilience); (iii) offers easy-to-read results; and, (iv) reduces the gap between researchers' analytical skills and decision-makers' needs. In the paper, a part of the framework, we present essential and practical notion of Rough Set theory and a case study based on Lombardy Region (Italy). Finally, some values and lacks of the method are discussed.

Integrating Agricultural Sustainability into policy planning: a geo-referenced framework based on Rough Set theory

1. Introduction

Policy-makers frequently use the term "sustainability" when declaring their objectives without taking into account the technical limitations this concept implies for the design of public intervention. As already underlined by various researchers, the problem resides in the need for a general and political definition of sustainability in agricultural, scientific, and analytical praxis (Francis et al., 1989; Pretty, 1995; Hansen 1996). This is because:

1. no unit can directly measure human well-being resulting from agricultural activity (McAllister, 1980);
2. economic profit, social welfare and environmental conservation, the three pillars of sustainability, can not be maximized contemporaneously due to the trade-offs between them (Brown *et al.*, 2001; Gaviglio *et al.* 2012);
3. the agricultural system is extremely heterogeneous by nature and includes different scales of analysis (Smit and Smithers, 1993);
4. today we are studying how to preserve resources for future generations, but today we cannot verify the reliability of our results (Gómez-Limón and Sanchez-Fernandez, 2010); and,
5. considering the anthropocentric focus of our studies, the goals of sustainability analysis change according to different stakeholders' points of views, so what is sustainable for one person, might actually be unsustainable for another.

Despite these difficulties, the concept of sustainability is widespread in agricultural science and researchers have developed two main interpretative schemes for it: the goal-prescribing and system-describing models (Hansen, 1996). According to the goal-prescribing model, agricultural sustainability is considered an alternative approach to agriculture; in this case, a scientists' work is focused on techniques that should improve agricultural sustainability. Alternatively, the system-describing model looks at sustainability as a (set of) feature(s) of agricultural activities. This model measures a "state" of sustainability, so it appears useful for identifying strengths and weakness of agricultural systems, helping in decision-making rather than indicating operative solutions. These two frameworks have stimulated the growth of literature on the assessment agricultural sustainability, but further efforts are still required for the development of new interpretive methods for its measurement, especially as regards its integration into policy planning (Gómez-Limón and Riesgo, 2009; Gómez-Limón and Sanchez-Fernandez, 2010).

The present paper contributes to the scientific discussion of this issue, proposing a geo-referenced framework for sustainability analysis based on the potential for approximate classification of data and information induction of Rough Set theory (RST, Pawlak, 1982). The initial assumption was that policy-makers cannot consider all the determining factors of sustainability, but they do have a correct basic understanding of it. It would therefore be helpful for them to have a tool that provides a summary of relevant issues in order to support decision-making. The "ideal" solution presented consists of a framework which: (i) integrates the three pillars of sustainability; (ii) proposes a simple measurement of a given agricultural system's ability to resist over time (agricultural resilience); (iii) offers easy-to-read results; and, (iv) reduces the gap between the analytical skills of researchers and the needs of decision-makers. In this respect the present paper introduces some novelties into the debate regarding the assessment of agricultural sustainability. The first is the presentation of Rough Set Theory as a methodical option to achieve these aims. Secondly, a simple and intuitive definition and interpretation of agricultural sustainability is proposed and discussed. Finally, the work is structured in order to illustrate the basics of RS Theory and develop some practical skills in its use.

The remainder of the text is organized into four sections. Section 2 presents the features of RST and reviews the literature on its applications in agricultural science. Section 3 presents materials and methods for the territorial case study of Lombardy (Italy), while the results and discussion are set out in Section 4. Finally, a concluding paragraph offers a summary of the proposed framework and some reflections on the potentialities and limitations of RST.

2. Rough Set theory for dataset analysis and its application in agricultural science

Scientific models do not always achieve satisfactory solutions for complex problems. Flawed results can easily be generated due to analytical problems like datasets inconsistencies and statistical constraints. In the early 1980s, the Polish professor Zdzisław I. Pawlak proposed a mathematical tool that could deal effectively with these two issues (Pawlak, 1982). He called his model Rough Set theory (RST), because it involves the partition of a set of items under study into subsets according to equalities within them, and an assessment of the overlapping portions (rough sets) which represent the inconsistencies of the database (see Figure 1 and its description in section 2.1.2 for further explanations).

Since its original formulation, the RST model has been successfully applied in descriptive and predictive procedures (Stefanowski, 2007). It helps describe regularities within data, uncovering hidden information and suggesting interpretation of dependencies between observed variables. It can be used as a technique for machine learning, knowledge discovery, and inductive inference (Pawlak, 1997) with valuable performance in data reduction, pattern recognition, data significance estimation, cause-effect link detection, automatic classification, and similarity/dissimilarity evaluation (Pawlak *et al.*, 1995).

The basic notions of RS theory and its utility will be discussed in the following paragraphs, with a brief review of applications in agricultural science at the end of the Section.

2.1. The Rough Set model

2.1.1 Basic notation and definitions

In Rough Set theory[1], data are organized in an information system $S = \langle U, Q, V, \rho \rangle$ composed of:

- U, the set of x objects described by a Q set of q attributes, that can be divided in *condition* attributes (set $C \neq \emptyset$) and *decision* attributes (set $D \neq \emptyset$), such that $C \cup D = Q$ and $C \cap D = \emptyset$. By definition, decision attributes split objects into sets pertaining to different decision classes $\{K_j : j = 1, \dots, k\}$
- $V = \bigcup_{q \in Q} V_q$, is the value set of the q attribute;
- $\rho(x, q): U \times Q \to V$, a total function such that $\rho(x, q) \in V_q, \forall\, x \in U, q \in Q$, called the *information function*.

RS induces information from this structure applying the *indiscernibility relation*, which states that given a non-empty subset of attributes $A \subseteq Q$, two objects $x_1, x_2 \in U$ and $\rho(x, a)$ defined as the value of attribute a taken by the object x, the objects are indiscernible if $\{(x_1; x_2) \in U \times U, \rho(x_1, a) = \rho(x_2, a), \forall a \in A\}$ and writing $x I_A y$. Indiscernible objects for particular values of a create subset of x objects in S; we call each of these subsets an *elementary set* in S or elementary class of equivalence, denoted by $I_A(x)$. Moreover, any finite union of elementary sets is called a *definable set*, and the entire family of equivalence classes of relation constructed over $x \in U$ (i.e. the union of all definable sets) is denoted by $U/I(A)$.

A hypothetical example related to determinants of adoption of biogas technologies by breeding farmers helps to present the method. The *decision table* in Table 1 represents information about the q characteristics of x farms and the decision (output) variable d, which states whether breeding farmers have or not installed a biogas plant. In this information system there are six objects (farms), three attributes (size of the farm, age of farmer, and type of breeding farm), and one decision attribute (decision about installation of biogas plant).

[1]The explanation of Rough Set Theory presented in paragraphs 3.1.1 and 3.1.2 follows Pawlak *et al.* (1995), Stefanowski (2007) and Slowinski et al. (2011). Researchers who would like to further investigate the formal characteristics of the method, and its early applications and developments, refer to Pawlak (1982), Kryszkiewicz, M. (1998), Yao (1998), and Pawlak and Skowron (2007).

Given the subset of attributes $A = \{Size, Age\}$, it is possible to find the following elementary sets: $\{x_1, x_2\}$, $\{x_3\}$, $\{x_4\}$, and $\{x_5, x_6\}$, and define the definable set $\{x_1, x_2, x_3, x_4, x_6\}$ by combinations of the attributes $Size = \{Big\}$ and $Age = \{Old\}$, $Size = \{Normal\}$ and $Age = \{Old\}$, $Size = \{Normal\}$ and $Age = \{Young\}$, or $Size = \{Small\}$ and $Age = \{Old\}$.

In order to reduce data and extract information, the *reducts* of attribute definitions are required. Considering the new set of attributes $B = \{Size, Age, Type\}$, the elementary sets are singletons $\{x_1\}$,$\{x_2\}$,$\{x_3\}$,$\{x_4\}$,$\{x_5\}$, and$\{x_6\}$, and they remain singleton if Age is removed by S. All the attributes that behave like the attribute Age can be considered redundant; the remaining subset of S whith no redundant attributes $P = \{Size, Type\}$ is called a *minimal set*. Furthermore, since P defines the same elementary set as B, we define P as a *reduct* of B, or we say P is covering B.

2.1.2 Rough Set theory, rule induction and data inconsistencies

Now the main problem is considered when dealing with data like that in Table 2, where the original information is reduced to the non-redundant attributes of $P = \{Size, Type\}$. In this step of the study, the aim is to discover which farm and farmer characteristics are linked to biogas plant installation, in order, for example, to forecast the likelihood of adoption of this technology in new agricultural areas.

First of all, X *decision classes* need to be constructed as the elementary sets of objects on the basis of decision attributes; in Table 2 these are $\{x_1, x_2, x_3\}$ and $\{x_4, x_5, x_6\}$; secondly, it is observed that all the elements of $U/I(P)$ are represented in one of the two classes, i.e. the decision "$Biogas$ equal to Yes" or "$Biogas$ equal to No" depends on the attributes $Size$ and $Type$ and neither of these two is redundant. Finally, the relations between decision and attributes can be expressed in the form of a lexical rule r such as “if P then d”. Note that this rule can be split into two parts:

- "if P …" represents the condition, i.e. the value of the q attributes pertaining to P, under which one object can be assigned to a certain decision class; and,
- " … then d" is the decision part, stating which decision class the object pertains to.

Thus, from Table 2 the following rules can be derived:

- if $Size = \{Big\}$, then $Biogas = \{Yes\}$;
- if $Size = \{Normal\}$ and $Type = \{Swine\}$, then $Biogas = \{Yes\}$;
- if $Size = \{Small\}$, then $Biogas = \{No\}$;
- if $Size = \{Normal\}$ and $Type = \{Cattle\}$, then $Biogas = \{No\}$.

The system information in Table 2 is consistent. The rules represent all the objects in U and there is no intersection between elementary sets and decision classes, i.e. each elementary set is a subset of some decision class. Unfortunately, in reality databases are often inconsistent, as if two more objects were being considering, x_7 and x_8, as in Table 3. In this dataset new elementary sets and decision classes are created, these being respectively $\{x_1, x_8\}$, $\{x_2\}$, $\{x_3, x_7\}$, $\{x_4\}$, $\{x_5\}$,and $\{x_6\}$ the elementary sets, and $\{x_1, x_2, x_3, x_8\}$ and $\{x_4, x_5, x_6, x_7\}$ the two decision classes. As shown in Figure 1, the pair of indiscernible objects $\{x_3, x_7\}$ for attributes $Size$ and $Type$ is not a subset of any decision class and represents an inconsistency that can be managed by RS theory.

RST attempts to calculate the greatest and the least *definable sets* for each X decision classes. The former is called the *lower approximation* and denoted $\underline{P}(X)$, while the latter is the *upper approximation* of X and denoted $\overline{P}(X)$. The subtraction $BN_P = \underline{P}(X) - \overline{P}(X)$ defines a particular set, called the *boundary region of* X. For example, from Table 3, considering $d = \{Yes\} \Rightarrow X = \{x_1, x_2, x_3, x_8\}$, then $\underline{P}(X) = \{x_1, x_2, x_3, x_8\}$ represents the definable set containing all the objects that can certainly be assigned to $d = \{Yes\}$, while $\overline{P}(X) = \{x_1, x_2, x_3, x_7, x_8\}$ is the union of the elementary sets defined by P that have no-empty intersection with X (such a pair of subsets represents the "rough set" the theory derives its name from), finally $BN_{d=\{Yes\}} = \{x_7\}$ can be computed. Despite the inconsistencies, rules can still be induced, but in this case, RS separates *certain* from *approximate* rules. The former are induced from lower approximations, the latter, instead, are induced from boundaries of decision classes. In Table 3, the certain rules are:

- if $Size = \{Big\}$, then $Biogas = \{Yes\}$;
- if $Size = \{Small\}$, then $Biogas = \{No\}$;
- if $Size = \{Normal\}$ and $Type = \{Cattle\}$, $Biogas = \{No\}$;

While the approximate rule is:

- if $Size = \{Normal\}$ and $Type = \{Swine\}$, then $Biogas = \{Yes\}$ or $Biogas = \{No\}$.

2.1.3 Rule extraction algorithms, data preprocessing, and model evaluations

Depending on whether the attributes that compose an information system are continuous or discretized, software packages can use different algorithms to induce information. A wide range of algorithms have been proposed, as reviewed by Stefanowski (1998) and Thangavel and Pethalakshmi (2009). In the present research the LEM2 (Grzymala-Busse, 1992) and modLEM (Stefanowski, 1998) algorithms were used in order to process information systems characterized by categorized or continuous attributes, respectively. They were chosen as the most popular rule induction techniques, with proven good performance in RST exercises. They both produce *minimal sets* of decision rules, which guarantee identification of the smallest number of rules explaining relations between attributes and decision variables applicable to all objects.

RST was applied to a categorized and a continuous variables dataset. In the case of categorized variables, the *recursive minimal entropy partitioning* algorithm proposed by Fayyad and Irani (1992) was applied to calculate the boundaries that split the attribute domains into classes in order to guarantee minimum class entropy considering all boundaries.

Note that, since the algorithm does not contain any constrains regarding the width of classes, the partitions created can be very asymmetric (Obersteiner and Wilk, 1999).

Normally RST models come together with a validation analysis that measures how the extracted rules fit to the original data. In the present research, validation was used to identify the best model. When dealing with small datasets, the leave-one-out method is prescribed among the numerous cross-validation techniques. Given a dataset of a number n of objects, this method performs an iterative and averaged measurement of fitting errors to the model. Operationally, the model is calculated n separate times using all the objects except for one and a prediction is performed for the excluded object. The method is iterative in the sense that each time this operation is repeated, reinserting the previously tested object into the training set, leaving-(another)-one-out. During each phase, an average error between the training and testing set results is computed, and finally the errors of all the steps are averaged and can be used to evaluate the rules extracted by the RST software.

2.2. Rough Set theory applications in agricultural science

Despite its potential, RST has rarely been used in Agricultural Science. Table 4 summarizes the characteristics of the contributions identified in the Scopus® database from a search for "rough set theory agriculture" in "article title, abstract and keywords" search fields. Seventeen items were found, only five of which were fully-fledged articles, the others being simply conference papers. This review does not satisfy the standards for a meta-analysis, but considering the relevance of the database examined, the rarity of this method within the agricultural science community is very obvious. RST has been applied since 2004 mostly by Chinese researchers, with just three contributions from India and one from Canada. No contributions were found from European or American scientists. Most applications involved rule extraction for validation of expert assessments, such as classification and detection of plant diseases (Li *et al.*, 2004; Jianping, 2009; Xue and Tie-Min, 2010; Phadikar *et al.*, 2013), quality evaluation of raw materials in agro-food (Wang *et al.*, 2011) and agri-business supply chains (Wenxiu *et al.*, 2009), forecasting farmer behavior (Jain, 2007), risk analysis in swine breeding farms (Xi *et al.*, 2010), and generating guidelines for coconut cultivation (Sabu and Raju, 2011). RST has also been applied in agricultural science as a tool for data reduction. It was applied specifically as an element in multi-stage models for multivariate and complex problem solving, like best partner selection in supply chains (Guo and Lu, 2013), agricultural topic tracking (Zhang *et al.*, 2011), evaluation of soil fertility (Chen and Ma, 2011), and agricultural Big Data management (Shi *et al.*, 2012; Liu *et al.*, 2009). Finally, three relevant applications of RST regard detecting cause-effect links in forecast modeling for rural area depletion (Wang et al., 2012) and agricultural water demand (Li *et al.*, 2010), and as a knowledge acquisition step in expert system formulations for agricultural problem solving (Li *et al.*, 2013).

3. Materials and methods

The purpose of the present research is to propose a georeferenced Rough Set theory-based framework for agricultural sustainability analysis. Emphasis is placed on the dual interest of evaluating the Rough Set model as a tool for agricultural policy planning, and proposing a new framework for assessing agricultural sustainability. Particularly, Rough Set theory is used for agrarian regions classification in order to identify patterns within agricultural territories based on similarities within them. In this sense, the goal of the framework could be achieved applying traditional clustering techniques, however, it is worth being noted that RST permits researchers to introduce a decision variable, thus the RST models seem to be less subjective than clustering methods and more powerful in term of information induction.

Furthermore, note that the indiscernibility-based RST has been applied in the paper and agricultural sustainability indicators have been considered as attributes rather than criteria[2], discarding their gain- or cost- like features. This choice is based on the idea that describing the underlying features of phenomena may be more efficient than search for variables relationships description or policy planning optimization. Other approaches have been proposed and efficiently applied, for example, when indicators of performance and objectives' of policy planning do not diverge too much, researchers could apply multi-criteria spatial analysis or multi-objective optimization instead of information induction techniques. Multi-criteria spatial analysis helps to consider the desirability of indicators' value, while multi-objective optimization methods would indicates possible optimal solutions; these two approaches present very interesting features, but suffer when dealing with problems of a complex nature, such as agricultural sustainability analysis.

The framework scheme is shown in Figure 2, while the agricultural indicators and case study are presented in Table 5. The framework is organized into three phases:

1. *database construction*: starting from a database of geo-referenced variables measuring different characteristics of an agricultural system, indexes are calculated and aggregated at agrarian regional level. This step concludes with the proposal of three different datasets on the basis of the degree of depletion of the agricultural area in question (RST decision variables), all of which are used with continuous or discretized variables (so finally 6 different datasets have been constructed);
2. *Rough set model*: RST analysis is performed using the Rose2 computer program (Poznan University of Technology - http://idss.cs.put.poznan.pl/site/rose.html). In this step each of the three models is further prepared in order to test continuous or

[2] Readers must be aware that a dominance-based RST (Greco et al., 2001) has been developed in order to deal with multi-criteria decision analysis. For further details, refer also to Greco et al. (2002).

four-class categorized indicators. The phase finishes with the assessment of the best model from among the six tested;

3. *data representation*: the results derived from the best model are geo-referenced in order to facilitate interpretation.

3.1. Agricultural sustainability indicators and decision variables

Agricultural sustainability analysis requires the selection of a set of attributes/indicators, based on the spatial scale and dimensions of sustainability considered. As agricultural sustainability derives from activities at multiple scales, ranging from field and farm to regional, national, and even international scale (Smith and McDonald, 1998), the selection of an adequate spatial scale is crucial. Numerous researchers have opted for a farm/local scale in their studies[3], because of the possibility this scale offers for in-depth investigation of farm environment and economic dynamics. However, this approach requires specific surveying to collect primary data, generating high costs, relatively small samples, and difficulties of repetition over the years. The present paper thus adopts a territorial-scale, based on Italian public data, which limits costs while ensuring transparency of data and repeatability of measurements.

Choosing how to represent agricultural sustainability is also of fundamental importance. According to literature, agricultural sustainability encompasses the economic, social, and environmental dimensions (Smith and McDonald, 1998; Van Cauwenbergh et al., 2007; Gomez-Limon and Sanchez-Fernandez, 2010). In order to establish consistent indicators for each dimension, the present research applies the framework proposed by van Cauwenbergh (2007) as revised by Gomez-Limon and Sanchez-Fernandez (2010). The three pillars of agricultural sustainability are considered, identifying the most important factors for each dimension (called s*ub-dimensions*), establishing the associated criteria, and assessing a set of

[3] See van Wenum et al., 1999; van der Werf and Petit, 2002; Pacini et al., 2003; van Passel et al., 2007; Meul at al., 2008; Bechini and Castoldi, 2009; Gomez-Limon and Riesgo, 2009; Thomassen et al. 2009; Fumagalli et al. 2011; Reig-Martinez et al. 2011; Ripoll-Bosch et al., 2012; Paracchini et al., 2014.

indicators, i.e. proxy variables for each sub-dimension. Finally, the Rough Set model requires selection of a “decision” variable capable of classifying objects (i.e. agricultural areas) according to their state of agricultural sustainability. In the present case, agricultural sustainability was interpreted as the ability of agriculture to endure over time, and three decision variables were created. Table 5 summarizes the attributes and decision variables used in the analysis.

3.1.1 Economic sustainability indicators

The economic dimension of agricultural sustainability includes three sub-dimensions, defined by four indicators.

The agriculture profitability sub-dimension is defined by the *INCOME* indicator, which is the Agricultural Standard Output (SO) per hectare of Utilized Agricultural Area (UAA) in the Agrarian Region.

$$INCOME = {}^{SO}/_{UAA}$$

According to the European Commission (EC) Regulation 1242/2008 the SO of an agricultural product is its average monetary value at farm-gate price, per hectare or per head of livestock.

CONTAGR is the proxy for economic weight of the primary sector relative to the whole economy. Because the official macroeconomic statistics do not provide a municipality level, the SO per inhabitant is adopted.

$$CONTAGR = {}^{SO}/_{population}$$

Agriculture competitiveness is assessed from two viewpoints: the degree of dependency on public subsidies, and the share of farms of minimum economic dimensions for competition in the market. In many developed countries, agricultural income depends partly on public subsidies. In the EU the quota of subsidies allocated to agriculture is significant and public support for farmers is always central to debate ,regarding budget revisions. It is clear that in periods of strained public finances heavy dependence on subsidies can be seen as a weakness.

Following Paracchini et al. (2014) the relative indicator, or *CAP,* is measured as the ratio of CAP Subsidies to Agricultural Standard Output.

$$CAP = {}^{Subsidies}/_{SO}$$

Finally, the share of professional farms among total farms, the *PROF* indicator, is considered as an element to quantify the economic sustainability of the agricultural sector. In the Census data a discrimination was made between professional and recreational or part-time farms by referencing their Standard Output. The division was set conservatively at 10,000 euros of Standard Output.

$$PROF = {}^{Professional\_farms}/_{Total\ farms}$$

3.1.2 Social sustainability indicators

Two sub-dimensions, each defined by two indicators, define the social dimension of agricultural sustainability. The dimensions highlight the role of agriculture in maintaining occupation in rural areas and the capacity of farmers to establish relations and networks with society, responding to new social demand for services related to quality of life, leisure, and the environment (Meul et al., 2008). Agriculture can satisfy new social demands by diversification of farm activities and development of new services and functions (Van Huylenbroeck et al., 2007; Jongeneel et al., 2008). Some of these services might assume the form of marketable private goods, but many of them, especially environmental services, exhibit an externality/public good aspect (OECD, 2001; Van Huylenbroeck and Durand, 2003).

The first indicator related to the contribution of agriculture to employment is *AGRILABO*, which represents the weight of agricultural labor within the social context as annual work units over inhabitants.

$$AGRILABO = {}^{AWU}/_{inhabitants}$$

A measure of work stability is offered by *LABOSTAB*, which is the average annual worked days per person working in the farm (PWF).

$$LABOSTAB = {}^{Worked\ days}/_{PWF}$$

Two indicators were calculated for the agriculture's multifunctionality sub-dimension: *MULTIF*, which represents the diversification of production and services offered by farms, and *RISKABAN*, which is linked to the risk of abandonment and depletion of agricultural human capital in a long-term perspective. Farm diversification strategies were exhaustively summarized by van der Ploeg and Roep (2003), and Meert at al. (2005). The former distinguished between *deepening* and *broadening* diversification strategies. In the present study, like Paracchini et al. (2014), both these strategies are contemplated by means of the *MULTIF* indicator, which is the share of farm work time devoted to diversification activities.

$$MULTIF = {}^{Farm\ worktime\ MF}/_{Total\ farm\ worktime}$$

RISKABAN, as proposed by Reig-Martinez et al. (2011), approximates the risk of abandonment of agricultural activities. It assesses the risk as increasing with a farmer's age and decreasing when the farm income grows. The indicator, ranging from 0 (zero risk) to 1 (maximum risk), is defined as follows:

$$RISKABAN = trans\left(\frac{trans(AGE)+(1-trans(INCOME))}{2}\right),$$

where:

$\mathrm{trans}(x) = \left(\frac{x-\min x}{\max x-\min x}\right)$;

$AGE = average\ AGE\ of\ farmers\ in\ agrarian\ region$;

$INCOME = {}^{SO}/_{UAA}$.

3.1.3 Environmental sustainability indicators

The environmental dimension of agricultural sustainability is based on two sub-dimensions related to the characteristics of agricultural environmental outputs, separated into positive and negative externalities. Examples of positive externalities linked to agricultural production are rural amenities, biodiversity, nutrient recycling, and carbon sequestration, while negative externalities are represented by all forms of pollution deriving from agricultural activity (OECD, 2001; Van Huylenbroeck et al., 2007). Finally, five indicators were calculated, mainly deriving from the Agricultural Census Data.

The first indicator of the positive externalities sub-dimension is *SPECIAL*, which is a proxy for agricultural biodiversity. Aiming to comply with current agro-environmental policies, this indicator recalls the 2014-2020 CAP of the EU, which provides a Greening Payment for crop diversification (EU Council Regulation 1307/2013, article 44; Cavicchioli and Bertoni, 2014). *SPECIAL* was calculated as the share of the most representative arable crop over the total arable land. High values indicate a trend toward monoculture, lower values signal the spread of positive environmental practices like poly-culture and crop rotation.

$$SPECIAL = {}^{Main\ arable\ crop\ surface}\!\big/_{Total\ arable\ land}$$

To define the quality of rural landscape, the *LANDSCAPE* indicator was proposed, this being a measurement of landscape diversity. *LANDSCAPE* is set to the Gini index of heterogeneity of agricultural land use. A total of 73 categories of agricultural land use were taken into account.

$$LANDSCAPE = 1 - \sum_{i=1}^{n} k_i^2$$

where k_i is the relative frequencies of the n agricultural land use categories.

AGROENV is the share of agro-environmental surfaces (AES) in the total Utilized Agricultural Area (UAA). *AGROENV* also refers to the CAP commitments. In fact, another compulsory practice for the Greening Payment is the establishment of Ecological Focus Areas (EFA) on arable land (EU Council Regulation 1307/2013, article 46). The list of AES was obtained by overlapping the list of EFA with the Census land use categories. Organic farming areas were also included, these being automatically entitled to the Greening Payment. As a consequence, AES take into account organic farming areas, nitrogen-fixing crops, crops under water, multi-annual temporary grass, permanent grassland and pasture, and fallow land.

$$AGROENV = {}^{AES}/_{UAA}$$

Negative externalities were described by two indicators: *WATERUSE*, related to agricultural water consumption, and *NITROGEN*, which is a proxy for agricultural environmental pressure, generated by animal raising activities. *WATERUSE* is the quantity of annual irrigation water (cubic meters) per hectare of irrigated area, while *NITROGEN* represents the number of Livestock Units (LSU) per hectare of Utilized Agricultural Area (UAA)

$$WATERUSE = {}^{Annual\ irrigation\ water}/_{Irrigated\ area}$$

$$NITROGEN = {}^{LSU}/_{UAA}$$

3.1.4 Decision variables

The Rough Set model requires a decision variable for rule induction that classifies objects along with their attributes. In the present research, the classification indicates as more sustainable those agricultural territories in which the utilized agricultural area (UAA) has decreased the least over the last 20 years. Three different decision variables are tested, the thresholds of which were chosen arbitrarily by researchers on the basis of their expertise, but they could be reasonably applied to agricultural regions of other Developed Countries. The operational choice was driven by the rational consideration that the framework would be used in the same way as by policy-makers, who have to indicate boundaries of "level of risk" of depletion of agricultural areas. Finally, the three variables were calculated as follows:

$$9010\_4 = \begin{cases} 1\ if\ \Delta \geq 100 \\ 2\ if\ 80 \leq \Delta < 100 \\ 3\ if\ 60 \leq \Delta < 80 \\ 4\ if\ \Delta < 60 \end{cases}$$

$$9010\_3 = \begin{cases} 1\ if\ 80 \leq \Delta \leq 100 \\ 2\ if\ 60 \leq \Delta < 80 \\ 3\ if\ \Delta < 60 \end{cases}$$

$$9010\_2 = \begin{cases} 1\ if\ 70 \leq \Delta \leq 100 \\ 2\ if\ \Delta < 70 \end{cases}$$

where $\Delta = \frac{UAA_{2010}}{UAA_{2000}} \times 100$.

Note that in variable *9010_4*, positive changes (i.e. increase) in UAA are included in the information system, while in *9010_3* and *9010_2* these values were excluded because considered likely to be inconsistent with Italian agricultural trends, representing errors in agricultural census data. Despite it contains census errors, 9010_4 dataset has been retained in the elaboration because, in authors' opinion, using this dataset may improve the present RST exercise by showing the importance of the cross-validation step (see section 4.1) and suggesting researchers how to deal with problems in primary database.

3.2. Data and definition of empirical application

The Lombardy Region in Northern Italy was chosen as a case study. Lombardy is the most inhabited Italian region with about 10 million residents (16.4% of the Italian population) and a population density of 410 inhab./sq.km. Its surface area totals 23,862 square kilometers (7.9% of the national area), of which 47% is occupied by lowlands, 40.5% by mountains, and 12.5% by hills. The UAA comprise 43% of the land surface, while 14.5% is urbanized. The UAA declined by 15% between 1982 and 2010 due both to urbanization (especially in the lowlands) and agricultural abandonment and re-naturalization processes in the mountain areas.

The primary sector accounts for only 1.0% of regional added value and 1.5% of total workers, but contributes about 11.0% to national agricultural added value and 1.5% of the EU-28 total. Agricultural activity is prevalently oriented toward livestock, which accounts for 62.5% of the agricultural production value. However, in recent years an increasing number of farms have decided to convert their business to multifunctional activities such as direct processing and sale, farm tourism, and educational farms.

The high intensity of Lombardy agriculture is demonstrated by the average productivity of agricultural work units, with the agricultural added value per work unit at 46% and 166% above the national and EU-28 parameters, respectively.

The observation unit for the analysis was the Italian Agrarian Region, which consists in the aggregation of municipalities of similar characteristics in terms of agricultural systems and specializations. The average dimensions are intermediate between LAU 2 (Local Administrative Units 2) and NUTS 3 level. The sample included all the 87 Agrarian Regions of Lombardy in the dataset with the 9010_4 decision variable, while in the 9010_3 and 9010_2 decision variable datasets, in which positive variations in UAA were excluded, observations fell to 79.

Following the criteria defined above, almost all the indicators were calculated utilizing Agricultural Census data, gathered periodically (every 10 years) and covering the whole of Italy in detail. Census data are based on a clearly codified and widely acknowledged methodology, and their use offers numerous savings for public finances. This decision incurred a loss of information due to the use of secondary rather than primary data, but this was offset by the fact that Census Data are released on a very detailed level: the LAU 2 level of the EU-28, former NUTS 5 level, which generally corresponds to municipality level.

4. Results and discussion

4.1. Rough set model evaluations

Table 6 shows the results from the RS model evaluations. All the classes of all the datasets considered approximate perfectly consistent decision classes (accuracy of approximation of all classes of all datasets equals 1.00), thus we refer to leave-one-out cross validation to evaluate the rule extraction. Among the six databases, the LEM2 datasets (composed by discretized indicators) perform better than the modLEM datasets (composed by continuous indicators). Considering the decision variables, the accuracy of correct classification as measured by leave-one-out cross validation increases with the decreasing of the number of classes. Thus, the 9010_4 LEM2 and 9010_4 modLEM datasets present the worst performance for discretized and continuous indicators respectively;this was expected, considering that these databases contain census errors. On the other hand, the best model is represented by 9010_2 LEM2, which excludes positive variation in utilized agricultural area, and contains a 2-class decision variable and discretized indicators (see Table 7 for an extended report of the intervals). Total results present an average accuracy of classification of 89.87%, which is somewhat better than the other models tested in the analysis. For simplicity, only the rules induced by this example and the approximated created clusters are presented and discussed.

4.2. Rules with numerical and graphical results

The extracted rules are reported in Table 8. The models induced five rules for the objects in the first class (9010_2=1, for $70 \leq \Delta \leq 100$), and four rules for those in the second class (9010_2=2, for $\Delta < 70$). *Strength* and *Accuracy* measure the importance of each rule and the imprecision of a rough set caused by its boundary region respectively. Strength is the ratio of the objects in the described set over the rule on the objects in the class, while accuracy is the ratio of the lower approximation objects over upper approximation objects of the class in question. The findings in Table 8 show that some of the rules strongly dominate others in term of relative strength, and that the rough set model induced perfectly consistent classes. Each rule identifies a subset of Agrarian Regions of Lombardy that can be considered clusters of objects classified by certain attributes. In order to characterize all the groups, the average values of the 13 sustainability indicators were calculated for each of them, paying particular attention if they significantly diverged from the mean of the sample, as shown in Table 9. All the clusters are georeferenced in Figure 3.

Class one rules

Rule 1 (PROF=3) identifies a very large cluster of 41 Agrarian Regions associated by a high incidence of professional farms (average PROF of 62% compared to a regional average of 46%). In this area, the loss of UAA has been smaller compared to the other groups (approximately -10% in 20 years). The good economic performance of this group is confirmed by the average value of the INCOME indicator (38% higher than the regional average) and a high contribution of agriculture to the overall economy (CONTAGR=2,410.85 euro/inhabitant). The social sustainability dimension gives an ambivalent result: the economic dimension of farms guarantees employment stability (LABOSTAB=163.84), but conversely the low MULTIF value highlights a limited degree of social interconnection. As regards the environmental dimension, there is a trade-off between economic performance and environmental quality. This is particularly highlighted by the low value of the AGROENV indicator (proxy for the proportion of natural areas), and simultaneously by high livestock density and high water consumption per hectare (respectively higher by 48.5% and 53.7% compared to the regional average). On the basis of these results, the cluster defined by Rule 1 is identified as "professional intensive agriculture".

Rule 2 (LANDSCAPE=2 and WATERUSE=2) identifies a cluster of 7 Agrarian Regions localized around the most urbanized areas, in which the twenty-year loss of UAA is about 14%. In these areas, where the economic performance of the farms is satisfying, multifunctionality has become an alternative development strategy to intensification (MULTIF=0.18, +63.6% compared to the regional average). Although here agriculture makes a very modest contribution to the economy and employment, the agricultural labor exhibits a good level of stability and the primary sector seems to depend less on public contributions. Furthermore, multifunctional agriculture systems perform better than intensive agriculture systems for certain environmental indicators like LANDSCAPE. Based on these results, the cluster defined by Rule 2 is clearly identified as "multifunctional agriculture".

Rule 3 (MULTIF=1 and WATERUSE=0) classifies a small cluster of 5 observations, clearly identifiable from a territorial point of view. They are located in the far north of the region (the mountain areas of Valtellina). Here the decline in UAA is 13.6%, relatively low compared to the other mountain areas of the Region. Despite per hectare profitability being quite low (INCOME=2,090), the contribution of agriculture to the economy and, especially, to the labor force is higher compared to the other mountain areas (AGRILABO is even higher than the regional average). In contrast, the stability of the agricultural work has the lowest value of all the clusters (LABOSTAB=108.2; -30% compared to the regional average). Considering taking into account the low percentage of professional farms, these occurrences represent a clear indication of a part-time farming system. Despite its complementarity in terms of household income, in this context agriculture is performing quite well, sufficiently to maintain the UAA and the agro-environmental systems linked to agricultural production. The farming here exhibits an high degree of environmental sustainability, revealed in the high values for variability of the agricultural landscape (LANDSCAPE=0.82), the presence of ecological areas (AGROENV=0.94) and conversely, low values for negative environmental impact. Based on these results, the cluster defined by Rule 3 is identified as "pluriactivity".

Rule 4 (AGROENV=1) identifies a group of 24 observations, configured mainly (20 of 24 Agrarian Regions) as a subsystem of Rule 1. Similarly to Rule 1, there is high performance (slightly attenuated) of the economic indicators. The social indicator values are also comparable to those of Rule 1. However, the environmental indicators are very different, with SPECIAL and LANDSCAPE in particular shifted towards greater environmental sustainability. SPECIAL is 13.5% lower than the average, while LANDSCAPE is significantly higher. Both reflect a high degree of crop diversification and a greater variability and complexity of the agricultural landscape. Consequently, even in a context of high-intensity and high-profitability agriculture, Rule 4 seems to indicate an evolution of such systems towards improved environmental compatibility and compliance with the new environmental standards required by the CAP 2014-2020. This cluster of Agrarian Regions identified by Rule 4 configures itself as "agricultural biodiversity".

Rule 5 (PROF=1) is configured as a residual rule, combining 4 Agrarian Regions around the territory that are characterized by a loss of UAA close to 20%.

Class two rules

Rule 6 (PROF=0 and MULTIF=2) identifies a cluster of 18 Agrarian Regions (69.2% of the observations of group 2) in the low mountain areas, directly adjacent to the highly urbanized and most industrialized areas of the region. The cluster identified by this rule shows the highest loss of UAA (nearly 50%). This area is characterized by very low profitability of agriculture (INCOME is equal to 59% of the regional average). The contribution of agriculture to the whole economy is particularly limited, and only 24% of farms have a professional economic dimension, signaling a marginalized sector. As often happens in highly populated areas, residual farms tend to move toward multifunctional activities, which benefit from a vast pool of potential users (MULTIF=0.19). Environmental indicators confirm the typical situation of extensive mountain farming, characterized by numerous positive externalities and a modest environmental pressure. Therefore, this group includes areas with a marked marginalization of agriculture, largely due to competition with other economic sectors. This competition acts not only on the workforce, but also on the land, with almost all the flat areas converted for industrial, commercial, and residential use, while agriculture persisted only in the residual mountain areas, characterized by a low productivity and high production costs.

Similar observations can be made for Rule 7 (PROF=0 and WATERUSE=1) which includes a cluster of 14 observations that overlap widely with those of the previous rule. For this reason, the clusters identified in Rule 6 and Rule 7 can be classified as areas of "inter-sectorial competition".

Rule 8 (PROF=2 and NITROGEN=1) identifies a small cluster of 3 Agrarian Regions, located in the high mountain region. INCOME_L is very low here, even if the average size of the farms and the contribution of the agriculture to the economy and labor force are higher than for Rule 6 and Rule 7. Interestingly, the stability of the agricultural jobs is above the regional average. The agro-environmental indicators have similar values as for Rule 6 and Rule 7, although with better landscape quality and incidence of ecological areas (close to 100% of the UAA). In this cluster, inter-sectorial competition is less pronounced, but territorial conditions determine a marked regression of the agriculture, which survives only in the most favorable areas for animal raising activities. Based on this analysis, the cluster defined by Rule 8 is defined as "Marginal mountain agriculture".

As in Rule 5 in the group with the decision variable 9010_2=1, Rule 9 (INCOME=2) represents a residual rule for the group with decision variable 9010_2=2.

4.3. Rough Set model supporting policy-making

As stated in the preliminary considerations, Rough Set Theory can assist Public Administrations for the definition of policies, providing reliable synthetic information. On the basis of the case study, some recommendations can be formulated. Firstly, the decision variable clearly divided the territory of the Lombardy Region into 2 main areas, whose respective objects are geographically more similar to each other than to those in the other group. Agrarian Regions with a contained loss of UAA (class one) are located in plain areas or in the main part of the high mountain areas, with the remaining objects located in hilly and remaining mountain areas (Figure 3).

Secondly, induced rules allowed identification of nine sets of Agrarian Regions. For each group, it was possible to characterize the agricultural sector according to the three pillars of sustainability. Therefore, rules help to describe different subsystems, whose strengths and weaknesses can be isolated in order to define specific intervention initiatives. For example, agricultural areas with higher profitability and a high proportion of professional farms have a greater chance of being preserved. However, in some cases, like areas of multifunctional or part-time agriculture, agricultural contexts with a lower rate of professional farms seem to persist quite strongly over time.

The results also show that good economic performance is generally associated with lower social and environmental indicators, but once again with some notable exceptions.

The professional agriculture model represented by the Rule 1 cluster is undoubtedly the most solid from the perspective of the decision variable. However, it also has the most environmental impact. In this context, policies should act to reduce the environmental impact, but without weakening the sector in terms of economic competitiveness and employment. There are two potential paths to achieve this objective. First, the use of traditional policy instruments to incentivize more extensive production, landscape quality enhancement, and crop diversification. For example, as described above, Rule 4 identifies Agrarian Regions that have partly evolved towards crop biodiversity and ecological areas, although still characterized by good economic performance. The second option involves support for investment in green technologies, aspiring towards sustainable intensive farming. This is especially applicable in the most intensive areas, the Agrarian Regions under Rule 1 and not Rule 4, where adaptation by marked extensification would incur very high opportunity cost.

Some agricultural systems other than intensive also exhibit resilience over time. The Rule 2 cluster identifies multifunctional farms that provide a wide range of services for the metropolitan area. This is a growing phenomenon in Lombardy, adequately supported by policies that encourage short food supply chains and the provision of services for the community (Gaviglio *et al.* 2014a, 2014b). This is not the case for the part-time farming systems in the mountains, for which the implementation of the policies is rather difficult. Here the status of part-time farmer and the small physical and economic size of this class of farms is a criterion for excluding, or at best hindering, access to policies. This phenomenon is not necessarily a problem when scarce public resources are preferentially assigned to farms with the best economic performance and potentially more able to develop higher added-value projects. However, problems may arise if part-time agriculture is very widespread, covering much of a territory. These contexts are defined by Rule 3, when part-time farming represents a viable socio-economic model and the disappearance of agriculture would cause a loss of social and environmental externalities associated with the agro-forestry systems, in addition to economic repercussions.

In areas with an high variation in UAA (mainly identified by Rules 6, 7, and 8), problems are caused by the lack of economic and social sustainability of mountain farming. From the economic point of view, it is obvious that the disparity with intensive farming areas is due to territorial factors, while from a social point of view the problem is mainly the progressive social marginalization of the role of farmers. In these cases policies should, for example, encourage the establishment of young and neo-farmers (land values in mountain areas are low and do not represent an excessive entry barrier for the sector) and support projects that involve close integration of agriculture with tourism.

5. Conclusions

Integration of the concept of sustainability into agricultural policy presents ontological and technical difficulties, but nevertheless decision-makers use the term frequently to indicate the general and main objectives of their plans. Sustainability carries a very clear message and is one of the most widely recognized political terms among citizens. Consequently, researchers are obliged to integrate this concept into their analysis, creating some methodological issues. The present paper proposed a georeferenced framework based on Rough Set Theory, aimed at supporting policy makers and researchers when dealing with agricultural sustainability. Some interesting points arouse from data processing and discussion, demonstrating the advantages and limitations of the method.

Figure 4 summarizes the most important results of the analysis, clearly demonstrating the potential of Rough Set Theory for data classification. Starting from municipality data and more than 300 variables, these were aggregated into agrarian regions with 13 sustainability indicators; nine rules were then extracted using RST, classifying the Lombardy territory by shared sustainability assessment attributes. In this way, a problematic and possibly redundant information system was scaled down to an easy-to-read georeferenced set of clusters, which can be interpreted by policy-makers with little knowledge of statistical and mathematical formalities. The proposed framework appears to offer simple representation and communication of complex results, which is one of the most arduous tasks for researchers when trying to bridge the gap between scientific and everyday language.

The paper also proposes a reflection on the interpretation of sustainability. Using this concept in scientific research sometimes requires the incorporation of multiple and divergent objectives, and so an attempt was made to arrive at a simple congruous definition of sustainable agriculture, suggesting that the longer an agricultural activity endures, the more it can be considered sustainable. Under this definition, the levels of other different sustainability factors do not necessarily need to be considered as performance targets, instead using them as performance indices to support policy-makers during decisional processes. While the proposed framework does not definitively resolve the problems of evaluation of agricultural sustainability on a territorial level, it can be considered a useful methodological option that introduces some novelties into this research area.

Some limitations also need to be considered. While rules demonstrated a capacity to classify objects by relevant attributes, they excluded some attributes from evaluation. Depending on the type of analysis, the researcher has to decide whether to discard further attributes (as would be the case if cause-effect links detection) or retain them (for example in pattern recognition studies). However, these subjective interventions can be justified within a philosophy of inductive reasoning.

Furthermore, while Rough Set Theory provides easy-to-read results, its application in agricultural science appears limited and heterogeneous. This indicates that the method is suitable for different purposes, but suffers from competition with other statistical and mathematical tools for data processing. One reason for this could be its origin: RST was developed (and is currently well-known) by researchers working on the development of algorithms for artificial intelligence. These were probably more interested in the formal mechanisms of the model and its potential for information induction, unlike agricultural scientists who would be more interested in RST's capacity to provide answers in applied research. Furthermore, RST can only be run by a few free software packages, which are little known and might not be trusted by researchers as reliable (Fuggetta, 2003). Notwithstanding these limitations, the capacity of RST to deal with vagueness and inconsistent datasets, without any statistical constraints, offers very interesting prospects for more widespread use in numerous agricultural science applications.

Acknowledgements

Research funded by Regione Lombardia, fondo per la promozione di Accordi Istituzionali, project BIOGESTECA 15083/RCC. The authors gratefully acknowledge the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

Bechini, L., & Castoldi, N. (2009). On-farm monitoring of economic and environmental performances of cropping systems: results of a 2-year study at the field scale in northern Italy. Ecological Indicators, 9(6), 1096-1113.

Cavicchioli, D., Bertoni, D., (2014). Effects of CAP green payments in Lombardy: a comparison of proposed and approved measures based on census data. Paper prepared for 51st Sidea Congress “La PAC 2014-2020: scenari per i sistemi agroalimentari e rurali europei”, 18-20 september 2014, Benevento (Italy).

Chen, G., & Ma, L. (2011). Research on Rough Set and DecisionTree Method Application in Evaluation of Soil Fertility Level. In Computer and Computing Technologies in Agriculture IV (pp. 408-414). Springer Berlin Heidelberg.

Farmar-Bowers, Q., Higgins, V., & Millar, J. (2012). Food security in Australia: Challenges and prospects for the future. Springer.

Fayyad, U. M., & Irani, K. B. (1992). On the handling of continuous-valued attributes in decision tree generation. Machine learning, 8(1), 87-102.

Francis, C. A., Hildebrand, P. E., Adriance, J., Actis, J. J., Lanyon, L. E., Traxler, G., & IFARD/LAC, Bogotá (Colombia). (1989). Farming systems research-extension and the concepts of sustainability. Farming System Research Extension Newsletter (EUA), 6-11.

Fuggetta, A. (2003). Open source software an evaluation. Journal of Systems and Software, 66(1), 77-90.

Fumagalli, M., Acutis, M., Mazzetto, F., Vidotto, F., Sali, G., & Bechini, L. (2011). An analysis of agricultural sustainability of cropping systems in arable and dairy farms in an intensively cultivated plain. European Journal of Agronomy, 34(2), 71-82.

Gaviglio A., Bertocchi M., Silvestri S., Pirani A., “The use of environmental, economic and social indicators as decision support for the advancement of farms towards sustainability”, Proceedings of the 50th SIDEA Conference – Sustainability of the Agri-Food System: Strategies and Performances (edited by: De Gennaro B. C. and Nardone G.), Lecce, 26-28 settembre 2013, Universitas Studiorum, Mantova, 2014, pp. 215-226.

Gaviglio A., Pirani A., Bertocchi M., "Development of the Environmental, Social and Economic Sustainability in the Peri-Urban Agricultural Areas: Governance Opportunities in the South Milan Agricultural Park", Advanced Engineering Forum, 11, 2014, pp. 417-423.

Gaviglio, A., Demartini, E., & Pirani, A. (2012). The economic sustainability of agri-environmental policies: the application of Reg. (EC) N. 1095/2007 in the rice sector. PAGRI/IAP, Politica Agricola Internazionale, 21:2012, 65-75.

Gómez-Limón, J. A., & Sanchez-Fernandez, G. (2010). Empirical evaluation of agricultural sustainability using composite indicators. Ecological Economics, 69(5), 1062-1075.

Gómez-Limón, José A., and Laura Riesgo. "Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain." Journal of environmental management 90.11 (2009): 3345-3362.

Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. European journal of operational research, 129(1), 1-47.

Greco, S., Matarazzo, B., & Slowinski, R. (2002). Multicriteria classification by dominance-based rough set approach. Handbook of data mining and knowledge discovery. Oxford University Press, New York.

Grzymala-Busse, J. W. (1992). LERS-a system for learning from examples based on rough sets. In Intelligent decision support (pp. 3-18). Springer Netherlands.

Guo F., Lu Q. (2013). Partner Selection Optimization Model of Agricultural Enterprises in Supply Chain. Advance Journal of Food Science &Technology, 5(10).

Guo, F., & Lu, Q. (2013). Partner Selection Optimization Model of Agricultural Enterprises in Supply Chain. Advance Journal of Food Science &Technology, 5(10).

Hansen J. W. (1996). Is agricultural sustainability a useful concept? Agricultural Systems, 50(2), 117-143.

Jain, R. (2007). Approximate Corebased RDT model for Classifying Technology Savvy Farmers. In IICAI (pp. 1669-1683).

Jianping, Z. (2009). Study on agricultural knowledge discovery based on rough set theory. In Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on (Vol. 3, pp. 701-704). IEEE.

Jongeneel, R. A., Polman, N. B., & Slangen, L. H. (2008). Why are Dutch farmers going multifunctional?. Land use policy, 25(1), 81-94.

Kassam A., Friedrich T., Shaxson F., Pretty J. (2009). The spread of conservation agriculture: Justification, sustainability and uptake. International Journal of Agricultural Sustainability, 7, (4).

Kryszkiewicz, M. (1998). Rough set approach to incomplete information systems. Information sciences, 112(1), 39-49.

Li A.-P., Sui P.-B., Jia Y., Liao G.-P., Wu Q.-Y. (2004). Knowledge acquisition method based on rough set in diagnostic expert system. Ruan Jian Xue Bao/Journal of Software, 15 (SUPPL.), pp. 13-19.

Li Z., Huang H. X., Zheng Y. L., Wang Z. Y. (2013, March). An intelligent diagnosis model based on rough set theory. In Fifth International Conference on Machine Vision (ICMV 12) (pp. 87842L-87842L). International Society for Optics and Photonics.

Li, X., Ding, L., Lv, J. (2010). Agriculture irrigation water demand forecasting based on rough set theory and weighted LS-SVM. 2nd International Conference on Communication Systems, Networks and Applications, ICCSNA 2010, 2, art. no. 5588826, pp. 371-374.

Liu, M., Zhang, X., & Wang, B. (2009, December). Research on agricultural decision support system based on rough set theory. In Bio Medical Information Engineering, 2009. FBIE 2009. International Conference on Future (pp. 102-109). IEEE.

Meert, H., Van Huylenbroeck, G., Vernimmen, T., Bourgeois, M., & Van Hecke, E. (2005). Farm household survival strategies and diversification on marginal farms. Journal of rural studies, 21(1), 81-97.

Meul, M., Van Passel, S., Nevens, F., Dessein, J., Rogge, E., Mulier, A., & Van Hauwermeiren, A. (2008). MOTIFS: a monitoring tool for integrated farm sustainability. Agronomy for sustainable development, 28(2), 321-332.

Obersteiner, M., & Wilk, S. (1999). Determinants of Long-term Economic Development: An Empirical Cross-country Study Involving Rough Sets Theory and Rule Induction (No. 11). Institute for Advanced Studies.

OECD (2001). Multifunctionality: towards an analytical framework (p. 158). Paris: OECD.

Pacini, C., Wossink, A., Giesen, G., Vazzana, C., & Huirne, R. (2003). Evaluation of sustainability of organic, integrated and conventional farming systems: a farm and field-scale analysis. Agriculture, Ecosystems & Environment, 95(1), 273-288.

Paracchini, M. L., Bulgheroni, C., Borreani, G., Tabacco, E., Banterle, A., Bertoni, D., Rossi, G., Parolo, G., Origgi, R., & De Paola, C. (2015). A diagnostic system to assess sustainability at a farm level: The SOSTARE model. Agricultural Systems, 133, 35-53.

Pawlak, Z. (1982). Rough sets. International Journal of Computer and Information Sciences, 11(5), 341-356.

Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. European journal of operational research, 99(1), 48-57.

Pawlak, Z., &Skowron, A. (2007). Rudiments of rough sets. Information sciences, 177(1), 3-27.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., &Ziarko, W. (1995). Rough sets. Communications of the ACM, 38(11), 88-95.

Phadikar, S., Sil, J., &Das, A. K. (2013). Rice diseases classification using feature selection and rule generation techniques. Computers and Electronics in Agriculture, 90, 76-85.

Pretty J. N. (1995). Participatory learning for sustainable agriculture. World Development, 23(8), 1247-1263.

Reig-Martínez, E., Gómez-Limón, J. A., & Picazo-Tadeo, A. J. (2011). Ranking farms with a composite indicator of sustainability. Agricultural Economics, 42(5), 561-575.

Ripoll-Bosch, R., Díez-Unquera, B., Ruiz, R., Villalba, D., Molina, E., Joy, M., Olaizola, A., & Bernués, A. (2012). An integrated sustainability assessment of mediterranean sheep farms with different degrees of intensification. Agricultural systems, 105(1), 46-56.

Sabu, M. K., &Raju, G. (2011). Rule induction using Rough Set Theory—An application in agriculture. In Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on (pp. 45-49). IEEE.

Shi, L., Ma, X., Duan, Q., Weng, M., & Qiao, H. (2012). Agricultural Data Classification Based on Rough Set and Decision Tree Ensemble. Sensor Letters, 10(1-2), 1-2.

Slowinski, R., Greco, S., & Matarazzo, B. (2012). Rough set and rule-based multicriteria decision aiding. Pesquisa Operacional, 32(2), 213-270.

Smit B, Smithers J (1993). Sustainable agriculture: Interpretations, analyses and prospects. Canadian Journal of Regional Science, 16, 499-524.

Smith C. S., McDonald G. T. (1998). Assessing the sustainability of agriculture at the planning stage. Journal of Environmental Management, 52, 15-37.

Stefanowski, J. (1998). On rough set based approaches to induction of decision rules. Rough sets in knowledge discovery, 1(1), 500-529.

Stefanowski, J. (1998). The rough set based rule induction technique for classification problems. In In Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT (Vol. 98).

Stefanowski, J. (2007). On combined classifiers, rule induction and rough sets. In Transactions on rough sets VI (pp. 329-350). Springer Berlin Heidelberg.

Thangavel, K., & Pethalakshmi, A. (2009). Dimensionality reduction based on rough set theory: A review. Applied Soft Computing, 9(1), 1-12.

Thomassen, M. A., Dolman, M. A., Van Calker, K. J., & De Boer, I. J. M. (2009). Relating life cycle assessment indicators to gross value added for Dutch dairy farms. Ecological Economics, 68(8), 2278-2284.

Van Cauwenbergh, N., Biala, K., Bielders, C., Brouckaert, V., Franchois, L., Garcia Cidad, V., Hermy, M., Mathijs, E., Muys, B., Reijnders, J., Sauvenier, X., Valckx, J., Vanclooster, M., Van der Veken, B., Wauters, E., & Peeters, A. (2007). SAFE—A hierarchical framework for assessing the sustainability of agricultural systems. Agriculture, ecosystems & environment, 120(2), 229-242.

Van der Ploeg, J. D., & Roep, D. (2003). Multifunctionality and rural development: the actual situation in Europe. Multifunctional agriculture: a new paradigm for European agriculture and rural development, 37-54.

Van der Werf, H. M., & Petit, J. (2002). Evaluation of the environmental impact of agriculture at the farm level: a comparison and analysis of 12 indicator-based methods. Agriculture, Ecosystems & Environment, 93(1), 131-145.

Van Huylenbroeck, G., & Durand, G. (2003). Multifunctionality and rural development: a general framework. Multifunctional agriculture: A new paradigm for European agriculture and rural development, 1-16.

Van Huylenbroeck, G., Vandermeulen, V., Mettepenningen, E., & Verspecht, A. (2007). Multifunctionality of agriculture: a review of definitions, evidence and instruments. Living Reviews in Landscape Research, 1(3), 1-43.

Van Passel, S., Nevens, F., Mathijs, E., & Van Huylenbroeck, G. (2007). Measuring farm sustainability and explaining differences in sustainable efficiency. Ecological economics, 62(1), 149-161.

van Wenum, J., Buys, J., & Wossink, A. (1999). Nature quality indicators in agriculture. Environmental indicators and agricultural policy (Brouwer F., Crabtree B., eds). CABI Publ, Wallingford, UK, 105-120.

Wang, F., Hasbani, J. G., Wang, X., & Marceau, D. J. (2011). Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set Theory. Computers, Environment and Urban Systems, 35(2), 116-125.

Wenxiu, Z., Shuangxi, L., Jinxing, W., Yuguo, K., &Futang, L. (2009). Research on dynamic identification technology of cottonforeignfibers. Chinese Journal of Scientific Instrument, 10, 034.

Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Xi, W., Sun, Y., Tian, X., Wang, H., & Jin, Q. (2010). Research on the SafetyRisk Structure and Early Warning System of Agriculture With Illustrations of Production of Live pigs to Farmers. Agriculture and Agricultural Science Procedia, 1, 462-468.

Xue, W., &Tie-min, M. (2010). The application of Edge Detection Algorithm Based on Rough Sets in the detection of soybean target leaf spot. In 2010 International Conference on Computer Application and System Modeling (ICCASM 2010) (Vol. 3).

Yao, Y. Y. (1998). Constructive and algebraic methods of the theory of rough sets. Information sciences, 109(1), 21-47.

Zhang, W., Wang, X., Li, S., &Xia, C. (2011, December). Application of data mining in agricultural topic tracking. In Computer Science and Network Technology (ICCSNT), 2011 International Conference on (Vol. 1, pp. 38-41). IEEE.

Tables and Figures

Table 1. A decision table

U \ Q	Attributes			Decision
	Size	*Age*	*Type*	*Biogas*
x1	Big	Old	Swine	Yes
x2	Big	Old	Cattle	Yes
x3	Normal	Old	Swine	Yes
x4	Normal	Young	Cattle	No
x5	Small	Old	Swine	No
x6	Small	Old	Cattle	No

Table 2. Decision table with no redundant attributes

U \ Q	Attributes		Decision
	Size	*Type*	*Biogas*
x1	Big	Swine	Yes
x2	Big	Cattle	Yes
x3	Normal	Swine	Yes
x4	Normal	Cattle	No
x5	Small	Swine	No
x6	Small	Cattle	No

Table 3. Decision table with inconsistencies

U \ Q	Attributes		Decision
	Size	*Type*	*Biogas*
x1	Big	Swine	Yes
x2	Big	Cattle	Yes
x3	Normal	Swine	Yes
x4	Normal	Cattle	No
x5	Small	Swine	No
x6	Small	Cattle	No
x7	Normal	Swine	No
x8	Big	Swine	Yes

Figure 1. Elementary sets, decision classes and inconsistency

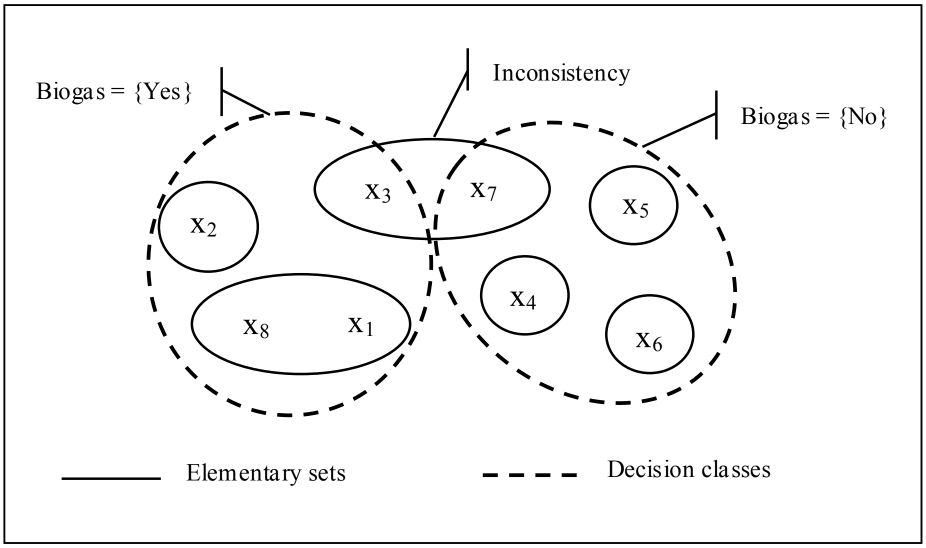


Table 4 – Use of Rough Set Theory in agricultural sciences

Reference	Document type	Use of RS	Description and Aims of analysis	Affiliation
Guo and Lu, 2013	Article	Data reduction	Selection of the best business partner	Hangzhou, China
Li *et al*., 2013	Conference paper	Knowledge acquisition	Expert system for complex problem solving	Guangzhou, China
Phadikar *et al*., 2013	Article	Rule extraction	Automatic classification of rice plants' disease	Salt Lake, India
Wan *et al*., 2012	Conference paper	Rule extraction	Automatic classification of soybean quality	Wuxi, China
Shi *et al*., 2012	Article	Data reduction	Reduction and classification of agricultural data	Zhengzhou, China
Zhang *et al*., 2011	Conference paper	Data reduction	Algorithm for agricultural topic tracking	Mudanjiang, China
Sabu and Raju, 2011	Conference paper	Rule extraction	Rules for optimization of coconut cultivation	Kerala, India
Wang *et al*., 2011	Article	Cause-effect detection	Evaluation of determinants of urbanization	Calgary, Canada
Chen and Ma, 2011	Conference paper	Data reduction	Evaluation of soil fertility levels	Jilin, China
Wenxiu *et al*., 2009	Conference paper	Rule extraction	Automatic classification of cotton fiber	Taian, China
Xue and Tie-Min, 2010	Conference paper	Rule extraction	Automatic classification of soy plants disease	Heilongjiang Province, China
Xi *et al*., 2010	Conference paper	Rule extraction	Risk detection systems in swine farms	Jilin, China
Li *et al*., 2010	Conference paper	Cause-effect detection	Forecast of agriculture water demand	Guangzhou, China
Liu *et al*., 2009	Conference paper	Data reduction	Reduction and classification of agricultural data	Wuxi, China
Jianping, 2009	Conference paper	Rule extraction	Diagnosis of plant diseases and insect pests	Hangzhou, China
Jain, 2007	Conference paper	Rule extraction	Forecast of farmers adoption of pesticides, insecticides or fungicides	New Delhi, India
Li *et al*., 2004	Article	Rule extraction	Automatic classification of rapeseed disease	Changsha, China

Figure 2. Framework proposed for agricultural sustainability assessment

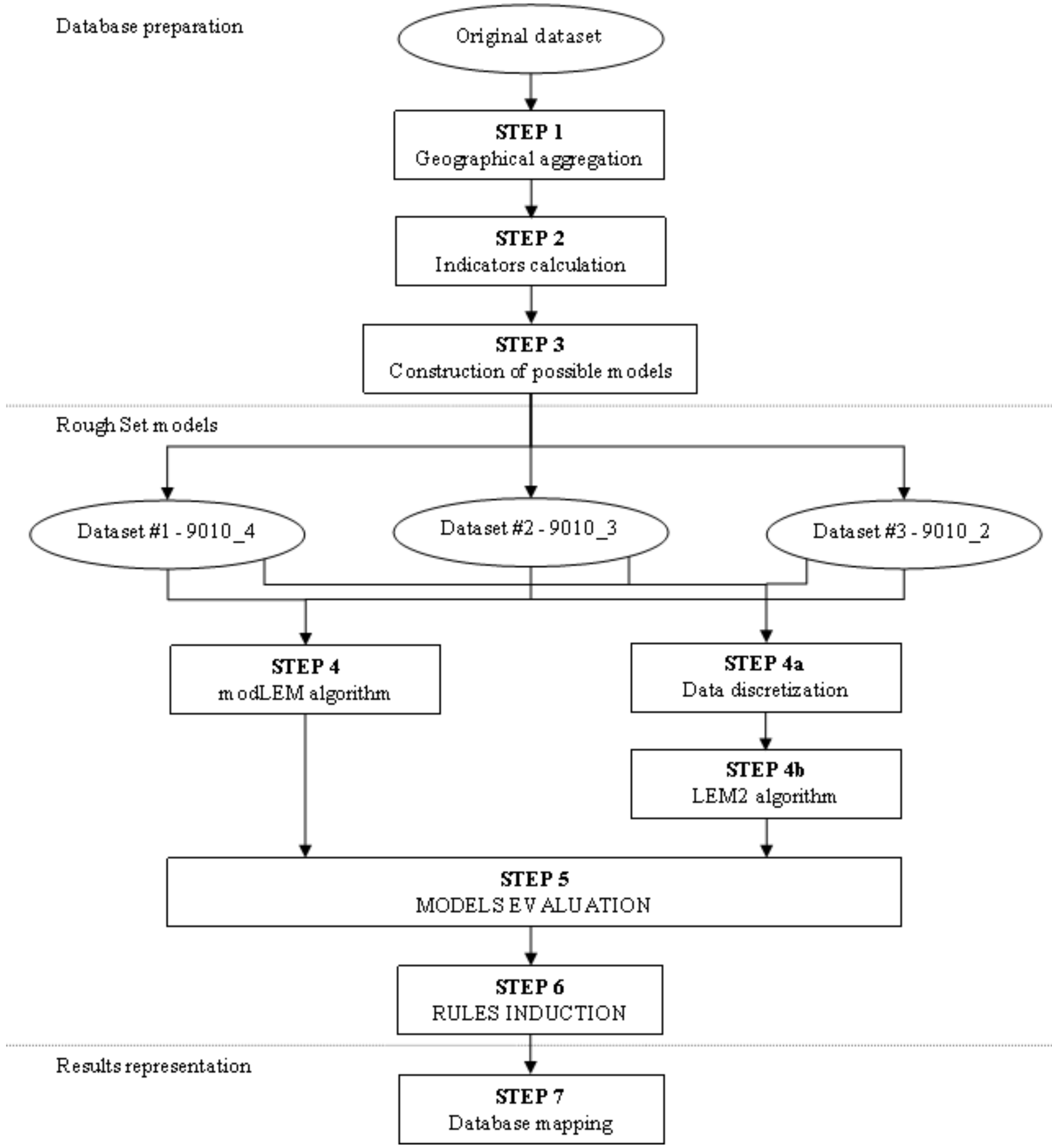


Table 5. Agricultural sustainability indicators and Rough Set decision variable used in the analysis

Dimension	Sub-dimension	Criteria	Indicator	Definition	Source
Economic	Agriculture profitability	Income maximisation	Agricultural income (INCOME)	Agricultural Standard Output per hectare of Utilized Agricultural Area (UAA)	ISTAT, Agricultural Census 2010
	Agriculture contribution to economy	Improving agriculture contribution to the whole economy	Agriculture economic weight (CONTAGR)	Agricultural Standard Output per inhabitant	ISTAT, Agricultural Census 2010 and Resident Population 2010
	Agriculture competitiveness	Improving farms competitiveness	Dependency on Cap subsidies (CAP)	The ratio of Cap Subsidies to Agricultural Standard Output	Regione Lombardia, Agriculture Information System; ISTAT, Agricultural Census 2010
			Ratio of professional farms (PROF)	The share of professional farms (farms having Standard Output > 10,000 euros) on total farms	ISTAT, Agricultural Census 2010
Social	Agriculture contribution to employment	Improving and stabilising occupation	Agriculture social weight (AGRILABO)	The number of agricultural Annual Work Units (AWU) per inhabitant (x 1,000)	ISTAT, Agricultural Census 2010
			Stability of workforce (LABOSTAB)	The average annual worked days per person working in the farm. Persons working in the farm include holder family and relatives, non-family labour regularly employed and other labour employed on a non-regular basis	ISTAT, Agricultural Census 2010
	Agriculture multifunctionalit y	Improving agriculture multifunctionality	Agriculture multifunctionality degree (MULTIF)	The share of time worked in farm multifunctional activities. Multifunctional activities include farm tourism, recreational and social activities, teaching farms, handicrafts, processing of agricultural products, direct selling, renewable energy production, etc.	ISTAT, Agricultural Census 2010
			Risk of abandonment of agriculture (RISKABAN)	Proxy of the probability of abandonment of agriculture based on the farmer's age and the farm's profitability	ISTAT, Agricultural Census 2010
Environmental	Agriculture positive externalities	Improving positive externalities of agricultural production	Crop diversification (SPECIALI)	The share of the main arable crop on the total arable land	ISTAT, Agricultural Census 2010
			Landscape heterogeneity (LANDSCAPE)	The Gini index of heterogeneity of the agricultural land use. 73 categories of agricultural land use have been taken into account.	ISTAT, Agricultural Census 2010
			Agro-environmental areas (AGROENV)	The share of the agro-environmental areas on the total Utilized Agricultural Area (UAA). Agro-environmental areas include organic farming area, nitrogen-fixing crops, crops under water, multi-annual temporary grass, permanent grassland and pasture and fallow land	ISTAT, Agricultural Census 2010
	Agriculture negative externalities	Reducing negative externalities of agricultural production	Irrigation water consumption (WATERUSE)	The volume of annual irrigation water (cubic meters) per hectare of irrigated area	ISTAT, Agricultural Census 2010
			Agriculture environmental pressure (NITROGEN)	The number of Livestock Units (LSU) per hectare of Utilized Agricultural Area (UAA)	ISTAT, Agricultural Census 2010
Decision variable	State of agricultural activities	Ability of agricultural activities to endure over time	Utilized Agricultural Area depletion #1 (9010_4)	Classification in 4 classes of UAA depletion over years 1990-2010	ISTAT, Agricultural Census 1990; ISTAT, Agricultural Census 2010
			Utilized Agricultural Area depletion #2 (9010_3)	Classification in 3 classes of UAA depletion over years 1990-2010, excluding database's inconsistencies	ISTAT, Agricultural Census 1990; ISTAT, Agricultural Census 2010
			Utilized Agricultural Area depletion #3 (9010_2)	Classification in 2 classes of UAA depletion over years 1990-2010, excluding database's inconsistencies	ISTAT, Agricultural Census 1990; ISTAT, Agricultural Census 2010

Table 6 – Leave-one-out cross-validation results for different model tested

Classes and Total	Average accuracy of correct prediction (%)					
	9010_4		9010_3		9010_2	
	modLEM	LEM2	modLEM	LEM2	modLEM	LEM2
Total	*40.23*	*51.72*	*50.63*	*64.56*	*77.22*	*89.87*
Class 1	0.00	12.50	66.67	83.33	79.25	94.34
Class 2	66.67	66.67	20.00	30.00	73.08	80.77
Class 3	5.00	30.00	47.06	58.82	--	--
Class 4	35.29	58.82	--	--	--	--

Note: modLEM algorithm is used on continuous indicators, while LEM2 refers to categorized indicators. Accuracy of approximation of all classes in all datasets equals 1.00.

Table 7 – Discretization intervals for indicators in model 9010_2 – LEM2

Indicator	Discretized Class	N. of obj.	Min	Max	Indicator	Discretized Class	N. of obj.	Min	Max
INCOME	0	10	642.50	2,179.86	RISKABAN	0	2	-	0.26
	1	41	2,214.87	7,156.77		1	16	0.28	0.54
	2	2	7,230.38	7,230.92		2	59	0.55	0.95
	3	26	7,420.93	16,456.69		3	2	0.96	1.00
CONTAGR	0	2	8.65	18.98	SPECIALI	0	5	0.26	0.34
	1	38	33.00	430.53		1	70	0.35	0.83
	2	5	451.55	601.56		2	3	0.84	0.94
	3	34	653.30	10,594.75		3	1	1.00	1.00
CAP	0	3	0.01	0.01	LANDSCAPE	0	6	0.50	0.62
	1	22	0.02	0.04		1	32	0.63	0.78
	2	53	0.05	0.23		2	16	0.79	0.82
	3	1	0.41	0.41		3	25	0.83	0.91
PROF	0	30	0.03	0.36	AGROENV	0	16	0.12	0.27
	1	4	0.37	0.38		1	24	0.30	0.54
	2	4	0.39	0.41		2	33	0.55	0.98
	3	41	0.43	0.83		3	6	0.99	1.00
AGRILABO	0	7	0.24	2.95	WATERUSE	0	9	606.40	1,005.08
	1	54	3.08	25.91		1	16	1,040.77	1,768.16
	2	8	26.31	32.91		2	23	1,776.94	3,336.58
	3	10	33.80	171.74		3	31	3,353.58	22,218.16
LABOSTAB	0	4	67.55	79.67	NITROGEN	0	6	0.08	0.24
	1	19	92.38	136.54		1	7	0.26	0.50
	2	3	137.24	140.72		2	42	0.52	2.15
	3	53	140.76	252.53		3	24	2.39	8.68
MULTIF	0	4	0.01	0.02					
	1	42	0.03	0.11					
	2	31	0.12	0.30					
	3	2	0.31	0.35					

Table 8 – Rules induced by model 9010_2 – LEM2

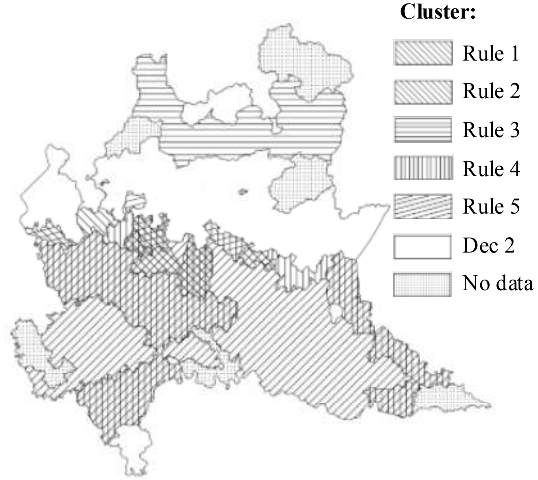
Class	ID	Rule	No. Objects	Strength
Class 1 Accuracy= 1.00	1	If PROF = 3 then 9010_2 = 1	41	77.36%
	2	If LANDSCAPE = 2 & WATERUSE = 2 then 9010_2 = 1	7	13.21%
	3	If MULTIF = 1 & WATERUSE = 0 then 9010_2 = 1	5	9.43%
	4	If AGROENV = 1 then 9010_2 = 1	24	45.28%
	5	If PROF = 1 then 9010_2 = 1	4	7.55%
Class 2 Accuracy= 1.00	6	If PROF = 0 & MULTIF = 2 then 9010_2 = 2	18	69.23%
	7	If PROF = 0 & WATERUSE = 1 then 9010_2 = 2	14	53.85%
	8	If PROF = 2 & NITROGEN = 1 then 9010_2 = 2	3	11.54%
	9	If INCOME = 2 then 9010_2 = 2	2	7.69%

Table 9 – Values of Agricultural Sustainability indicators per Rule and ANOVA test

Indicator	9010_2 =1					9010_2 = 2				Total
	R1	R2	R3	R4	R5	R6	R7	R8	R9	
	no. = 41	*no. = 7*	*no. = 5*	*no. = 24*	*no. = 4*	*no. = 18*	*no. =14*	*no. = 3*	*no. = 2*	*no. = 79*
INCOME	7,860.37 ***	6,339.17	2,089.99 **	7,175.91 *	5,683.82	3,363.82 **	3,521.31 **	1,644.49 *	7,230.65	5.692.32
CONTAGR	2,410.85 **	315.65	732.49	1,697.65	173.64	209.69 **	190.07 **	709.86	422.50	1.499.08
CAP	0.09	0.05	0.05	0.07	0.03	0.08	0.05	0.08	0.03	0.08
PROF	0.62 ***	0.40	0.26 **	0.55	0.38	0.24 ***	0.26 ***	0.40	0.28	0.46
AGRILABO	21.89	7.10	26.56	21.18	7.46	14.21	12.54	29.38	13.12	20.05
LABOSTAB	163.84	158.72	108.18 ***	160.94	157.25	151.60	157.15	173.74	125.77	155.22
MULTIF	0.09 *	0.18 **	0.06	0.11	0.15	0.19 ***	0.16 **	0.11	0.08	0.11
RISKABAN	0.65	0.72	0.73	0.69	0.59	0.60	0.66	0.54	0.86	0.65
SPECIALI	0.56	0.54	0.65	0.51 **	0.57	0.63	0.58	0.64	0.68	0.59
LANDSCAPE	0.75	0.81	0.82	0.81 **	0.82	0.75	0.77	0.80	0.77	0.76
AGROENV	0.37 ***	0.43	0.94 **	0.40 ***	0.67	0.87 ***	0.88 ***	0.96 **	0.60	0.60
WATERUSE	6.625.77 **	2.396.58	920.63	4.359.58	2.040.24	1.608.57 **	1.417.77 **	1.503.33	2.619.37	4.310.33
NITROGEN	2.88 **	1.98	0.53	2.32	1.30	1.01 **	1.09 *	0.35	1.48	1.94
9010_2	90.36	85.87	86.38	85.88	81.38	52.85	56.27	57.17	55.25	77.24

*Note: ANOVA test have been run for all the indicators per each Rule vs the Total (sample) mean - Sign. *=0.10**=0.05 ***=0.01.*

Figure 3 – Geo-r eferenced results



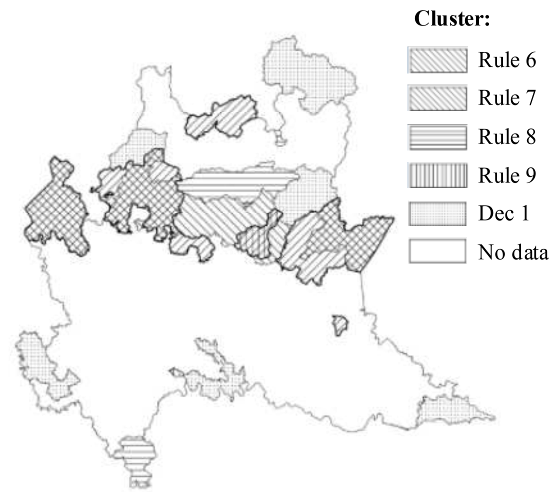


Figure 4 – Geo-referenced framework for agricultural sustainability analysis