Simulation Modelling Practice and Theory 102 (2020) 102068

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Agro 4.0: A data science-based information system for sustainable agroecosystem management

Eugênio Pacceli Reis da Fonseca[*,a], Evandro Caldeira[a], Heitor Soares Ramos Filho[a], Leonardo Barbosa e Oliveira[a], Adriano César Machado Pereira[a], Pierre Santos Vilela[b]

[a] *Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - Pampulha, Belo Horizonte, Minas Gerais (31270-901), Brazil*
[b] *FAEMG, Federação da Agricultura e Pecuária do Estado de Minas Gerais, Avenida do Contorno, 1.771 - Belo Horizonte, Minas Gerais (30110-005), Brazil*

ARTICLE INFO

Keywords:
Sustainability
Agroecosystems
Intelligent IoT
Data analysis

2019 MSC:
00-01
99-00

ABSTRACT

One of the solutions for handling and treating the diverse data related to the sustainability of an agroecosystem is the use of Information Systems and Internet of Things. In this work, we adopt a methodology called Indicators of Sustainability in Agroecosystems (*Indicadores de Sustentabilidade em Agroecossistemas* – ISA), implement an information system based on Internet of Things and apply Data Science and simulation techniques over the gathered data, from 100 real rural properties. As a result, we have developed a set of tools for data collection, processing, visualization, simulation and analysis of the sustainability of a rural property or region, following the ISA methodology. Two experiments were applied on the dataset collected by the tools: environmental change scenarios simulations on targeted agroecosystems to predict how they affect two ISA scores (Soil Fertility and Water Quality) of involved agroecosystems; Evaluation of Feature Selection models searching for subsets of features good enough to predict the two ISA scores for the dataset with a smaller amount of data necessary. We have that with only 7 of the 21 Indicators present in ISA we can identify the level of sustainability in more than 90% of cases, allowing for a new discussion about shrinking the amount of data needed for the computation of ISA, or remodeling the final computation of the Sustainability Index so other Indicators can be more expressive. Users of the solutions developed in this work can identify best practices for sustainability in participating agroecosystems.

1. Introduction

Sustainability is when one can work on the present development without compromising the development of future generations [1]. The idea of sustainable development became widely known after the United Nations Conference on Environment and Development in Rio de Janeiro in 1992 (ECO-92). Since then, there have been variations in the definition of sustainability [2–5], but they all converge to a definition where sustainability implies a medium- and long-run profitability, as well as agricultural practices with sustainable environmental impacts. Public awareness of the negative impacts of human activity on our environment is at an all times high, and, according to specialists' predictions, no more time can be wasted [6].

[*] Corresponding author.
E-mail addresses: eugenio.pacceli@dcc.ufmg.br (E.P.R. da Fonseca), evandro@dcc.ufmg.br (E. Caldeira), ramosh@dcc.ufmg.br (H.S. Ramos Filho), leob@dcc.ufmg.br (L. Barbosa e Oliveira), adrianoc@dcc.ufmg.br (A.C.M. Pereira), psvilela@faemg.org.br (P.S. Vilela).

https://doi.org/10.1016/j.simpat.2020.102068
Received 31 August 2019; Received in revised form 9 January 2020; Accepted 10 January 2020
Available online 13 January 2020

Achieving the growing challenges of the next decade [7] is a complex task due to a large number of variables. In addition to a solid model for measuring sustainability, relevant data needs to be well structured for retrieval, storage and analysis. There are also other problems related to understanding the factors that affect sustainability. Therefore, there is a dire need for a system able to: a) characterize, visualize, and analyze collected data; and b) implement smart strategies that can measure sustainability in agroecosystems.

As a solution to this problem, we have designed, developed, and evaluated a sustainability management system in agroecosystems based on data science and internet of things. The system, dubbed Agro 4.0[1] [8], provides tools for the collection, storage, analysis, visualization and scenarios simulations of sustainability-related information of rural properties. Agro 4.0 enables, for instance, different properties to be compared regarding their Sustainability Index. Besides, the system can pinpoint agro-ecosystems with critical levels of sustainability and then suggest managers measures to reverse the situation. It is worth saying that Agro 4.0 is based on the Agroecosystems Sustainability Index (ISA) [9] methodology.

Agro 4.0 is inserted in a multidisciplinary and multifaceted context. First of all, it is an *Green Information System* that supports *Decision Making* with data, information extraction and visualization. The system facilitates the work of groups and entities that seek to improve the sustainability of properties of a given profile. This profile can be, for instance, for the type of product produced. The system has different interactive interfaces (for the browser, tablets and desktop), and asks for and accepts data potentially generated by other sensors (cameras, soil and water sensors, drones, etc). The system makes use of *Data Science*, which enables managers of projects to perform more complex analysis over a set of rural properties. *Agroecosystems* and *Sustainability* are at the core of the ISA Methodology, that Agro 4.0 implements.

This work has three main contributions: (i) the identification of which indicators of the ISA Model are more relevant or expressive for the Sustainability of a rural property, opening a discussion about the amount of data that ISA requires and the possibility to reduce the input (as it is right now, there are hundreds of fields a technician has to fill in order to obtain the final scores, and some require analysis of samples in laboratories and portable test kits); (ii) A tool to simulate environmental changes scenarios and their impact on the sustainability score and other indicators of the ISA methodology on participating agroecosystems; (iii) a data science-based information system for sustainability management of agroecosystems that allows to:

1. collect, to structure and validate data about the sustainability of agroecosystems using the ISA Methodology;
2. manage information about sustainability and support decision in agroecosystems;
3. identify and characterize the most relevant factors for sustainability;
4. perform visualization and analysis over aggregate data in a user-friendly way;
5. simulate environmental changes and measure their impact on ISA indicators and sustainability indexes.

Besides, we have validated Agro 4.0 by using data from one hundred rural properties in the state of Minas Gerais, Brazil. Minas Gerais was Brazil's greatest producer of both coffee and milk in 2018. Brazil was the world's greatest producer and exporter of coffee in 2018 and in previous years as well. Brazil was also the worlds' greatest exporter of both beef and chicken in the same year. There were more cattle heads than people in Brazil in the year 2018, according to FAS/USDA.

The rest of this work is organized as follows. In Section 2 we present the theoretical basis of our work describing concepts of agroecosystems sustainability, and data science. In Section 3 we discuss the related work, concepts of agroecosystems, sustainability and data science. In Section 4 we present our platform and its architecture. In Section 5 we describe a case study based on rural properties located in the countryside of Brazil. Finally, in Section 7, we conclude our work.

2. Fundamentals

In this section, we describe the theoretical fundamentals needed for the comprehension of this paper. Section 2.1 describes the main concepts related to agroecosystems sustainability, focusing on the ISA model and methodology.

2.1. Sustainability

One of the current challenges faced by productive systems is balancing sustainable production with societies' needs, or market demands. Certifications are used, on the industrial sectors, to reduce the environmental impacts of such activities and guide the processes involved towards improvements, so they become more efficient lessening their impact on the environment. Some of those certificates are the ISO 14001 and the EMAS [10–12](*Eco-Management and Audit Scheme*), currently being practiced in the European Union. The EMAS is more rigid, precise and yet more reaching [10] than the ISO, that being the reason it was picked for implementation by the European Union.

The elaboration of Sustainability Indicators regarding agriculture is a complex task that begins with the definition of parameters to be monitored (soil erosion, soil acidity, production efficiency, among others). The definition of these parameters and the meaning of the indicators can also be influenced by regionality or geography, noting that some parameters cannot be applied uniformly for every region, like, for instance, water salinity [13]. Acknowledging the complex nature of such task (elaboration of environmental indicators) and taking that into account, Rogmans and Ghunaim [14] and Coteur et al. [15] proposed guidelines for the definition of

[1] Demonstrations of the Agro 4.0 system available at: https://agro.sybers.dcc.ufmg.br/promotion

such indicators. Singh et al. [16] studied 41 methodologies for computing and estimating sustainability indicators, each with recommended scenarios and use cases, citing three: indicators for urban development, environmental vulnerability for cities and indicators for green policies effectiveness.

2.2. Green information system

The Green Information System - or Green IS - label has a broad definition. According to Watson et al. [2008], it refers to an Information System that supports or enables sustainable initiatives and addresses environmental issues, aiming to reduce an activity's impacts on the environment. The Green IS thus has an indirect impact on the environment, through the positive impact it has on the activity it supports [17,18].

In literature, a related term is Green Information Technology, or Green IT, which is an Information Technology practice or study focused on reducing the first order impacts of IT activities on the environment. Examples of Green IT practices are introducing energy-efficient hardware to an IT operation or providing a sustainable framework to handle the disposal of IT equipment. Green IT is often related to hardware; it is related to software when the software focuses on mitigating the immediate impacts of an IT activity [17,19,20].

Examples of Green IS, referring to Information Systems that indirectly affect the environment by improving the sustainability of activities those give support to, can be: IS that aims to provide support to supply chains, optimizing routes and transportation; IS that monitor environmental variables such as water and energy consumption, waste, emissions, toxicity and carbon footprints of an industry, among others [21].

We argue that the system presented in this work is a Green Information System, precisely because it is used to register and monitor, yearly, variables such as water contamination and quality, usage of agrotoxics, soil quality and contamination, size and status of legal preservation areas, waste management, among others, of an agricultural economic activity, so technicians can help producers in improving their business' sustainability, lowering their impacts on the environment.

2.3. Indicators of sustainability in agroecosystems

A way to evaluate the sustainability of rural properties and farming businesses is the System of Sustainability Evaluation (SAS) [22], applied to measure the sustainability of ethanol and sugar cane productive businesses on the state of São Paulo. This methodology, despite being well detailed and accurate in some aspects, regarding air quality measurements, for instance, is not generic enough to apply to rural businesses with other productive profiles.

The ISA project is an initiative of the State Secretary of Agriculture, Pecuary and Supplies of Minas Gerais (SEAPA), Brazil. The methodology proposed by the ISA project allows a detailed check on a target rural property, highlighting a compounded analysis of their production systems, information management, water and soil qualities, natural habitat preservation, employment conditions and quality for the workers, among other characteristics. The ISA Platform is accessible on http://www.epamig.br/projeto-isa/. Environmental Sustainability is also a factor of economic interest for the municipalities that house rural businesses. For instance, some Brazilian states, such as Paraná, São Paulo and Minas Gerais reward municipalities that take good care of their natural environment by providing tax benefits, as measured by the system [23].

Table 1
Sub-indexes and indicators of ISA.

Sub-index	Indicators computed for the score
1. Economic Balance	Productivity
	1. Income Diversification
	2. Assets Development
	3. Degree of Indebtedness
2. Social Balance	5. Basic Services Availability
	6. Scholarship
	7. Work/Employment Quality
3. Business Management	8. Business Management
	9. Information Management
	10. Residues Management
	11. Work Security indicators
4. Soil Productive Capacity	12. Soil Fertility indicator
5. Water Quality	13. Water Quality
	14. Contamination Risks (concerning the usage of pesticides)
6. Handling of the Production Systems	15. Soil Degradation Evaluation
	16. Conservation Practices Adoption
	17. Roads Quality indicators
7. Ecology of the Rural Landscape	18. Native Vegetation
	19. Permanent Preservation Areas
	20. Legal Reserve Area
	21. Landscape Diversification indicators

ISA is composed of 21 indicators which values are in the interval [0; 1] [9,24,25], grouped by sub-indexes, as detailed in Table 1. Those indicators are computed by the application given the user input for each equivalent section of the interface.

Each ISA Methodology Indicator is calculated by different associations between their required data input [25,26]. Each input is usually fed to an impact curve, which represents how that input affects the Indicator's score, and such curves are represented by multi coefficient equations (summations), which yield a final number. The same kind of data input can have its final impact calculated by different impact curves varying for different value thresholds of itself or other related variable(s) (for instance, for the Soil Fertility Indicator, the amount of clay - *dag/kg* - in the soil sample dictates which impact curves all the other variables that compose that indicator will be applied to before arriving at the final Soil Fertility formula).

The coefficients are constants obtained by empirical experimentation through sensibility and probability tests for each Indicator, and also by the input of specialists about how a variable should affect an Indicator [25,26]. Those coefficients can also be interpreted as constants in weighting matrixes *C*, each row *r* representing an equation to be used for a threshold of the value *i* (for input) or other value within the scope of the related Indicator, targetted by the matrix and each column *j* to *n* being the coefficients of the multi coefficient equation, which form the impact curve of that input variable on it's designed indicator. So the impact curve is summed to a single final number, given that a row *r* was selected by the input *i* threshold, by the formula:

$$\sum_{j=0}^{n} C_{r,j}\cdot i^{j} \tag{1}$$

In the end, an indicator is calculated by a function that maps from 0 to 1 some relation made of the results of those multi coefficient equations applied to each of the Indicators' inputs [9,24,26,27]. The full calculations for the Water Quality and Soil Fertility Indicators are displayed in the Appendix A, as they are used for the simulations in Section 5. All indicators calculations are detailed, discussed and implemented in FAEMG's ISA Reference Spreadsheet [25].

2.4. Agro 4.0

In [8,28] Agro 4.0 was introduced. The ISA Methodology was conceived by the Enterprise for Agropecuary Research of Minas Gerais (*Empresa de Pesquisa Agropecuária de Minas Gerais* - EPAMIG) and it's original implementation consisted of using Microsoft Excel for the questionnaire collecting the data for each property, with a complex Excel worksheet that computed all the indicators, sub-indexes and other immediate results on the go [9,24]. The Excel approach was deemed not suitable for analysis extrapolating more than one property. The sheets could slightly vary in format and patterns of form filling, not all data could be guaranteed to be properly validated, and there was no tool to aggregate and extract information regarding a set of properties, limiting the analysis to technicians manually extracting information from various sheets with hundreds of fields each.

Agro 4.0 is an Information System that offers a centralized data collection and visualization approach for the ISA Methodology. Solutions were developed to make it easier for agrarian technicians to collect the ISA questionnaire data when they visit a target rural property and later interpret the results obtained. Data from various properties are collected, structured and processed to generate different reports and diagnostics for each participating rural property. There are web and desktop modules to access, query data and write to the system, detailed in Section 4.

2.5. Feature selection

Feature Selection is a machine learning/statistical technique whose main objective is to reduce the amount of data fed to models by selecting a subset of variables on which to base the models on, this way simplifying the models while keeping or even improving their accuracy or another desirable metric. This technique is usually done by selecting the most relevant attributes (not items!) present on the dataset or by dropping redundant or noisy ones, the idea is to have a subset of the initial variables good enough or even better than using the total amount of variables, to operate the prediction models. Feature Selection is a problem in the class of $O(2^n)$ time complexity: the naive approach is to check the performance of the models for every permutation of variables possible, an exhaustive search, and that is usually very time prohibitive. What is done in the literature and every day practice to remedy this nature of the problem is to use heuristics and other tools developed in the statistical and machine learning fields [29,30]. In this work, we have used the InfoGain [31] and CFS [30] techniques as part of our simulations.

The feature selection algorithm InfoGain [31] works by measuring the information gain, or loss of entropy (H, Eq. (2)) with respect to the classes (*X*), for each feature of the dataset. Entropy is a measure of impurity, ranging from 0 to 1, the closer to 0 the entropy of a group is, the more homogeneous it is.

$$H(X) = -\sum_{i\in X} P(Class = i)\cdot \log P(Class = i) \tag{2}$$

To measure the Information Gain for a given feature - $H(Class|Feature)$, Eq. (3), InfoGain separates the initial group, putting each object in a group representing one of the discretized values (*V*) of the target feature, then measures the entropy (*H*) in each group, and sums those values of entropy weighted by the size of each group by the initial number of samples. Entropy is calculated by the formula below:

$$H(Class|Feature) = \sum_{v_i\in V} P(Feature = v_i)\cdot H(Feature = v_i) \tag{3}$$

$$InfoGain(Class, Feature) = H(Class) - H(Class|Feature) \tag{4}$$

Selecting features by InfoGain consists of computing the information gain (Eq. (4)) for every feature and picking the features with the best scores, selecting scores by a minimum threshold or ranking and picking the top *n*.

CFS (Correlation-based Feature Selection) [30] is a correlation-based heuristic algorithm that selects features on the basis of two main observations: a feature is considered relevant if it has the potential to be predictive of the classes (meaning that it should exist at least one feature value v_i and a class *c* that satisfies the Eq. (5)); redundant features (ones that are highly correlated to other features in the set) are undesirable.

$$p(C = c|V = v_i) \neq p(C = c) \tag{5}$$

CFS computes a merit score based on the pair-wise Person's correlation coefficient when all the variables have been standardized and an equation adapted from test theory (Eq. (6): M_s is the merit of a *k* features subset *S,* r_{ff} the average correlation between the features of *S* and r_{fs} the mean features in *S* and class correlation).

$$M_s = \frac{k \cdot \overline{r_{fs}}}{\sqrt{k + k \cdot (k-1) \cdot \overline{r_{ff}}}} \tag{6}$$

To select the best features subset, CFS employs heuristic search strategies testing for different combinations of features in *S* while trying to increment the computed M_s, *S* with the best M_s is considered the most desirable feature selection taking into account the initial two main observations. Different heuristic search strategies can be used as well as different correlation metrics for computing r_{ff} and r_{fs}.

We use the Weka software [32,33] implementations of both algorithms.

3. Related works

The article [34] made a review of the literature for big data applications in farming and agriculture. Thirty-four articles were analysed for the tools they used and the problems they tackled. The authors note that s although Big Data is quite successful and popular as a domain, there are still very few cases of its applications on agriculture, especially on small farming, as the numbers of scientific publications and commercial initiatives show. The authors note that the five Vs of Big Data - Volume, Velocity, Variety, Veracity and Value - [35] et al. are often misunderstood as people value Volume over the rest, which is equally important. The authors also point out that the sources of data for the solutions studied are very varied: drone images, governmental institutions, weather sensors, historical information, surveys, the web, among other sensors of different natures. The review documents the most used Big Data tools, including algorithms, databases, GIS systems, statistical tools, among others. Machine learning is often used in predictions, and database solutions are very varied. The survey identifies problems that may be slowing down the pace of Big Data in agriculture, such as: privacy issues raised by farmers - regarding the ownership of the data -, security and accuracy doubts, the possibility of the creation of monopolies as valuable data is collected and concentrated by complex solutions, the access to ground information by the team behind the answers, among others. It is also noted that there is a gap of expertise and access to infrastructure in third world countries and small farms. The authors note that many farmers in all parts of the globe are organizing in cooperatives or communities, a move that empowers them and increases sharing of information and data, possibly opening new windows to introduce Big Data into their operations (in fact, the data analytics tool presented by our work was implemented on ground by FAEMG's affiliated cooperatives). The analysis of data by experts can help guide farmers, such as in one of the surveyed works [36] et al., 2011, in which the analysis of crops' responses to fertilisers allowed the farmers to manage better which fertilisers to use. The article ends by stating that Big Data has the potential to boost productivity and the development of smarter farming, allowing for an increase in production in an environmentally friendly way.

Article [37] highlights the growing importance of data analytics and informatisation in agriculture, and how the introduction of Big Data in that sector in the United States is reshaping market relations between companies and farmers. Some of the main challenges agriculture faces nowadays, as pointed by the article, is sustainability. What differentiates the current paradigm of agriculture, called Precision Agriculture, from the traditional Conventional Agriculture, is its an emphasis on data collection and analysis to guide decision making and the overcome of challenges. Soil Fertility, over fertilisation, water contamination, water availability and greenhouse gasses emissions are some of the sustainability challenges pointed by the authors of the work. All those items are taken into consideration by the ISA Methodology. Traditional farm supplier firms equip themselves with Data Analytics solutions as part of their commercial arsenal, and those solutions involve the systematic analysis of data to provide valuable information to their clients. The article points out how significant players, as well as startups, are making considerable investments in Data Analytics and Big Data solutions, seeing the potential of growth in this segment.

The article [38] implements and argues for Business Intelligence - BI - models that factor in the sustainability of a company. In the work, a Sustainability dimension is added to an existing BI model, considering the economic, social and ecologic sustainability dimensions of an exemplary generic corporate business. An example data model for monitoring Sustainability Projects through BI is also shown and discussed. The author argues that the management of corporate sustainability should rely on BI, as such a tool can provide valuable information for analysis. The author also argues that sustainability - all dimensions: economic, social and ecologic - should be part of a corporate business strategy, and so corporate data should necessarily include sustainability data, and a BI model should reflect that.

The article [39] studies how institutional pressures influence the adoption of Green IS in organisations of different nature. The authors also used a classification of Green IS/Green IT that identifies three groups, based on the contribution the deployment of such technologies has in an organisation. Pollution prevention Green IS the adoption of IS to reduce the pollution caused by other activities of the organisation. Product stewardship Green IS is the adoption of IS to enhance the lifecycle management on the supply chain. Sustainable development Green IS is the adoption of IS to transform business activities, reducing their impact on the environment. A survey involving 75 organizations is done and analysed, resulting in a suggestion that both mimetic - the imitation of behavior of other organizations, such as partners or competitors, that resulted in success for them - and coercive - regulations, contracts, market demands - pressures are essential and often result in a firm implementing Green IS to mitigate it's activities impacts on the environment.

The authors of [19] begin the article by comparing the impacts of information technology and systems on the environment, and then dividing those into two categories: First order impacts are the negative impacts of using and disposing of information systems on the environment, the effort of mitigating those is called Green IT (Green Information Technology); Second order impacts are the positive impacts of using Green IS (Green Information System) as a tool to improve the sustainability of an operation, activity or business. The article studies the reasons and results behind the adoption of Green Information Systems and uses Melville's Belief-Action-Outcome framework (BOA) to evaluate the impacts of the adoption of Green IS in multiple business environments. The authors surveyed 508 managers from various businesses in Malaysia and concluded that the managers' perception and attitude to Green IS as well as coercive pressure (by regulatory bodies, market or business partners) to the firms pushing them to become more environmentally friendly play a big part in the adoption of Green IS. The article also suggests that the adoption of Green IS had a positive impact on the environmental performance of the firms.

An Information System tackling environmental sustainability issues allows managers involved in the productive chain to make more qualified decisions, resulting in benefits regarding the social, economic and environmental aspects of their activities. Those systems, when properly implemented, can bring advantages to the groups that use them [40]. An example of that is the usage of an Information System in the management of energy, resulting in costs reduction. Another use case would be the deployment of sensors in a project for more efficient irrigation systems that could consume less water and energy [41].

A general evaluation of systems employed to help on the measurement of sustainability of agricultural and farming properties happened in Denmark [42]. Solutions for the assessment of sustainability - based on indicators - were compared regarding the process and complexity of employing them. More than 40 solutions were evaluated, and only 4 of them met all the desirable criteria and took into account the environmental, social and economic dimensions of sustainability. The RISE [43] solution was the one with the best results, and it is used to measure the sustainability of farms. The experiment concluded that the usability, complexity of the solution, language use and meeting the expected use value - by developers and farmers - of the information outputted by the solutions are factors that are weighted for the adoption or rejection of the solution.

The AESIS (*Agro-Environmental Sustainability Information System*) was initially applied to organic agriculture and then expanded to other crops [44]. That solution comprises many subsystems that generate environmental indicators for each interest point. They formulated possible answers to sustainability questions, together with critical points for the agricultural sectors of the local economic and agroecological zones, identifying thresholds for indicators and setting systems of management with the proper political parameters. This format is similar to ISA [24], but here the indicators are divided into subgroups, the critical threshold is the same for all indicators and the actions for tackling the discovered issues are defined in the correction plan.

SAFE [45] (*Sustainability Assessment of Farming and the Environment*) structures the information regarding an agro-economic system in a hierarchical manner, to evaluate its sustainability. Three levels, called Portion, Farm and Landscape are defined. That framework also aims to explore the agroecological system's data in a more generalist way to obtain a more concise result of its stipulated sustainability. On the environmental aspect, they take into account data labelled by the groups Air, Soil, Water, Energy, and Biodiversity. On the economic perspective, the financial viability of the business is factored in. For the social issue, food production quality and safety, workers' and families quality of life, social and cultural acceptance of the activity are factored in.

Regarding the adoption of Information Systems in large properties, and in large-scale, three different types of systems are identified in [46]. The first type is responsible for the prediction of future land uses, based on the extrapolation of current tendencies. The system employs measurements verified in the past to identify future states. This process requires quality and precise measurements [47] so that they can create simulations with an acceptable degree of trust. The second type is focused on extensive research to define the types and possible land usages. Initially, the methodology performs studies of the biophysics of the system. The land usage optimization is then made by taking into account all the objectives aimed by the employment of those lands. The third kind of system aims at the identification of policies that benefit certain and specified land usages. The definition of the objectives and specific land usages can be performed by taking into account the financial market to determine future demands and products on the rise [48].

The low usage of Information Systems by farmers and other rural properties owners could be explained by the immediate economic impact produced by the adoption of the technologies. Beyond the economic factor, it is also noted that age (of the people whose technological solution is aimed at), educational level and the size of the rural properties are also important factors that weight in the adoption or rejection of new technologies [49]. In the year 2000, it was predicted that industry restrictions and environmental regulations would force the adoption of support technology by farmers and other rural properties owners [50]. Besides restrictions and resistance by smaller producers [49], the usage of FMIS (Farm Management Information Systems) is indispensable for high precision agriculture [51] (High precision agriculture defined as in "eletronic monitoring and control applied to data collection, processing and usage for support in decisions regarding the temporal and spacial allocations of supplies for crops" [52]).

Fountas et al. [53] analyzes 141 international FMIS packages, grouped into 11 categories according to the activity the packages support: Field Operations, Better practices, Finances, Inventory, Trace-ability, Reports, Local Specific, Sales, Machinery Management, Human Resources and Quality Assurance. Commercial solutions from France (10 solutions), Germany (16 solutions), Italy (16 solutions), The United States (62 solutions) and Canada (4 solutions) were evaluated. In the same article it was stipulated that 75% of the solutions are developed for personal computers, 10% only worked on mobile platforms, 9% are developed as web systems only, and only 6% of the solutions provide modules for both the mobile and web platforms [53]. In the end, the authors proposed a new classification to group the packages, by complexity and activities covered by their features: Basic, Sales Oriented, Local Specific and Complete.

The use of sensors facilitates the automation of various processes in agricultural properties. Abbasi et al. [54] analyzes different use cases for wireless sensors in signals monitoring. Some of the kinds of signals that can be monitored are temperature, humidity, rain, water levels, conductivity, salinity, hydrogen, CO_2, winds' speed and direction, atmospheric pressure, among others. Acquisition costs, network types and their use scenarios are also analysed.

Efficient management of water resources can also be achieved with the employment of auxiliary Information Systems based on the Internet of Things. Kim et al. [55] describes the implementation of a wireless network - composed of wireless sensors and specialized software - for the control of a precision irrigation system. Six sensors stations were installed and distributed on the targeted field following a soil's properties map. Periodic samplings produce data sent to a processing center. That central unit analyses the situation and decides on irrigating specific points (georeferenced by *sprinklers*) on the field at a given time or not.

Precision agriculture is not solely defined by the adoption of precise tools, as its implementation has other impacts on the way the farms work, and how the farmers labor. Precision agriculture changes the main practices and laboring methods on a rural property. It employs a diverse range of technologies, from GPS devices (Global Positioning System), to GIS (Geographic Information System) and IoT (Internet of Things), the first is used for the elaboration of the topographies of the rural properties, and precisely positioning sensors. The later two can be implemented as a georeferenced database that stores relevant information regarding soil, and its relief, for example. The monitoring of production can be executed using sensors logically scattered along a field, remote sensing can provide satellite images and other area data for the identification of problems on crops, and with that sort of data and others collected by sensors, the exact quantity of nutrients or defensive chemicals can be administered to target areas [56].

In [57], the authors built a C#.NET application as an environment to simulate virtual machine migration scenarios inside data centers. An optimization problem was formalized, aiming to reduce the monetary costs of the operation while utilizing as much green energy as possible while decreasing the overall energy cost too. The winner strategy was Joint Optimal Planning, which was not only green-energy aware, predicting the amount of green energy available before starting an operation, but also took into consideration the temperature of the air and the cost of the cooling systems of the data center, using stochastic search to reach close-to-optimal solutions.

The work [58] presents a mathematical modeling of water quality attributes of the Izvorul Muntelui Lake in Romania, tackling the body of water's eutrophication issues. Simulations on the model were executed, for different scenarios, aiming to identify solutions for the rehabilitation of the lake. Among the variables taken into consideration, were Nitrogen (mg m^3), Phosphorus (mg m^3), phytoplankton biomass (growth rate, day^{-1}) and Turbidity. The model was validated with real-world data from 2007 and 2008, and among other solutions, it was noted with the simulations that the restoration of the lake may involve phosphorus inactivation or artificial circulation.

In [59] a message exchange in distributed systems model was proposed, aiming to assist in the minimization of the sum of request and response messages on a system, specifically during its management phase, while optimizing energy consumption and the overall system performance. Using graph theory, the model participated in simulations of different scenarios of message exchanging and network setup.

The authors of [60] proposed a sim-heuristic solution for scenarios of agri-food supply chains when there's a single central warehouse providing items such as food and related biological products prone to deterioration to nodes of stochastic demand. The decision problem involves inventory management, minimization of food waste and optimization of delivery routing. The authors' solution, in the form of an algorithm, use Monte Carlo simulations to mimic the stochastic aspect of the demand and uses a mix of integer programming and local search meta-heuristic to find the optimal minimal cost solution. Their work is applicable to similar scenarios involving perishable inventories.

In [61], the authors review traditional data-fusion algorithms, as well as newer machine-learning and deep-learning-based ones, on the task of diagnosing and predicting mechanical faults. The authors go into detail discussing the different approaches and data treatment implied by the discussed techniques. Generally, data is collected using IoT sensors, preprocessed then fed to prediction algorithms. A simulation was developed and performed to benchmark the performance of six data-fusion algorithms based on neural networks and the challenges of data-fusion applied to multi-source sensing for fault diagnosis were discussed.

4. The application

Agro 4.0 was presented originally in [8,28]. Development has continued in this work. In the following session, we introduce the system's core concepts, briefly describe the system's architecture as well as the new contributions of this work (hierarchical users grouping, a Simulations module, and visualizations).

4.1. Architecture

Agro 4.0 [8] was deployed and tested with participating rural properties from the state of Minas Gerais, Brazil, from the second semester of 2016 until the first semester of 2017. The Federation of Agriculture and Livestock of the State of Minas Gerais (*Federação da Agricultura e Pecuária do Estado de Minas Gerais* - FAEMG) was the institution that applied the system developed in this work in real

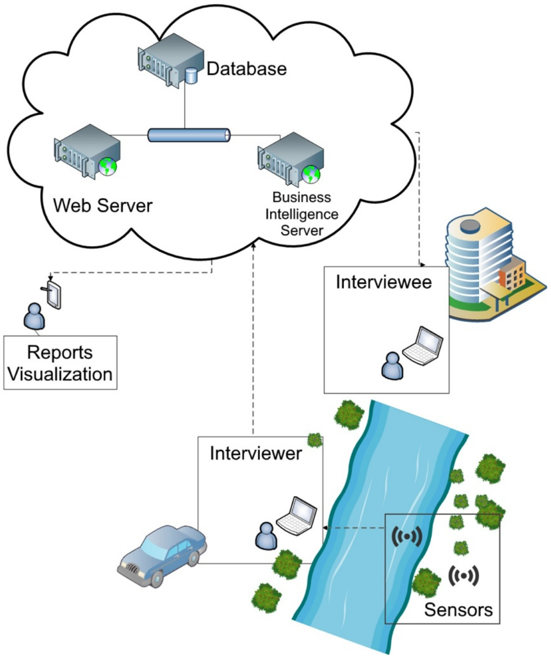


Fig. 1. System architecture.

properties, through their technicians.

Fig. 1 shows the architectural aspects of the Agro 4.0 solution. Agro 4.0 is a multi-platform system, featuring independent components for different tasks (data collection, storage, processing, mining, and visualization). Fig. 2 later shows the software stack of the system, in which each block depends on the ones bellow them in the same column. In Appendix B an UML process diagram details the most relevant human interactions and information delivery processes of the Agro 4.0 system.

The data is primarily collected on the property through an interview made by an able technician with a representative of the rural property. The interview is intended to collect all the data required to fulfil an ISA questionnaire. The questionnaire is composed of hundreds of fields (detailed in [24]), many of them requiring technical and precise data (i.e. the pH acidity level of the water streams on the property, the proportion of metals collected by soil sampling, age and training level of employees, among others) [9,24], thus the interview process can take more than a day. A ISA questionnaire refers to the state of a rural property in a given year, however, in Agro 4.0, each questionnaire is also linked to a project.

An Agro 4.0 project is a set of properties grouped by geographical regions and closed time windows [8]. In the questionnaire, to be filled through a Java desktop client, the Interviewer attributes it to a project, date and the property the questionnaire refers to. A project is also always associated to an institution, that is responsible for it.

The module that receives and validates input data is a Java 8 and JavaFX desktop application. This client sends formatted and

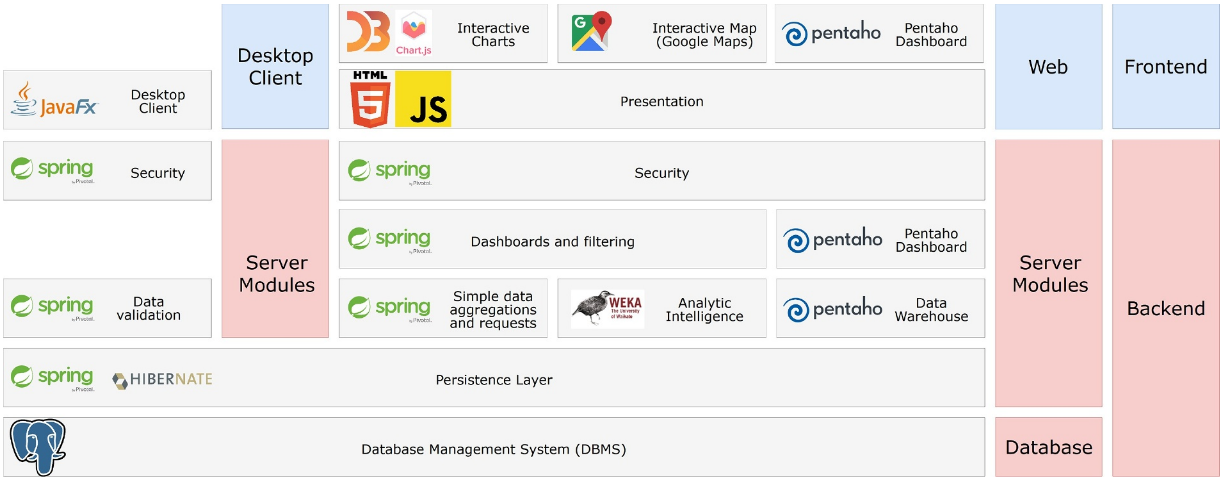

JavaFx Desktop Client
Desktop Client
D3 Chart.js Interactive Charts
Interactive Map (Google Maps)
pentaho Pentaho Dashboard
HTML JS Presentation
Web
Frontend
spring Security
Server Modules
spring Security
spring Dashboards and filtering
pentaho Pentaho Dashboard
spring Data validation
spring Simple data aggregations and requests
WEKA Analytic Intelligence
pentaho Data Warehouse
Server Modules
Backend
spring HIBERNATE Persistence Layer
Database Management System (DBMS)
Database

Fig. 2. Agro 4.0 architecture: software stack.

validated data to a remote database, or stores it in the user's (agrarian technician) disk - with the classic save to file button approach - until they decide it is complete and appropriate to be sent. This client also allows the user to download data sent in the past and update it if they wish.

The data is collected in the format of a questionnaire, with multiple tabs and fields. This data collection client is not a web application because some or many rural properties may not be equipped with an Internet connection. The database that receives the data collected by the system is used by the data warehouse, mining and visualization modules for the generation of reports and analysis.

When the questionnaire is ready to be sent, the technician does so, and the data is thus sent a Java EE server (built with Spring MVC and Hibernate, running on Wildfly 15), which stores it in a PostgreSQL database shared with all the other modules. The other modules, running on the same server, are now able to generate the reports, perform the intelligence analysis solutions, feed the data to the data warehouse module and aggregate the questionnaire's header data (date, water basin and municipality) in the sets available for user-guided filters. The server also allows the technicians to fill the Adequation Plan through the browser (the plan to increase a property's sustainability index, as described by the ISA model [9,24]) and the representatives to see it.

One of the advantages of this new solution, when compared to the traditional method of application of the ISA methodology (Excel sheets manually filled by agrarian technicians), is the centralized and structured data storage on the Agro 4.0 server, using a Relational Database Management System (RDBMS), while also having additional rules for data integrity validated by the persistence layer of the Java EE server (Hibernate). This advantage allows for the automatic processing of data, creation of more complex and trustable analysis and also makes the system more flexible to the additions of new modules in the future.

Reports are generated for each questionnaire that is collected and sent to the server on the go. A user with sufficient permissions can also request dynamically generated reports for a collective of properties: it is possible to filter sets of properties by specific characteristics (location, year of data input, associated institutions, and others) on the fly. Those reports can be accessed through any modern web browser that supports HTML5. The charts and graphs are plotted using either D3.js or Charts.js - depending on the report.

A Data Warehouse (DW) solution is used to provide more complex and sophisticated data analysis to Managers. The Data Warehouse module is implemented with the data analytics server Pentaho Community Edition, by Hitachi Software. Additionally, with this data warehouse software, more capabilities for further integration with other databases, or future hypothesis and cause-effect investigations, are available.

A software stack overview can be seen in Fig. 2. Each grey block on the image represents a layer that depends on the blocks layered bellow them on the same column. The top blocks, at the same level of the "Frontend" blue block, represent the browser and desktop clients stack. The front-end blocks depends on the back-end stack to exchange data. All access is regulated by a security layer, managed by the Spring Security software, and all data ultimately comes and goes to the central database, managed by Postgres.

Through the web application module, allowed users can visualize useful information for various sets of properties, interactively adjusting and filtering by different properties' or questionnaires' characteristics. They can also request reports for a single questionnaire. All those reports and information aggregation are the results of processing the questionnaires sent by the Interviewers, for each property and year. The system provides three dashboards for data aggregation and visualization. Each dashboard is dedicated to different kinds of information: obtained by simple aggregation and processing; obtained by data mining; generated by the data warehouse solution. A list of each report generated by each questionnaire received is also available so that the user can request for its results, and a map for each questionnaire and property, showing the locations to which those refers to (a form in the questionnaire asks for the properties' coordinates, they are used to plot the property in that map) is also present. The Interviewers can visualize, aggregate and query for all the properties' questionnaires they sent, while the Interviewees or property managers can only see the reports and data related to their properties.

In this work, a hierarchy of visualization and write permissions as well as of users was formalized and introduced into the Agro 4.0 system. In [8] and in the previous implementations of the ISA Methodology used by FAEMG, there were only Interviewers and Interviewees. We made it so Agro 4.0 involves four kinds of human actors, described in Table 2. All human actors have login credentials in the system. The Managers and Interviewees mainly use it to visualize data and information, the Interviewers use to send

Table 2
Actors of the Agro 4.0 system.

Actor	Description
The Interviewer	Actor, most of the times a trained agrarian technician, responsible for filling - through an interview with people responsible for a property - and sending the questionnaires for each property. The technician interviewer is also responsible for formulating a diagnosis and suggesting actions to be taken, by a responsible for the rural property, to further increment the sustainability indicators.
The Interviewee	The person interviewed during the data collection stage, aiding the technician to fill in the forms of the questionnaire for the property they are being questioned about. They also receive diagnosis and suggestions by the technician, and can check results and technician notes for their property on the system.
A Project Manager	An Actor, for example, an Agrarian Engineer -, that can visualize the gathered data and its results for a (or multiple) sets of properties in one or more projects they manage. Can add new Interviewers to the system.
The Institutional Manager	A representative of a cooperative, union or other kinds of associations of rural properties, has the ability to visualize the gathered data and its results for properties and projects linked to their institution(s). Can register new Institutional Managers, Project Managers and Interviewers to the system.

the questionnaires collected through the client, visualize their reports and write Correction (Adequation) Plans. the Managers are able to visualize and aggregate more data and information than the technicians and interviewees, which are limited to reports and information regarding their questionnaires. The managers can also add new user to the system and set them to act like one of the four actors. The technicians are the only actors writing ISA Methodology related data to the system other than the automatic solutions for data analysis. They are also the only actors that interact with the system outside of the web browser (they use the Java desktop client to send data).

According to the original ISA Methodology, The workflow of the Interviewer comprises of three steps. For the first step, the Interviewer goes to a designated rural property and applies the questionnaire, filling it in an interview with a representative of the said property. Once that task is done and there is internet available, the second step comes in: the Interviewer submits the questionnaire to the Agro 4.0 server and checks later, through the browser, the report the system generated for the questionnaire. The third step consists of the Interviewer writing the Adequation Plan, also a part of the ISA Methodology. The interviewee can access the Adequation Plan written by the technician on the web, it contains orientation on how to increase the sustainability of a property given the results of the information collected through the questionnaire.

In our system, a Project has a Project Manager, an actor responsible for the management and supervision of a project. Managers can register, associate or remove the Interviewers from the projects they manage, they can also visualize reports of questionnaires associated to the Projects they manage. They have access to the same dashboards the Institutional Managers have, but limited to display data and information from questionnaires associated to their Projects. A Project can have more than one Project Manager.

The Project and Institutional Managers are responsible for overseeing the application of the ISA Methodology on the targeted rural properties for the Projects they manage. During the application of the ISA Methodology through the Agro 4.0 system, those professionals can monitor the work of the technicians by checking the reports and Adequation Plans written for each questionnaire submitted. They can also analyze and find patterns on the reports for the properties of a region or project, and perform other kinds of data interpretation they wish through the dashboards they have access to. Managers also manage the users on the system and the allocation of Interviewers to Projects.

Fig. 3 shows the process stream, beginning with the appliance of the ISA Questionnaire through the Agro 4.0 desktop client, and ending with the data mining techniques processing on the collected questionnaires, to aid and generate new information to be visualized in the web application. In the first level, colored in green, the steps for the appliance of ISA are presented. Initially, an agrarian technician visits a participating property, to start the appliance of the ISA Methodology. On this visit, a person (owner, manager, specialist or other with knowledge regarding the rural property) is interviewed and aids the technician with the fulfilment of the ISA questionnaire, a partial result over the collected data is generated and available in the data collection client. Data storage and management are presented in the second level. In the third level, yellow, the data visualization steps are shown. There are included: the reports generated for each questionnaire submitted, data warehouse and data mining reports processed and generated over aggregations of those reports, all accessible through different screens and sections of the web application. In the last level, red, the steps to execute the data mining techniques (involving machine learning) are displayed.

A Simulations module was added to the Agro 4.0 system, allowing for the operator to apply broad environmental changes on top of the data of ISA reports of a target group of rural properties. The operator (any user of the system with permission to do so) can pick a set of properties and select an Indicator to have it's composing variables incremented or decremented proportionally by some desired percentage. As the operator increments or decrements the sliders, the effects, and consequences are computed in real-time for all the properties participating in the experiment. A picture of the user interface of the module is shown in Fig. 4.

The Simulation module has a control panel, containing the sliders to apply proportional changes to the variables, histograms (for the variables, related formulas values, and Indicators values, up to the user's choice), and a Google Maps panel with the properties markers colored in a scale chosen by the user (colors vary in a scale by Sustainability Index Score or other Indicator). The histogram and the map are updated as soon as the user applies changes to some variable, so they can test different combinations of changes in

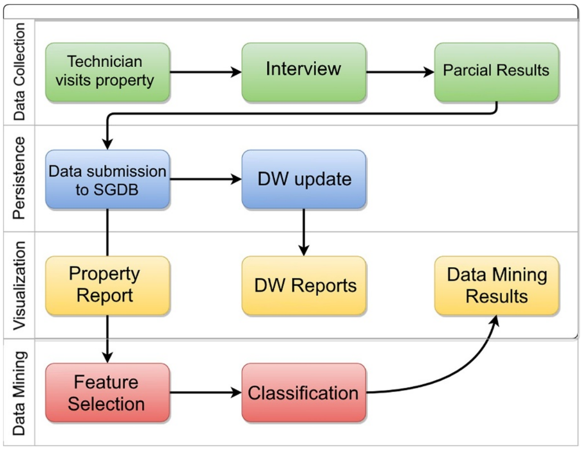
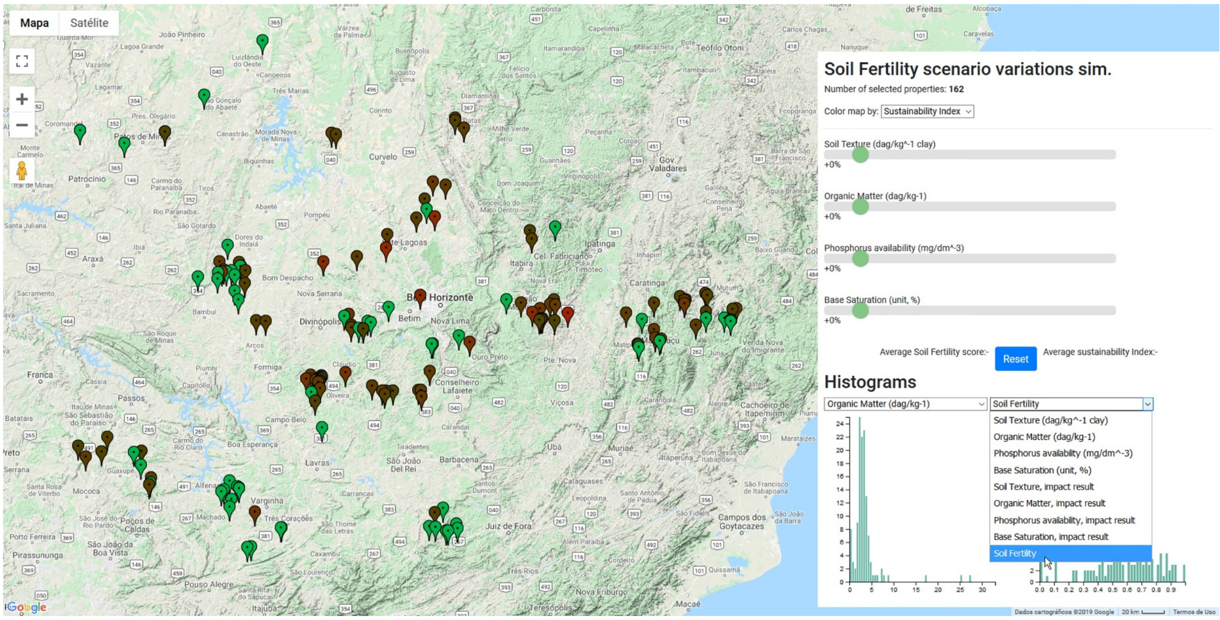


Fig. 3. Agro 4.0's steps of data processing.



Soil Fertility scenario variations sim.

Number of selected properties: **162**

Color map by: Sustainability Index

Soil Texture (dag/kg^-1 clay) +0%

Organic Matter (dag/kg-1) +0%

Phosphorus availability (mg/dm^-3) +0%

Base Saturation (unit, %) +0%

Average Soil Fertility score:- Average sustainability Index:-

Histograms

Organic Matter (dag/kg-1) | Soil Fertility

Fig. 4. A screenshot of the simulation module in Agro 4.0, on a panel to simulate changes on variables related to Soil Fertility.

target variables and see in real-time how the properties' scores are affected. The formulas and procedures to compute each indicator were rewritten and implemented in Javascript, the procedures are available and executable in the client's browser (for the subset of properties they have permission to see).

4.2. Analytical intelligence

Analytical Intelligence methods are employed to generate more complex and detailed charts, aimed to aid Managers in deciding what actions to take to increase their properties' Sustainability Indexes further. The data (questionnaires) fed to those methods can be filtered by: Project, Year, Main Income, SENAR Region of the properties, Coffee Producing Region, State of the Union, Meso Region, Micro Region, Municipality, Water Basin. Only managers can access the panels generated by such methods.

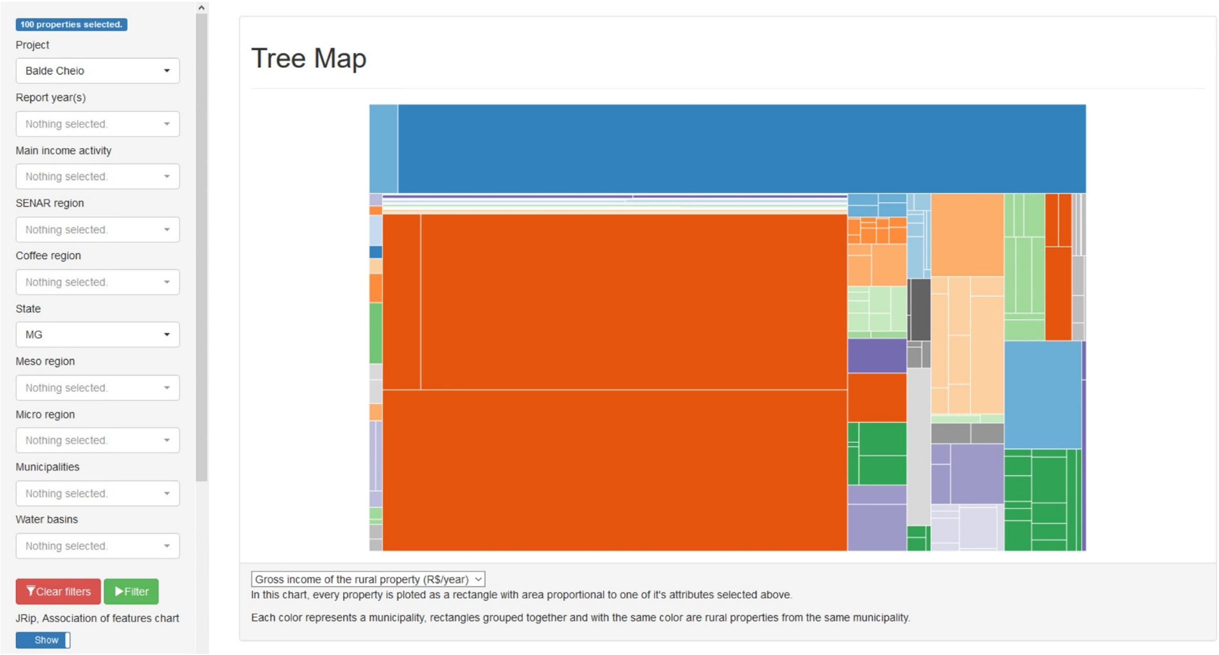
The Analytical Intelligence techniques present in this module of the system for the user to interact with include charts displaying the results of JRip (RIPPER, an algorithm which finds relations and rules of association between features that generate a predictable classification outcome [62]) and CFS [30], computed in the backend by Weka [32,33]. Other tools are available in the format of charts and graphs, such as a Word Cloud (of most common words filled into the free text forms of the questionnaires), TreeMaps and Scatter charts that can be generated for any set of rural properties, displaying their performance for anyone of the 21 Indicators (Table 1).

Finding association of attributes rules that imply satisfactory Sustainability Indexes can also help the identification of good practices to be implemented on properties currently presenting unsatisfactory Indexes. Through the JRip Sankey Diagram, one can see trends of attributes' values leading to similar Sustainability Index in the filtered data set. The drawing shows which attributes are associated for each rule, how many properties presented that trend, and each rule's relevancy for a bad, average or good Sustainability Index. An example can be seen in Section 6, Fig. 14.

An example of the system's Analytical Intelligence capabilities is shown in Fig. 5. The TreeMap chart allows for comparisons between the geographical area of each rural property with their performance in some Indicator, total area or yearly gross income, a criterion chosen by the user. In the example Figure, the chart is displaying the 100 properties in the Balde Cheio dataset, collected for 2016, as rectangles with areas proportional to their yearly gross income. When the user hovers the mouse pointer over the rectangles, the interactive chart displays information about that property. The chart shows that the top five municipalities with greater gross income represent more than 50% of the total quantity of money registered in that dataset. Those are all located in the south and south-west of the state of Minas Gerais, known for more industrialized milk-producing techniques and highly skilled workforce when compared to the rest of the state, also shown in the chart.

5. Methodology

In this section, we present experimental case studies using the dataset of rural properties collected using Agro 4.0, the system presented in Sections 2 and 4, by a program called Balde Cheio. The data set is made of one register per year by rural property, and it contains the data and fields specified by the ISA Methodology for each property. In Section 5.1, a characterization and analysis of the data set is explored. These analyses are relevant for all further discussion presented in this paper. Then, two kinds of simulations are



100 properties selected.

Project: Balde Cheio

Report year(s): Nothing selected.

Main income activity: Nothing selected.

SENAR region: Nothing selected.

Coffee region: Nothing selected.

State: MG

Meso region: Nothing selected.

Micro region: Nothing selected.

Municipalities: Nothing selected.

Water basins: Nothing selected.

Gross income of the rural property (R$/year)

In this chart, every property is ploted as a rectangle with area proportional to one of it's attributes selected above.

Each color represents a municipality, rectangles grouped together and with the same color are rural properties from the same municipality.

Fig. 5. Agro 4.0's TreeMap plot of the properties in the Balde Cheio 2016 dataset, area proportional to each rural property's yearly gross income. All the properties are located in the state of Minas Gerais. The top five municipalities with greater gross incomes in total are, according to that chart, in descending order, the municipalities of Patos de Minas (shock orange), Coromandel (deep blue), Paraguaçu (lighter orange), Estrela do Indaiá (orange) and Luz (green).(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conducted using the dataset.

The first simulation consists of Feature Selection experiments on different cuts of the dataset to simulate the calculation of the ISA Sustainability Index using a reduced amount of data about each property. The aim is to present models that, classifying the properties between acceptable and undesirable Suitability's Indexes using reduced data, can be a cheaper alternative to the current ISA Methodology, in terms of time and costs of collecting the necessary data.

The motivation comes from the observation that some data is expensive to get, and the ISA Methodology requires dozens of data for the computation of different indicators. For instance, pH of close water bodies, phosphorus availability in the property's soil, nitrate availability in the property's soil, coliforms by cubic meter of water, among others, are examples of data expensive to collect: for the water properties there are kits, but for the soil properties sometimes an analysis of the soil sample in laboratories is necessary [26].

The second experiment consists of a simulation varying data that make up two important indicators, Soil Fertility and Water Quality, testing for enrichment or catastrophic scenarios on properties distributed on a target geographical reason, to analyze how those ISA Methodology indicators behave given the simulated scenarios. How the Indicators are calculated is briefly described in Subsection 2.3, and the two mentioned Indicators' calculation is fully presented at the Appendix A.

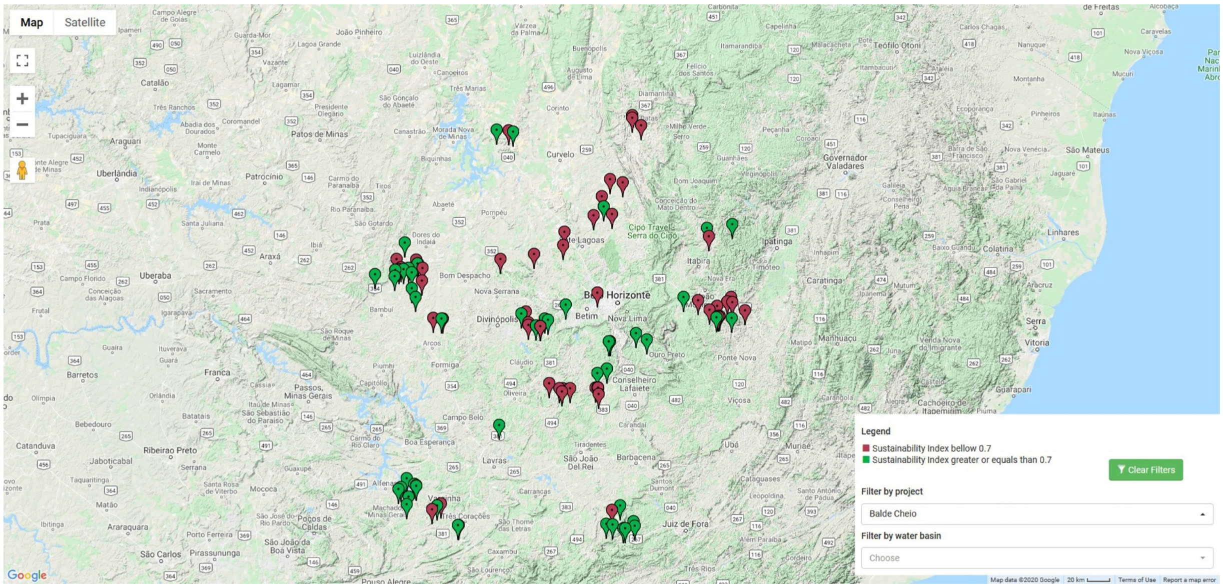
With this simulation tool, property owners, technicians and cooperative managers, for instance, can simulate environmental change scenarios or the systematic application of a strategy to mitigate some problems in the properties and predict how the performance of some set of properties is affected in the ISA Methodology. This can be a very useful tool to help people understand what they need to improve in their relations with their agroecosystems. In this work, we will perform an experiment regarding soil enrichment and a catastrophic scenario modifying the water quality of affected properties.

5.1. Dataset

The *Embrapa Pecuária Sudeste* is responsible for the creation of the Balde Cheio project. This project carries out technology transfer to milk producers and related entities. The desired results of the project are the development of the sector and the increase of the profitability for the participating rural producers. With the increase of profitability of the property, the permanence and professionalization of the workers involved in the milk production becomes more viable [63,64]. Balde Cheio is held in Minas Gerais by the FAEMG system, an institution which promotes efforts to spread the program for all regions of the state. The program consists of providing training for technicians contracted by partner entities and testing new technologies to assist producers of milk. The new technologies are also monitored for their environmental, economic and social impact.

The data used in this work was collected by the project Balde Cheio, which was applied by FAEMG, using the system presented in the former section, Agro 4.0. In another words, Agro 4.0 was one of the technologies tested by Balde Cheio in 2016.

Fig. 6 presents the properties of the state of Minas Gerais where the Balde Cheio Program has been applied and had the Agro



Legend

🟥 Sustainability Index bellow 0.7

🟩 Sustainability Index greater or equals than 0.7

Clear Filters

Filter by project: Balde Cheio

Filter by water basin: Choose

Fig. 6. Agro 4.0 : Geographical Visualization Chart, displaying participating properties in Agro 4.0 where Balde Cheio is being applied.

4.0 questionnaire sent to our system. The map was generated by Agro 4.0 , using Google Maps to plot the markers. Green markers mean the Sustainability Index for that property is greater or equals than 0.7, red markers mean it is below 0.7. In total there are 317 municipalities and 1929 participating properties in Balde Cheio, the number of properties both in Balde Cheio and Agro 4.0 , however, is 100. The maximum number of participating properties in a municipality is 112. Of the 317 municipalities, 75% of them have up to 7 properties participating in the program. In our system, there are 100 questionnaires, for 100 unique properties.

FAEMG collected data for Agro 4.0 from 100 rural properties enlisted in Balde Cheio, having their technicians visit the property and interview a person responsible for it, meeting the ISA Methodology requirements. The properties are scattered across several sub-basins as shown in Table 3. The collection was carried out between August 2016, and January 2017 in a group of properties and municipalities defined by FAEMG.

5.2. Treatment for the features selection experiments

The details of the available variables are defined in [9,24,26], our model uses 87 of those. Fig. 7 summarizes the methodology of this work: The first step consists of formatting the attributes from the ISA questionnaires data, the attributes are mapped to features and that set with data from all the participating questionnaires is called “FeaturesDS”; During the second step, the 21 sustainability indicators for each questionnaire are calculated; In the third step, the data set “IndicatorsDS” is generated by formatting in the 21 sustainability indicators computed for each entry of “FeaturesDS”; For the fourth step, *Features Selection* is applied separately to both “FeaturesDS” and “IndicatorsDS”; The fifth step consists of generating reduced versions of both data sets, filtering out the attributes that did not get picked by the *Features Selection* step; For step six, the classification algorithms are executed with both data sets; the last and seventh step consists of analysis of the results.

Table 3
Hydrographic sub-basins.

Prop. count	Sub-basin
19	Tributaries do Alto São Francisco
16	Basin of Paraopeba River
13	Basin of Pará River
11	Basin of Das Velhas River
9	Basin of Piracicaba River
8	Basin of Furnas Reservoir
5	Basin of Verde River
4	Basin of Alto Rio Grande
4	Basin of tributaries from Minas Gerais of the rivers Preto and Paraibuna
4	Basin of Piranga River
3	Basin of Santo Antônio River
3	Basin of waters around the dam of Três Marias
1	Hydrographic basin of São Francisco River

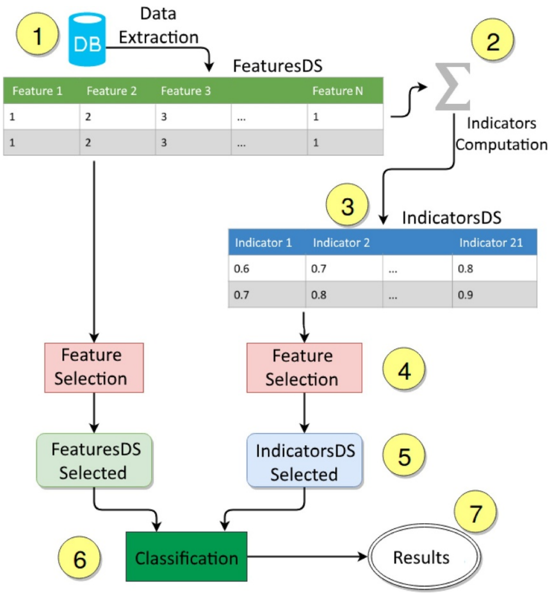


Fig. 7. Methodology.

With the database initially collected two data sets were sorted and grouped. The first set FeaturesDS is made of 87 attributes; meanwhile, the second one, IndicatorsDS, contains the 21 indicators described in the ISA Model plus the Sustainability Index (SI). In both data sets, the SI was categorized in three levels: Low for when the SI is between the interval [0; 0.5], Medium when it's between [0.5; 0.7] and High for values between [0.7; 1.0]. The High level is, according to the ISA Model, the interval in which the value is considered satisfactory, the other two levels being insufficient and meaning that the property needs intervention. In the collected data, all the properties had a SI evaluated as in Medium or High level, meaning we have a problem regarding the binary classification.

For each one of the two data sets, we applied the attribution selection techniques *Correlation-based Feature Selection* (CFS) [30] and *Information Gain* (InfoGain) available in the Weka Software [32,33]. The hypothesis was that a smaller set of attributes could reduce the time needed for the application of the ISA Model, implying the collection of a smaller set of data of a property to know its Sustainability Index, with the same or satisfactory results when compared with the data required by the full methodology.

5.3. Steps for the scenarios simulations

Fig. 8 shows the steps of this second experiment. After an exploration of the dataset and an initial analysis, interesting scenarios for simulations were thought and discussed, regions of properties on the dataset were selected to participate in the simulated scenario and then the experiment was tested on the Agro 4.0 simulation tool.

A part of the experiment consists of finding a good proposal in the form of a solution to increase at least one of the indicators for a set of properties on some region, and, indirectly, their overall Sustainability Index. Simulations on increases of positive variables were tested on different regions to see how the scores of those properties would behave.

The other part of the simulation experiment is to hypothesize a scenario that would harm significantly at least one of the indicators of another set of properties. Examples of such scenarios are soil or water contamination and drop in prices and fires. Variables with negative connotations were increased and tested in different regions.

6. Overall dataset analysis and experiments results

In this section, an initial analysis of the dataset is presented and discussed, then the results of both simulations are detailed. The first simulation (Feature Selection, briefly explained in Section 2.5) was done by using the Weka software [32,33], the second simulation was executed using the Analytical Intelligence tooling developed in this work for the Agro 4.0 software.

In Fig. 9 we can see correlations between the sustainability indicators. Some groups of variables stand out because they present

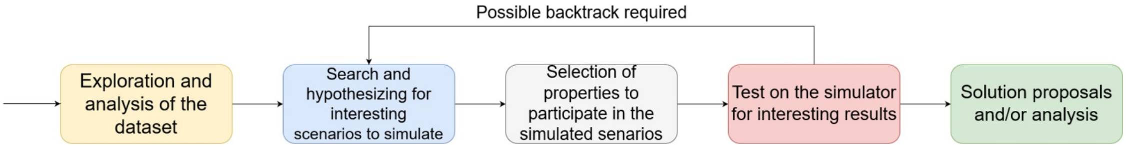

Possible backtrack required
Exploration and analysis of the dataset
Search and hypothesizing for interesting scenarios to simulate
Selection of properties to participate in the simulated senarios
Test on the simulator for interesting results
Solution proposals and/or analysis

Fig. 8. Steps of the second experiment.

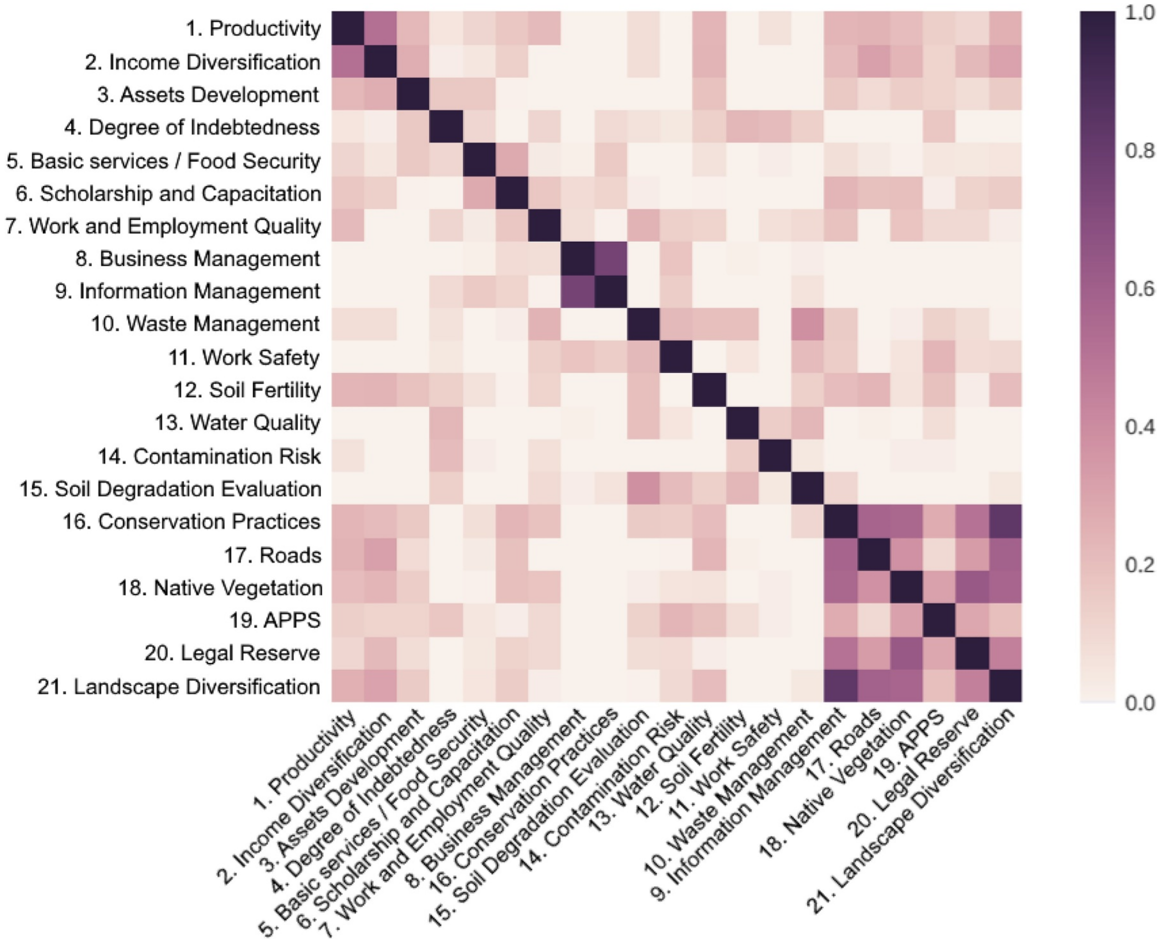


Fig. 9. Indicators Heatmap.

high correlation, such as the following groups: Income Diversification and Productivity; Information Management and Property Management; Conservation Practices, Roads, Native Vegetation, APPs, Legal Reserve and Landscape Diversification.

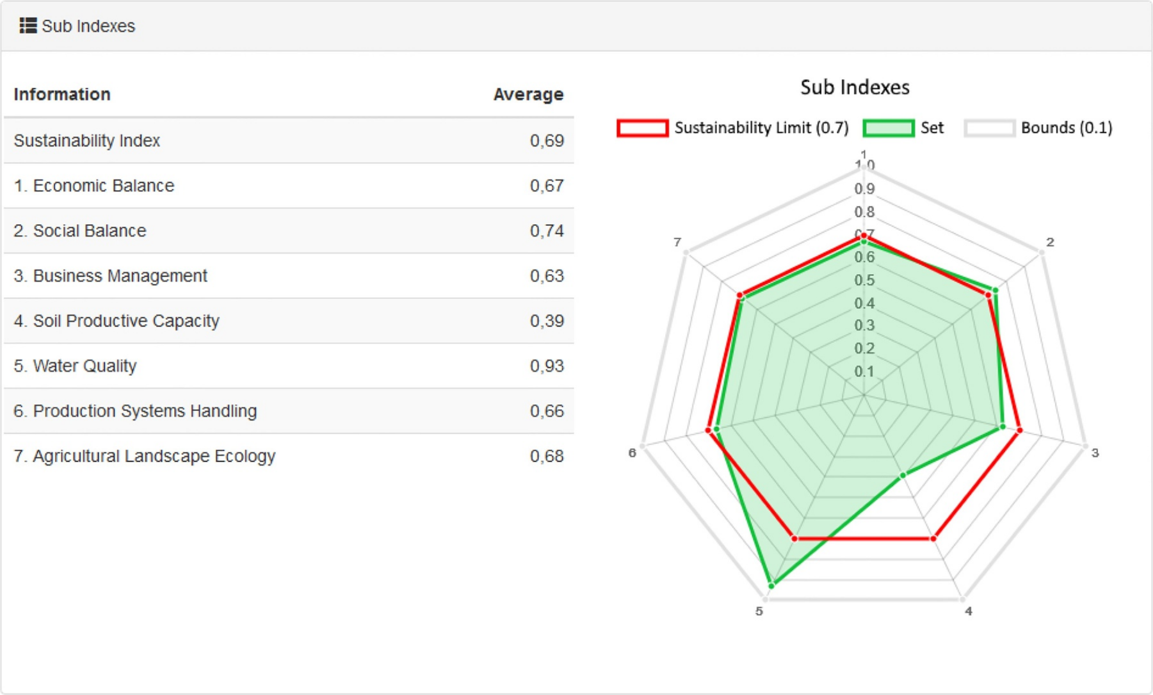
Using the general dashboard available in Agro 4.0 , we get the average Indicator and Sub-Indexes scores for the set of 100 properties participating in the Balde Cheio project, as shown in Fig. 10. We can see that this set of properties averages badly regarding *Soil Productive Capacity* in the Sub-Indexes, which is explainable by the fact those are milk-producing properties. Thus they are enrolled in a program for milk producers (Balde Cheio). The 100 properties of Balde Cheio tend to exceed in the *Water Quality Indicator*, and as the chart shows, their average is slightly below the recommended mark when it comes to *Production Systems Handling, Business Management* and *Economic Balance*. The other Sub-Indexes are very close to the desired line, on average.

Fig. 11 presents the Socioeconomic Indicators. The system show that the properties of this set present below desirable averages for *3. Assets Development, 10. Residues Management* and *11. Work Safety. 3. Assets Development* is an indicator of the Sub-Index *Economic Balance* and the other two indicators are aggregated in the *Business Management* sub-index. It's possible to see that those 3 indicators impact the results expressed in Fig. 10.

For the environmental Indicators, displayed in Fig. 12, Balde Cheio's properties do generally well in *14. Water Contamination Risk (containment)* - a part of the *Water Quality* sub-index, while also generally being located in areas of hard access and which presents, on average, a bad soil for farming - as the Indicators *17. Roads* (part of *Handling of the Production Systems*) and *12. Soil Fertility* show. On average, for those properties, the *19. Permanent Preservation Areas* (an indicator weighted in the *Ecology of the Rural Landscape* sub-index) score is low, meaning the areas reserved for preservation of the native vegetation are either below the recommended in size or are in bad conservation state.

The dashboard also displays a box plot of the Sub-Indexes in the shape of blue bars, presenting the final Sustainability Index in the last column, represented by a grey bar (Fig. 13). As pointed out by the radar charts, the *Soil Productive Capacity* indicator performed the worse for those milk producing properties, the majority of properties are below the healthy (0.7) line for *Business Management, Production Systems Handling* and *Economic Balance* as well. *Water Quality* is very high overall. As the first radar chart points out, the set has most of its properties above or near the minimum desirable Sustainability Index.

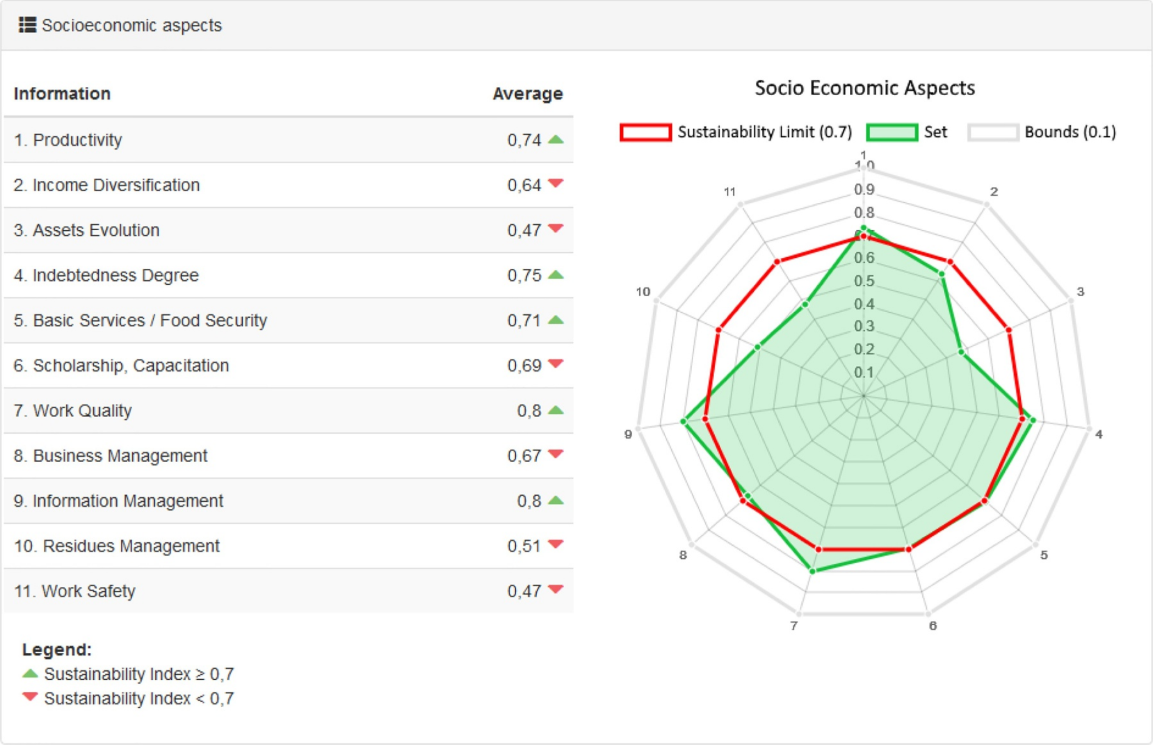
It is worth mentioning that the main dashboard panel is only capable of generating averages or sums for the sets it receives. The



Sub Indexes

Information	Average
Sustainability Index	0,69
1. Economic Balance	0,67
2. Social Balance	0,74
3. Business Management	0,63
4. Soil Productive Capacity	0,39
5. Water Quality	0,93
6. Production Systems Handling	0,66
7. Agricultural Landscape Ecology	0,68

Fig. 10. Agro 4.0 : Averages Radar Chart of the Sub-Indexes for the 100 properties participating in both Agro 4.0 and Balde Cheio.



Socioeconomic aspects

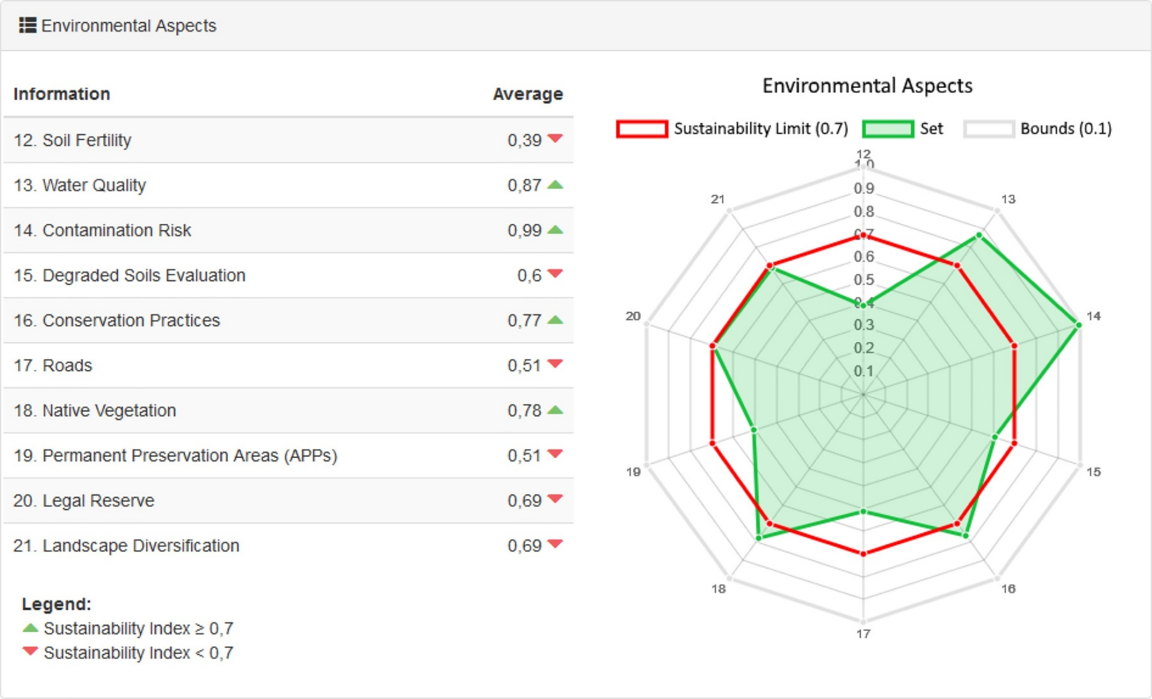
Information	Average
1. Productivity	0,74 ▲
2. Income Diversification	0,64 ▼
3. Assets Evolution	0,47 ▼
4. Indebtedness Degree	0,75 ▲
5. Basic Services / Food Security	0,71 ▲
6. Scholarship, Capacitation	0,69 ▼
7. Work Quality	0,8 ▲
8. Business Management	0,67 ▼
9. Information Management	0,8 ▲
10. Residues Management	0,51 ▼
11. Work Safety	0,47 ▼

Legend:
▲ Sustainability Index ≥ 0,7
▼ Sustainability Index < 0,7

Fig. 11. Agro 4.0: Averages Radar Chart of the first 11 Indicators for the 100 properties participating in both Agro 4.0 and Balde Cheio.

Analytical Intelligence panel, on the other hand, can show us both the results of JRip [32,33] (RIPPER, Cohen [62]) and CFS for that subset.

The JRip Sankey Diagram (Fig. 14) generated by Agro 4.0 for that Balde Cheio's set of properties shows that if the property has a



Environmental Aspects

Information	Average
12. Soil Fertility	0,39 ▼
13. Water Quality	0,87 ▲
14. Contamination Risk	0,99 ▲
15. Degraded Soils Evaluation	0,6 ▼
16. Conservation Practices	0,77 ▲
17. Roads	0,51 ▼
18. Native Vegetation	0,78 ▲
19. Permanent Preservation Areas (APPs)	0,51 ▼
20. Legal Reserve	0,69 ▼
21. Landscape Diversification	0,69 ▼

Legend:
▲ Sustainability Index ≥ 0,7
▼ Sustainability Index < 0,7

Fig. 12. Agro 4.0 : Averages Radar Chart of the Environmental Aspects related Indicators for the 100 properties participating in both Agro 4.0 and Balde Cheio.

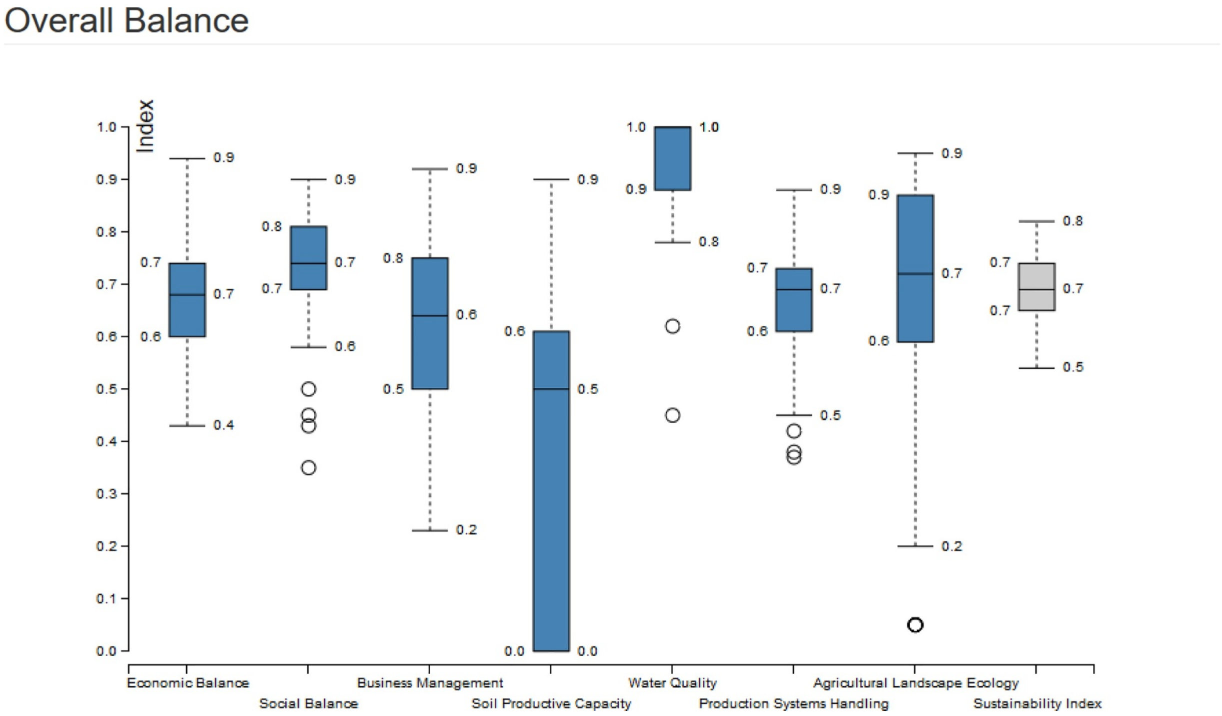


Fig. 13. Agro 4.0 : Box Plot of the Sub-Indexes and Sustainability Index for the 100 properties participating in both Agro 4.0 and Balde Cheio.

high *Business Management* Indicator it is highly likely to be ranked with a high Sustainability Index. As shown in the box plot and the radar charts for averages, the majority of those properties score below recommended for the *Business Management* Indicator.

The associative rules found by JRip (Fig. 14) for our data set shows that those - 25 properties - which managed to score greater or equal the recommended value (0.70) for *Business Management* are highly likely to also score an above recommended Sustainability

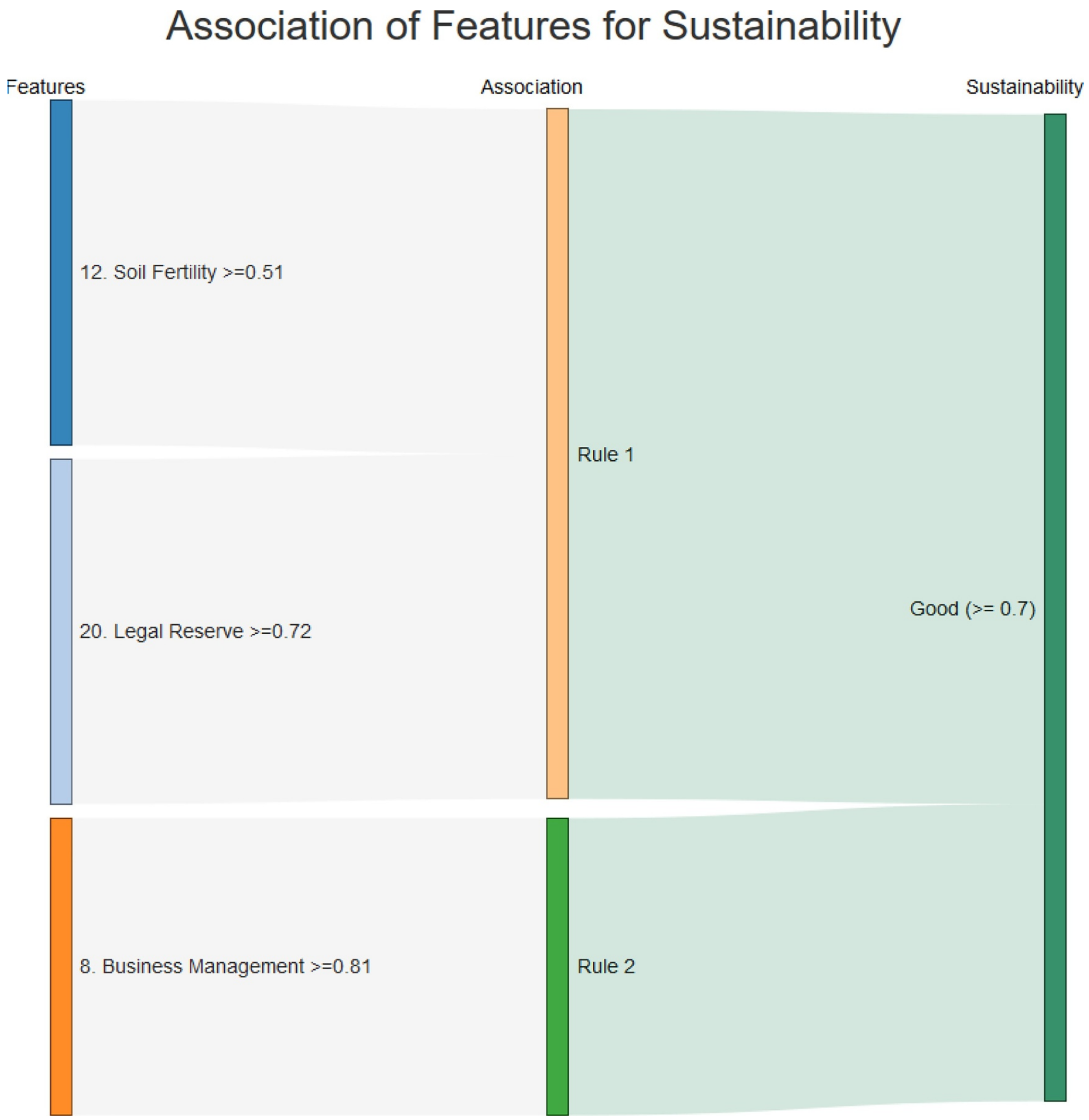


Fig. 14. Agro 4.0 : Sankey Diagram, generated by Agro 4.0, displaying associative rules found by JRip for the 100 properties participating in both Agro 4.0 and Balde Cheio.

Index. As the Sustainability Index is the average of all the Indicators, those properties also tend to do well for the other Indicators as well.

This Figure shows the most critical rules associations that lead to a good sustainability index (SI) discovered by JRIP. A good SI is characterised by a score equals to or greater than 0.70. For this dataset, the combination of good ratings on Indicators 12, 20 and 8 are characteristic of the majority of properties that also present a good SI. The association of high scores for Indicators 12 and 20, the first rule displayed in the graph, is covered by 58% of the rural properties with a good SI score. The second rule in the graph, which is having a *Business Management* score equals to or greater than 0.81, has coverage of 50%. The intersection of these rules shows us the effectiveness of the *8.1 Access to Technical Assistance* provided to the agricultural property manager: with proper assistance of a technician, the producer can better understand the soil quality of the property, the best suitable culture to plant on it and also how to treat or deal correctly with any imperfections on the soil, all affecting the 12th Indicator, *Soil Productive Capacity. 8. Business Management* is also important to ensure the financial viability of the entrepreneurship. A good compliance to *20. Adequacy of Permanent Preservation Areas* by rural properties is also essential for the preservation of the lands and water resources.

The Fig. 14 shows that for properties that got a *Soil Fertility* Indicator score near the median of the group (0.50) and simultaneously scored a *Legal Reserve* slightly higher than the recommended value (0.70) also tend to present a satisfactory Sustainability Index.

Table 4
Selected features among FeaturesDS by CFS and InfoGain.

Selection by CFS:	
Questionnaire 11.	Gross Income of the business R$/year
Questionnaire 11.	Sources of the gross income of the rural business
Questionnaire 14.2	Number of natural lakes and ponds
Indicators 8.5	Environmental Regulation
Indicators 9.1	Researches information to optimize the sells of products
Indicators 9.4	Capacity to innovate or participate in leadership within the community
Indicators 12.3	Phosphorus Availability - Index
Indicators 12.3	Phosphorus Availability - Results
Indicators 12.5	Exchangeable Magnesium - Index
Indicators 12.4	Exchangeable Calcium - Index
Indicators 12.7	Active acidity (pH) - Index
Indicators 12.9	Effective CTC - Index
Indicators 12.10	Base Saturation (%) - Index
Indicators 16.2	Level of adoption of strategies for conservation and preservation of water bodies in the rural property.
Selection by InfoGain:	
Questionnaire 11	Sources of the gross income of the rural business
Questionnaire 12.1	Facilities and other betterments (R$)
Indicators 8.5	Environmental Regulation
Indicators 9.3	Adoption of innovative techniques
Indicators 9.4	Capacity to innovate or participate in leadership within the community
Indicators 12.3	Phosphorus Availability - Index
Indicators 12.3	Phosphorus Availability - Results
Indicators 12.4	Exchangeable Calcium - Results
Indicators 12.5	Exchangeable Magnesium - Index
Indicators 12.5	Exchangeable Magnesium - Results
Indicators 12.7	Active acidity (pH) - Index
Indicators 12.9	Effective CTC - Index
Indicators 12.10	Base Saturation Base Saturation (%) - Index
Indicators 16.2	Level of adoption of strategies for conservation and preservation of water bodies in the rural property

6.1. Results of the feature selection process

The CFS analysis made through Agro 4.0 for that subset, shows that such algorithm selected the indicators *20. Legal Reserve, 12. Soil Fertility, 19. APPs, 16. Conservation Practises* and *1. Productivity* as the most relevant Indicators that weighted in the Sustainability Index of a property. *20. Legal Reserve* and *19. APPs* are a part of the *Ecology of the Rural Landscape* sub-index, *12. Soil fertility* has a Sub-Index of it's own called *Soil Productive Capacity, 16. Conservation practices* is one of the factors for the *Handling of the Production systems* Sub-Indexes and *1. Economic Balance* is partially composed by *1. Productivity*.

When using CFS with the data set FeaturesDS, that contains every attribute, the search method *BestFirst* was deployed. This method does a greedy search on the data set checking all the possibilities. The attributes that were categorical, such as the ones listed on Section 5.1, were treated as numeric, assuming the values 0, 0.5 and 1 (inexistent, partial and satisfactory, respectively). For CFS on IndicatorsDS the same setup was deployed, excluding the categorical attributes, and all data is in the interval [0; 1]. The validation method was cross-validation with 5 *folds*. The selection found by CFS for FeaturesDSis displayed in Table 4. With IndicatorsDS, the selection found by CFS is displayed in Table 5.

Environmental Regulation takes into account the use of water, Legal Reserves, Permanent Preservation Areas and proper licensing and obedience to environmental norms. The "Researches information to optimize the sells of products field" can have three values: 0, for nonexistent, 0.5 for partial and 1.0 for sufficient. Part of the Information Management sub-index, which field represents the perception of how much effort the producer or other people responsible for the property put into research information to increase sales and/or attempt to diversify and reach new buyers).

Some of the features selected contain the suffix "Index" and "Result" in the name. Those attributes are related to the soil

Table 5
Selected features among IndicatorsDS by CFS and InfoGain.

CFS	InfoGain
4. Degree of Indebtedness	4. Degree of Indebtedness
8. Business management	7. Work Quality
9. Information management	8. Business management
12. Soil Fertility	9. Information management
16. Conservation Practices	12. Soil Fertility
19. APPs	16. Conservation Practices
20. Legal Reserve	19. APPs
	20. Legal Reserve

Table 6
Machine learning methods setup.

Algorithm	Setup
AdaBoost	10 iterations, the classifier used being *Decision Stump*.
Naïve Bayes	No additional parameter, Weka's default.
J48	Confidence factor of 0.25 and the minimum number of objects per sheet was 2.
JRip	Weka's default. Quantity of data used for the validation *folds* was 3.
MLP	Learning rate was 0.3, momentum 0.2 and the number of hidden layers was half of the quantity of attributes summed with the number of classes.
RandomForest	100 iterations with unlimited height.
SVM	Implementation of LibSVM with classification type C-SVC, radial kernel, $\mu = 0.5$, $\gamma = 0$, $\epsilon = 0.001$ and loss function 0.1.

productive capacity of the property and are part of the computation that generates Indicator 12. The attributes with the suffix "Index" are obtained by measurement or analysis, made by a technician, using a physically collected soil sample from the property. The attributes with suffix "Result" are results from equations that receive the "Index" suffix attributes as parameters. The values with the suffix "Result" are intermediary to the computation of Indicator *12. Soil Productive Capacity*.

For InfoGain a *ranking* was used to select the attributes that better define sustainability. That *ranking* is constructed with individual evaluations for each attribute. Like CFS, the categorical attributes of the FeaturesDS data set were treated as numeric. For InfoGain, the data set IndicatorsDS was deployed with the same configuration, except in this case there are no categorical attributes and all data is in the interval between [0; 1]. The evaluation method was cross-validation with 5 *folds*. With the usage of FeaturesDS, the selection found by InfoGain is featured in Table 4. Using IndicatorsDS, the selection found by InfoGain is listed in Table 5.

As for the Machine Learning techniques, we set up the algorithms using the default parameters in Weka, as in Table 6. Table 7 shows the results of the execution of all the algorithms tested. The best results of each one are highlighted in bold, the very best result being the one underlined. The metrics used to evaluate the results were precision and recall. Precision is the ratio of relevant instances to the selected ones and recall is the ratio of selected relevant instance to the total quantity of relevant instances. The technique that obtained the best precision was Random Forest, with 0.94 of precision and recall.

The algorithm Naïve Bayes (NB) presented its best result when CFS was employed over the IndicatorsDS data set, with a precision score of 0.863. With InfoGain over the same data set, the best result was 0.852 of precision. The worst result obtained with CFS was running it over the data set FeaturesDS scoring a precision score of 0.764. It has a probabilistic approach and assumes that the features are independent with the target value. This is not all true since some indicators have influence in another one. For example, the business management influences the indicators of the sub-index Economic Balance.

For Multilayer Percepton (MLP) its best result was obtained using InfoGain, presenting a precision score of 0.861 for the FeaturesDS data set. For IndicatorsDS the results averaged 0.850 of precision.

The Support Vector Machine (SVM) had its best results being executed on the IndicatorsDS data set and without the feature selection, scoring 0.892 of precision. With the data set FeaturesDS the results were unsatisfactory, with 0.303 of precision without feature selection and 0.605 with CFS. It performed better than the MLB and NB even without the need of the extra step of CFS or InfoGain.

AdaBoost's has the best result in IndicatorsDS data set had 0.823 of precision. With the data set FeaturesDS the results didn't present many variations, and the best score was 0.770 using CFS. This classifier uses a set of weak classifiers that combined creates the final model. Each weak classifier is a stamp that is tree with only two leaves called as stump. These stumps are weighted and

Table 7
Results of executing each algorithm over the data sets IndicatorsDS and FeaturesDS, aided by CFS or InfoGain. Random Florest (RF), Naïve Bayes (NB), Precision (Prec.), Recall (Rec.) and attributes (attrib). The best precision scored by each algorithm is written in bold font. The best result is underscored.

Algorithm		Precision			Recall		
		All attr.	CFS	InfoGain	All attr.	CFS	InfoGain
NB	FeaturesDS	0.832	0.764	0.804	0.830	0.760	0.800
	IndicatorsDS	0.792	**0.863**	0.852	0.790	0.860	0.850
MLP	FeaturesDS	0.781	0.790	**0.861**	0.780	0.790	0.860
	IndicatorsDS	0.852	0.850	0.850	0.850	0.850	0.850
SVM	FeaturesDS	0.303	0.300	0.635	0.550	0.540	0.580
	IndicatorsDS	**0.892**	0.852	0.852	0.890	0.850	0.850
AdaBoost	FeaturesDS	0.732	0.770	0.740	0.730	0.770	0.840
	IndicatorsDS	**0.823**	0.820	0.820	0.820	0.820	0.820
JRip	FeaturesDS	0.648	0.655	0.710	0.650	0.650	0.710
	IndicatorsDS	0.740	**0.802**	0.784	0.740	0.800	0.780
J48	FeaturesDS	0.689	0.700	0.704	0.690	0.700	0.700
	IndicatorsDS	0.695	**0.800**	0.780	0.690	0.800	0.780
RF	FeaturesDS	0.840	0.790	0.795	0.840	0.790	0.790
	IndicatorsDS	0.891	**<u>0.940</u>**	0.920	0.890	0.940	0.920

combined reducing the error at each iteration. The AdaBoost has the advantage to be very fast with good result even with those simple classifiers. It can also provides the relevance of each feature, indicator in our case, in the final classification.

JRip presented its best result for the IndicatorsDS data set, with 0.802 of precision for CFS. Using FeaturesDS which is another data set, the best precision obtained was 0.710, using InfoGain. One of the main advantages of this algorithm is that it produces readable rules as C4.5 rules and also is well fitted on continuous datasets [62] like our case. Comparing it with J48 can been that the result were closer each other. Using Random Forest (RF) the best result was obtained using the IndicatorsDS data set and CFS, with 0.940 of precision. With FeaturesDS the best precision was obtained without the feature selection, with 0.840 of precision. The RF uses a combination of classification trees to produces the classification of the model. This technique generalizes well delivering consistent good results in new data sample. The only disadvantage of it, in our case, is that it is hard to check the rules that produce a result.

Thus, the Random Forest (RF) algorithm presents the best performance (94.0%) of all algorithms, using seven features selected by CFS feature selection method: *4. Degree of indebtedness, 8. Business management, 9. Information management, 12. Soil fertility, 16. Conservation practices, 19. APPs* (Adequacy of Permanent Preservation Areas) and *20. Legal Reserve (Legal Reserve)*. This result is important because with only 7 of the 21 indicators we can infer with 0.94 of precision the sustainability level of a rural property.

6.2. Results of the scenarios simulations

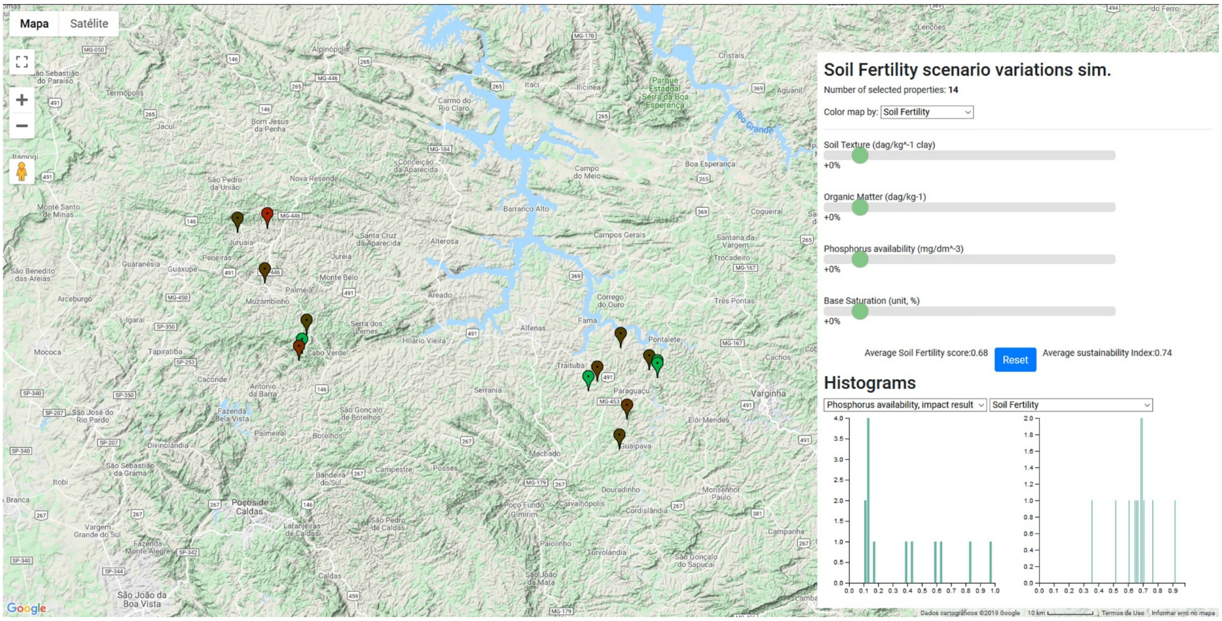
As the initial data analysis suggested, the rural properties present in the Balde Cheio dataset generally present poor Soil Fertility indicators and elevated Water Quality indicator performances. The simulations chosen and executed were related to those two indicators, one experiment aiming to increase the performance of fourteen properties on the Soil Fertility indicator and the other simulating an unfavorable scenario where there was a contamination of the waters used by a group of eight properties.

The 14 (fourteen) participating properties for the Soil Fertility simulations were handpicked and are from the municipalities of Cabo Verde, Muzambinho, Juruaia, Traituba, Paraguaçu, Guaipava and Pontalete, all in the south of the state of Minas Gerais, shown in Fig. 15. The systematic soil enrichment phenomenon simulated here can be translated to the real world as a collective effort from a cooperative or community to adopt soil enrichment practices, or a governmental policy to subsidize fertilizers, for example.

Their average Soil Fertility in the year of 2016 was 0.68 (std. dev. 0.15), as a group, they were only slightly below the desired 0.7, however, 10 out of 14 properties were below the desired threshold. We tested multiple soil enrichment scenarios and the most consistent variable with the best results was the measurement of the availability of phosphorus in the productive soil of the properties (mg/dm^3). Enriching lands with phosphorus is a common practice, many fertilizer solutions contain phosphorus in the formula.

By enriching those properties' soils with phosphorus by a proportion of +13% mg/dm^3 each, the selected set of properties reached the desired 0.7 (std. dev. 0.14) score for the Soil Fertility, with half of them below the threshold and the other half above or equal. Increasing the same variable by +66%, we have an average of 0.73 (std. dev. 0.12) for Soil Fertility, with 4 out of 14 below the threshold. With +78% of phosphorus enrichment, the properties present an average of 0.74 (std. dev. 0.12) for Soil Fertility and only 3 out of 14 below the desired threshold. Fig. 16 shows the properties simulated for an increase of +80%, it doesn't change significantly from the result obtained by increasing +78%.

The second simulation is a scenario of water contamination by thermotolerant coliforms, measured in CFU/100 mL (Colony Forming Units in 100 mL of sampled water). It was also taken into consideration if such infection would affect the turbidity



Soil Fertility scenario variations sim.

Number of selected properties: **14**

Color map by: Soil Fertility

Soil Texture (dag/kg^-1 clay) +0%

Organic Matter (dag/kg-1) +0%

Phosphorus availability (mg/dm^-3) +0%

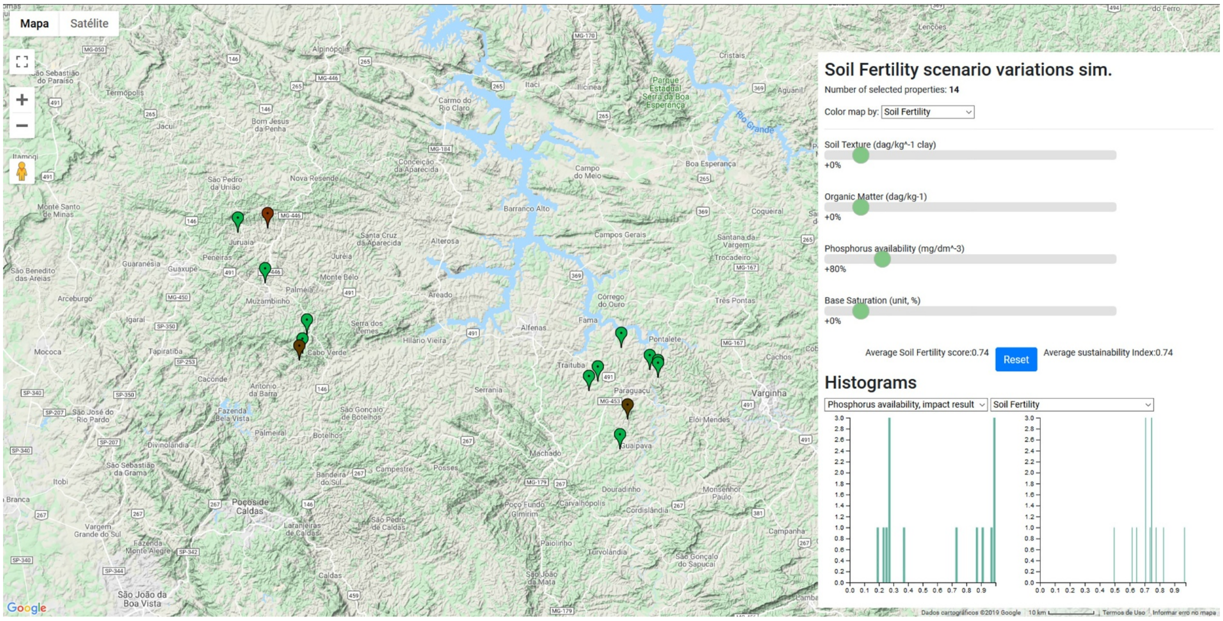
Base Saturation (unit, %) +0%

Average Soil Fertility score:0.68 Reset Average sustainability Index:0.74

Histograms

Phosphorus availability, impact result | Soil Fertility

Fig. 15. Agro 4.0 : Screenshot of the selected properties for the Soil Fertility experiment, without any modification on the original values.



Soil Fertility scenario variations sim.

Number of selected properties: **14**

Color map by: Soil Fertility

Soil Texture (dag/kg^-1 clay)
+0%

Organic Matter (dag/kg-1)
+0%

Phosphorus availability (mg/dm^-3)
+80%

Base Saturation (unit, %)
+0%

Average Soil Fertility score:0.74 Reset Average sustainability Index:0.74

Histograms

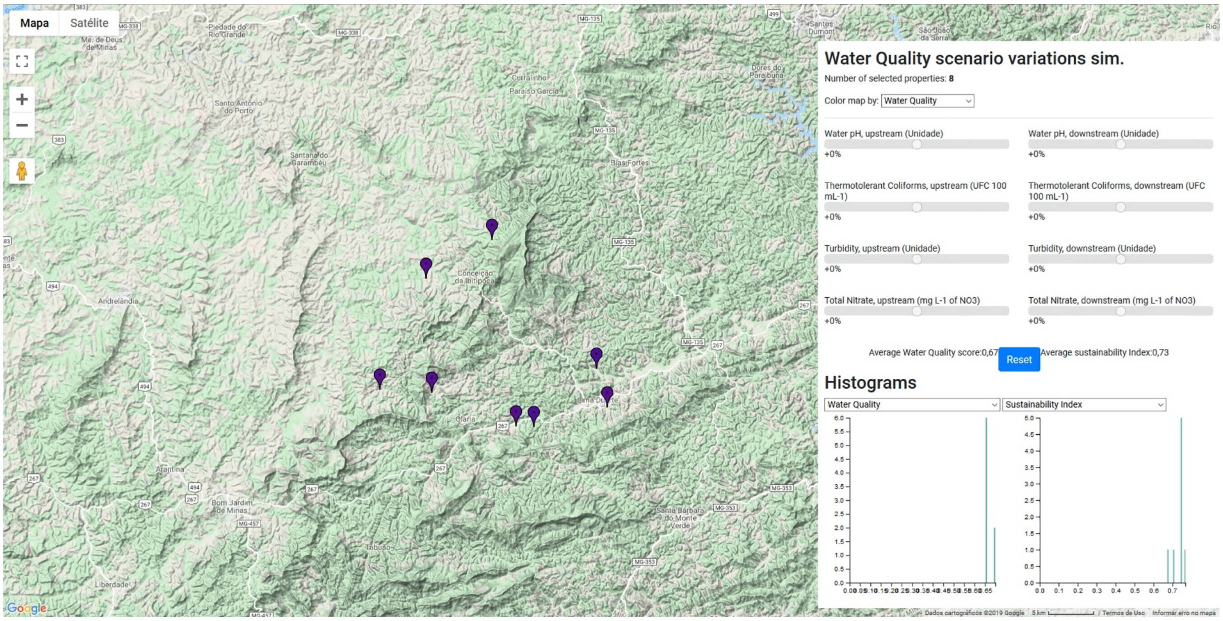
Phosphorus availability, impact result | Soil Fertility

Fig. 16. Agro 4.0 : Results of the Soil Fertility experiment after enriching the availability of phosphorus (mg/dm^3) on the soils of each property by 80%.

(cloudiness, measured in NTU - Nephelometric Turbidity Unity) of the water or not because that observation depends on conditions of the infection (for instance, sewer contamination, the water turbidity would also probably change).

Figs. 17 and 18 show the selected 8 properties before and after the simulation of the scenario, respectively. The agroecosystems are located in the municipalities of Conceição da Ibitipoca, Olaria and Lima Duarte, also in the south of Minas Gerais and to the east of the municipalities presented in the first scenarios simulation experiment.

The selected 8 (eight) properties presented an average Water Quality Indicator of 0.67 (std. dev. 0.02), 2 out of 8 of them above the desired 0.7 threshold. Increasing the thermotolerant coliforms values by +3% for each property, the average Water Quality of the group decreased to 0.53 (std. dev. 0.09). Increasing the turbidity of the water alone by +3% applied to each item, the average Water Quality of the group went to 0.49 (std. dev. 0.12). Increasing both the thermotolerant coliforms and turbidity values by +3% in each property's water samples, the average Water Quality of the group decreased to 0.36 (std. dev. 0.20). Contamination by coliforms and



Water Quality scenario variations sim.

Number of selected properties: **8**

Color map by: Water Quality

Water pH, upstream (Unidade)
+0%

Water pH, downstream (Unidade)
+0%

Thermotolerant Coliforms, upstream (UFC 100 mL-1)
+0%

Thermotolerant Coliforms, downstream (UFC 100 mL-1)
+0%

Turbidity, upstream (Unidade)
+0%

Turbidity, downstream (Unidade)
+0%

Total Nitrate, upstream (mg L-1 of NO3)
+0%

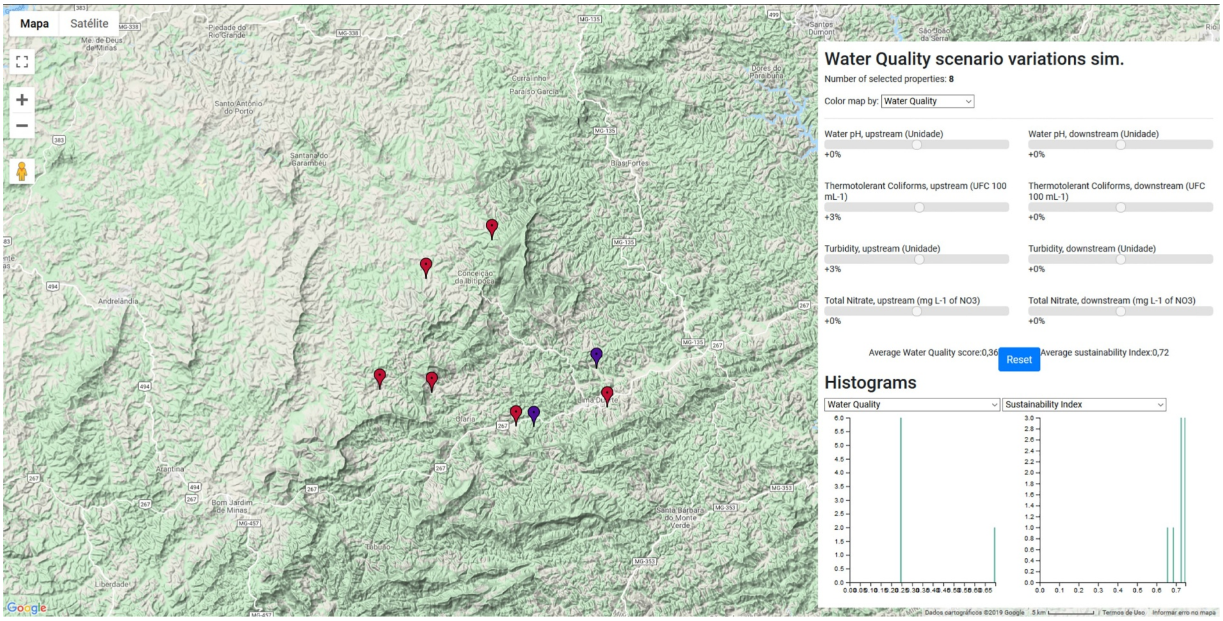
Total Nitrate, downstream (mg L-1 of NO3)
+0%

Average Water Quality score:0,67 Reset Average sustainability Index:0,73

Histograms

Water Quality | Sustainability Index

Fig. 17. Agro 4.0 : Screenshot of the selected properties for the Water Quality experiment, without any modification on the original values.



Water Quality scenario variations sim.

Number of selected properties: **8**

Color map by: Water Quality

Water pH, upstream (Unidade) +0%

Water pH, downstream (Unidade) +0%

Thermotolerant Coliforms, upstream (UFC 100 mL-1) +3%

Thermotolerant Coliforms, downstream (UFC 100 mL-1) +0%

Turbidity, upstream (Unidade) +3%

Turbidity, downstream (Unidade) +0%

Total Nitrate, upstream (mg L-1 of NO3) +0%

Total Nitrate, downstream (mg L-1 of NO3) +0%

Average Water Quality score:0,36 Average sustainability Index:0,72

Fig. 18. Agro 4.0 : Results of the Water Quality experiment after increasing both the quantity of colony forming units in 100 mL of water and the water turbidity for each property in 3%.

impure liquids quickly manifests itself in the Water Quality indicator.

For both experiments, it was noted that the changes in the overall Sustainability Index were insignificant, since it’s an average between all the 21 indicators and only one indicator in each experiment would have its related variables changed. The worst-case scenario of the Water Quality experiment caused the Sustainability Index for that group of properties to be decreased only by 0.01. It could be argued that the scenarios simulated here would also affect and imply in changes in other variables measured in other indicators.

7. Conclusion

Public awareness of the negative impacts of human activity on our environment is at an all times high. Technological efforts to increase the sustainability of productive Agroecosystems are being studied, developed and applied in many different places. In this work, we adopted a Brazilian methodology called Indicators of Sustainability in Agroecosystems (*Indicadores de Sustentabilidade em Agroecossistemas* – ISA), implemented an information system based on it and apply Data Science techniques over the gathered data - from 100 real rural properties - to compute which are the most relevant ISA Indicators for the final ISA Sustainability Index Score and ran environmental changes simulations on two targeted locations while measuring the impact of such changes on their experiment-related Indicators.

Initially, the ISA methodology for the calculation of sustainability in agroecosystems was presented. Based on this reference methodology, Agro 4.0 was developed. This new tool makes it possible for the methodology to be applied in a greater scale, allowing for quicker evaluation of participating properties, regarding their agroecosystems’ sustainability.

Agro 4.0 facilitates the work of professionals applying the ISA Methodology on rural areas, it reduces the time needed to collect information, makes it more difficult for human errors to happen and generates out of the box useful reports and dashboards with aggregated queries and on demand summarizations for sub groups of agroecosystems or regions. A simulation module was also developed and it allows for environmental changes scenarios and their impacts to be evaluated using the ISA Methodology.

With the use of data mining techniques, it was possible to identify that using only 7 out of the 21 indicators - originally required by the ISA Model - it is possible to identify with 94% precision the level of sustainability of a rural property. For this, the Random Forest classification technique was used. The indicators used are: *4. Degree of indebtedness, 8. Business management, 9. Information management, Soil fertility, Conservation practices, 19. APPs* (Adequacy of Permanent Preservation Areas) and *20. Legal Reserve (Legal Reserve)*. These indicators were selected using the CFS feature selection technique.

Some of the future work involves integration with new databases. Data can be collected from IBGE, such as data sets that can be useful to identify new variables with relevant correlations regarding sustainability. The National Water Agency (ANA) has updated maps of the basins and sub-basins of the country. This information can be cross-checked and be used to try to predict possible water scarcity for rural properties in some geographical region, for example. The integration of databases with railways, waterways and highways can help identify new production outflows routes or facilitate the creation of producers networks. Another future work possible, because demand was identified, is the development of a mobile application for georeferenced photographic registration of problems or solutions found in a property.

Appendix A. Soil fertility and water quality indicators computation

This appendix contains the procedures to calculate the Soil Fertility and Water Quality Indicators. The idea behind the procedures are explained in Section 2.4. The procedures come from the reference ISA Methodology spreadsheet [25] made and used by FAEMG since 2014.

The procedure to compute the Soil Fertility Indicator and related partial results for a rural property, is as follow, as specified in the spreadsheet [25] and discussed in the main ISA Methodology publications [9,24,26,27]:

$$limit(x) = \begin{cases} 1, \text{ if } x > 1 \\ 0, \text{ if } x < 0 \\ x \end{cases} \tag{A.1}$$

$$PROM(m, t) = \begin{cases} 0.100 + 0.3597 \cdot m + 0.1926 \cdot m^2 - 0.0216 \cdot m^3, \\ \text{if } \ t \leq 15 \ \text{ and } \ m < 1.5 \\ 1.15 - 0.10 \cdot m, \\ \text{if } \ t \leq 15 \ \text{ and } \ m \geq 1.5 \\ 0.100 + 0.0024 \cdot m + 0.2183 \cdot m^2 - 0.0397 \cdot m^3, \\ \text{if } \ 16 \leq t \leq 34 \ \text{ and } \ m < 3 \\ 1.15 - 0.05 \cdot m, \\ \text{if } \ 16 \leq t \leq 34 \ \text{ and } \ m \geq 3 \\ 0.100 - 0.0452 \cdot m + 0.1269 \cdot m^2 - 0.0161 \cdot m^3, \\ \text{if } \ 35 \leq t \leq 60 \ \text{ and } \ m < 4.5 \\ 1.15 - 0.0333 \cdot m, \\ \text{if } \ 35 \leq t \leq 60 \ \text{ and } \ m \geq 4.5 \\ 0.100 - 0.0486 \cdot m + 0.0984 \cdot m^2 - 0.0107 \cdot m^3, \\ \text{if } \ t > 60 \ \text{ and } \ m < 5.2 \\ 1.15 - 0.0288 \cdot m, \\ \text{if } \ t > 60 \ \text{ and } \ m \geq 5.2 \end{cases} \tag{A.2}$$

$$PRB(b) = \begin{cases} 0.3 + 0.0067 \cdot b, \\ \text{if } \ b \geq 30 \\ 0.1 + 0.0067 \cdot b + 0.0002 \cdot b^2, \\ \text{if } \ b < 30 \end{cases} \tag{A.3}$$

$$PRP(p, t) = \begin{cases} 0.10 + 0.0124 \cdot p, \\ \text{if } \ t \leq 15 \ \text{ and } \ p \leq 12.1 \\ -0.4748 + 0.0721 \cdot p - 0.001 \cdot p^2, \\ \text{if } \ t \leq 15 \ \text{ and } \ 12.1 < p < 25.1 \\ 0.1946 + 0.0201 \cdot p, \\ \text{if } \ t \leq 15 \ \text{ and } \ 25.1 \leq p \leq 40 \\ 1.0233 - 0.0008 \cdot p, \\ \text{if } \ t \leq 15 \ \text{ and } \ p > 40 \\ 0.1 - 0.0087 \cdot p + 0.0023 \cdot p^2, \\ \text{if } \ 16 \leq t \leq 34 \ \text{ and } \ p \leq 15.1 \\ -0.3015 + 0.0639 \cdot p - 0.0007 \cdot p^2, \\ \text{if } \ 16 \leq t \leq 34 \ \text{ and } \ 15.1 < p < 30 \\ 1.0333 - 0.0011 \cdot p, \\ \text{if } \ 16 \leq t \leq 34 \ \text{ and } \ p \geq 30 \\ 0.1 - 0.0045 \cdot p + 0.0067 \cdot p^2, \\ \text{if } \ 35 \leq t \leq 60 \ \text{ and } \ p \leq 8.1 \\ 0.1034 + 0.0483 \cdot p + 0.0000856 \cdot p^2, \\ \text{if } \ 35 \leq t \leq 60 \ \text{ and } \ 8.1 < p < 18 \\ 1.04333 - 0.0019 \cdot p, \\ \text{if } \ 35 \leq t \leq 60 \ \text{ and } \ p \geq 18 \\ 0.1 - 0.1041 \cdot p + 0.0492 \cdot p^2, \\ \text{if } \ t > 60 \ \text{ and } \ p \leq 4.1 \\ 0.1076 + 0.0928 \cdot p + 0.0007 \cdot p^2, \\ \text{if } \ t > 60 \ \text{ and } \ 4.1 < p < 9 \\ 1.0333 - 0.0037 \cdot p, \\ \text{if } \ t > 60 \ \text{ and } \ p > 9 \end{cases} \tag{A.4}$$

$$soilFertility(b, m, p, t) = \frac{limit(PROM(m, t)) + limit(PRP(p, t)) + limit(PRB(b))}{3} \tag{A.5}$$

The functions for the Soil Fertility Indicators are based on regression analyses made using the software SigmaPlot with data from a Soil Analysis document by EMBRAPA (2003) [65] and by the Commission of Soil Feritlity of the State Of Minas Gerais. The input variable b represents the amount of base saturation (%) found on soil samples from the rural property by tests conducted on laboratory, m represents the amount of Organic Matter (dag/kg) on those samples, p represents the amount of phosphorus ((mg dm^3)) and t is the amount of clay (dag/kg) found on the examined samples.

The equations with the suffix "Reg" bellow are based on the documents Resolution CONAMA 357/05 [66] of CONAMA and Ordinance 518/04 [67] of the brazilian Health Ministry. The Water Quality Indicator for surface water in a rural property is computed as follow [9,24,26,27]:

$$phReg(ph) = \begin{cases} 0.7, \text{ if } 6 \le ph \le 9 \\ 0.3, \text{ if } ph < 6 \text{ or } ph > 9 \end{cases} \tag{A.6}$$

$$turbidityReg(t) = \begin{cases} 0.7, \text{ if } t \le 100 \\ 0.1, \text{ if } t > 100 \end{cases} \tag{A.7}$$

$$coliformReg(c) = \begin{cases} 0.7, \text{ if } c \le 999 \\ 0.1, \text{ if } c > 999 \end{cases} \tag{A.8}$$

$$nitrateReg(n) = \begin{cases} 0.7, \text{ if } n \le 9.99 \\ 0.1, \text{ if } n > 9.99 \end{cases} \tag{A.9}$$

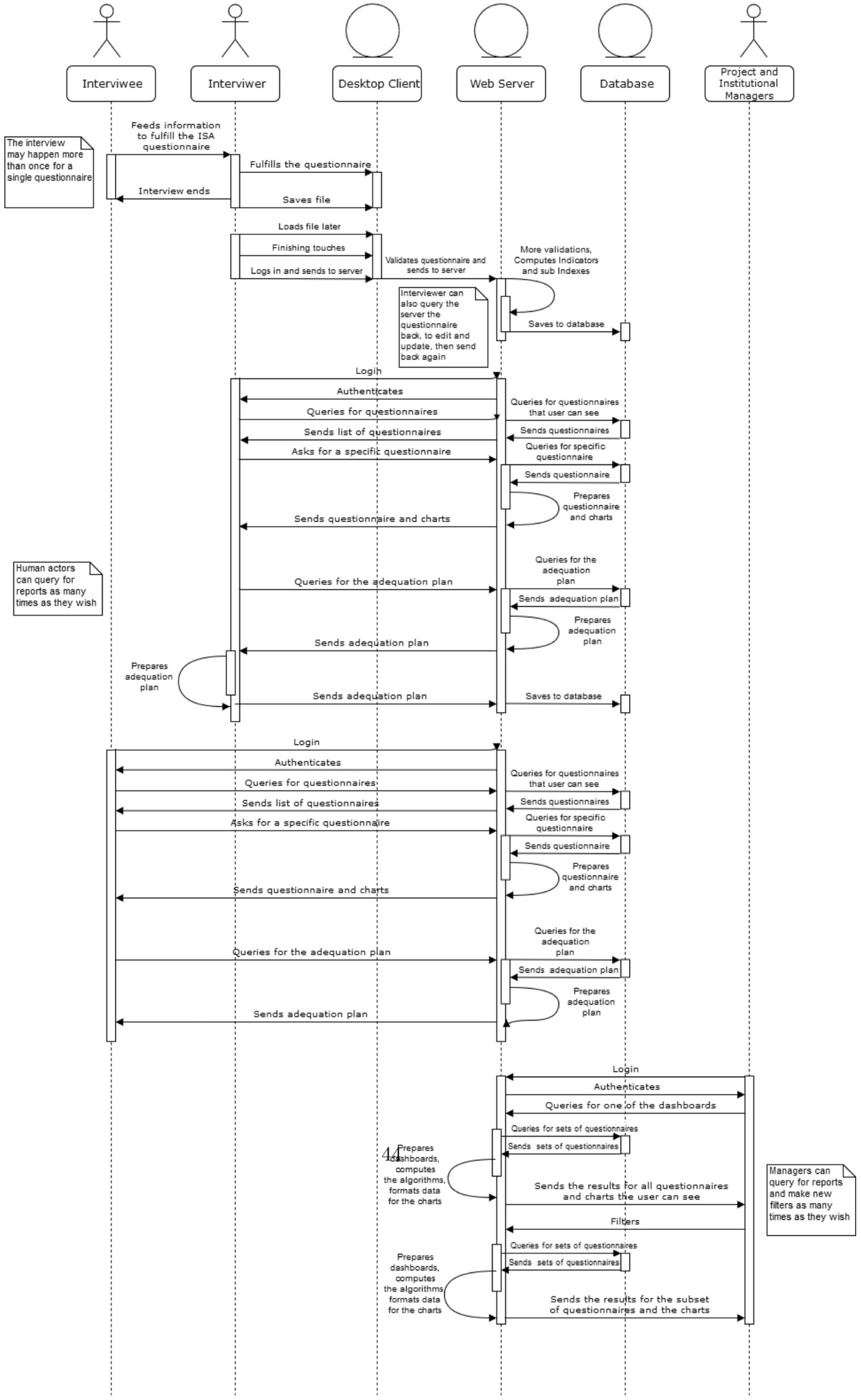
$$upstream(pH, t, c, n, s) = \frac{\begin{array}{c} phReg(pH) + turbidityReg(t) \cdot 4 \\ + coliformReg(c) \cdot 3 + nitrateReg(n) \cdot 2 + s \cdot 4 \end{array}}{14} \tag{A.10}$$

$$downstream(pH, t, c, n, s) = \frac{\begin{array}{c} phReg(pH) + turbidityReg(t) \cdot 4 \\ + coliformReg(c) \cdot 3 + nitrateReg(n) \cdot 2 + s \cdot 4 \end{array}}{14} \tag{A.11}$$

$$waterQuality = min(upstream, downstream) \tag{A.12}$$

There is sampling predicted for downstream and upstream water bodies in laboratories in the ISA Methodology. The pH variables refer to the pH levels of the water samples, t refers to the samples' turbidity levels (measured in NTU – Nephelometric Turbidity Unity), c represents the amount of thermotolerant coliforms found on the samples, measured in CFU/100 mL (Colony Forming Units in 100 mL of sampled water), n is the amount of nitrate (NO_3) in those samples, mg/L. The s variables are the scores of the waters of the rural property obtained by visual and odor testing by a technician using the reference spreadsheet.

Appendix B. Agro 4.0 main user interaction processes


Interviwee
Interviwer
Desktop Client
Web Server
Database
Project and Institutional Managers
The interview may happen more than once for a single questionnaire
Feeds information to fulfill the ISA questionnaire
Fulfills the questionnaire
Interview ends
Saves file
Loads file later
Finishing touches
Logs in and sends to server
Validates questionnaire and sends to server
More validations, Computes Indicators and sub Indexes
Interviewer can also query the server the questionnaire back, to edit and update, then send back again
Saves to database
Login
Authenticates
Queries for questionnaires
Queries for questionnaires that user can see
Sends questionnaires
Sends list of questionnaires
Asks for a specific questionnaire
Queries for specific questionnaire
Sends questionnaire
Prepares questionnaire and charts
Sends questionnaire and charts
Human actors can query for reports as many times as they wish
Queries for the adequation plan
Queries for the adequation plan
Sends adequation plan
Prepares adequation plan
Sends adequation plan
Prepares adequation plan
Sends adequation plan
Saves to database
Login
Authenticates
Queries for questionnaires
Queries for questionnaires that user can see
Sends questionnaires
Sends list of questionnaires
Asks for a specific questionnaire
Queries for specific questionnaire
Sends questionnaire
Prepares questionnaire and charts
Sends questionnaire and charts
Queries for the adequation plan
Queries for the adequation plan
Sends adequation plan
Prepares adequation plan
Sends adequation plan
Login
Authenticates
Queries for one of the dashboards
Queries for sets of questionnaires
Sends sets of questionnaires
44
Prepares dashboards, computes the algorithms, formats data for the charts
Sends the results for all questionnaires and charts the user can see
Managers can query for reports and make new filters as many times as they wish
Filters
Queries for sets of questionnaires
Sends sets of questionnaires
Prepares dashboards, computes the algorithms, formats data for the charts
Sends the results for the subset of questionnaires and the charts

References

[1] B.R. Keeble, The brundtland report: ‘our common future’, Med. War 4 (1) (1988) 17–25.
[2] J. Bebbington, R. Gray, An account of sustainability: failure, success and a reconceptualization, Crit. Perspect. Account. 12 (5) (2001) 557–587.
[3] M. van Marrewijk, Concepts and definitions of CSR and corporate sustainability: between agency and communion, J. Bus. Ethics 44 (2) (2003) 95–105, https://doi.org/10.1023/A:1023331212247.
[4] P. Glavič, R. Lukman, Review of sustainability terms and their definitions, J. Cleaner Prod. 15 (18) (2007) 1875–1885.
[5] N. Dempsey, G. Bramley, S. Power, C. Brown, The social dimension of sustainable development: defining urban social sustainability, Sustain. Dev. 19 (5) (2011) 289–300.
[6] J.R. Lamontagne, P.M. Reed, G. Marangoni, K. Keller, G.G. Garner, Robust abatement pathways to tolerable climate futures require immediate global action, Nat. Clim. Change (2019), https://doi.org/10.1038/s41558-019-0426-8.
[7] D. Tilman, K.G. Cassman, P.A. Matson, R. Naylor, S. Polasky, Agricultural sustainability and intensive production practices, Nature 418 (6898) (2002) 671–677, https://doi.org/10.1038/nature01014.
[8] E. Fonseca, E. Caldeira, L. Oliveira, A.C.M. Pereira, P.S. Vilela, Agro 4.0: uma ferramenta web para gestão e análise da sustentabilidade em agroecossistemas, Anais Do XXIII Simpósio Brasileiro de Sistemas Multimídia E Web: Workshops E Pôsteres, (2017), pp. 184–188.
[9] K. Marzall, J. Almeida, Indicadores de sustentabilidade para agroecossistemas estado da arte, limites e potencialidades de uma nova ferramenta para avaliar o desenvolvimento sustentável, Cad. Ciênc. Tecnol. 17 (1) (2000) 41–59.
[10] E. Commission, Emas - factsheet, 2008, URL http://www.emas.de/fileadmin/user_upload/04_ueberemas/PDF-Dateien/Unterschiede_iso_en.pdf.
[11] E. Commission, Emas, a premium environmental management toolfor organisations, 2018, URL https://ec.europa.eu/environment/emas/pdf/other/EMAS%20presentation%20for%20organisations_2018.pdf.
[12] E. Commission, Emas and biodiversity, 2016, URL https://ec.europa.eu/environment/emas/pdf/other/EMAS_Biodiversity_Guidelines_2016.pdf.
[13] D. Freebairn, C. King, Reflections on collectively working toward sustainability: indicators for indicators!, Anim. Prod. Sci. 43 (3) (2003) 223–238.
[14] T. Rogmans, M. Ghunaim, A framework for evaluating sustainability indicators in the real estate industry, Ecol. Indic. 66 (2016) 603–611, https://doi.org/10.1016/j.ecolind.2016.01.058.
[15] I. Coteur, F. Marchand, L. Debruyne, F. Dalemans, L. Lauwers, A framework for guiding sustainability assessment and on-farm strategic decision making, Environ. Impact Assess. Rev. 60 (2016) 16–23, https://doi.org/10.1016/j.eiar.2016.04.003.
[16] R.K. Singh, H. Murty, S. Gupta, A. Dikshit, An overview of sustainability assessment methodologies, Ecol. Indic. 9 (2) (2009) 189–212, https://doi.org/10.1016/j.ecolind.2008.05.011.
[17] J. Dedrick, Green is: concepts and issues for information systems research, Commun. Assoc. Inf. Syst. 27 (2010), https://doi.org/10.17705/1CAIS.02711.
[18] Z. Liu, H. Wang, P. Li, The antecedents of green information system and impact on environmental performance, SSRN (2018), https://doi.org/10.2139/ssrn.3177907.
[19] R. Gholami, A.B. Sulaiman, T. Ramayah, A. Molla, Senior managers’ perception on green information systems (is) adoption and environmental performance: results from a field survey, Inf. Manag. 50 (7) (2013) 431–438, https://doi.org/10.1016/j.im.2013.01.004.
[20] T.A. Jenkin, J. Webster, L. McShane, An agenda for ‘green’ information technology and systems research, Inf. Organ. 21 (1) (2011) 17–40, https://doi.org/10.1016/j.infoandorg.2010.09.003.
[21] A.J Chen, M.-C. Boudreau, R.T. Watson, Information systems and ecological sustainability, J. Syst. Inf. Technol. 10 (3) (2008) 186–201.
[22] S. Rodrigues Filho, A.J. Juliani, Sustentabilidade da produção de etanol de cana-de-açúcar no estado de são paulo, Estudos Avançados 27 (78) (2013) 195–212.
[23] W.B. Hempel, A importância do icms ecológico para a sustentabilidade ambiental no ceará, REDE Rev. Eletrôn. PRODEMA 2 (2) (2009).
[24] J.M.L. Ferreira, J.H.M. Viana, A.M. da Costa, D.V. de Sousa, A.A. Fontes, Indicadores de sustentabilidade em agroecossistemas, Inf. Agropecu. Belo Horiz. 33 (271) (2012) 12–25.
[25] ISA, FAEMG, Isa 2014 complete excel spreadsheet, reference and application, 2014, URL https://agro.sybers.dcc.ufmg.br/promotion/assets/ISA%202014.xlsx.
[26] C.C. Geraldo Stachetti Rodrigues, Sistema integrado de avaliação de impacto ambiental aplicado a atividades do novo rural, Pesquisa Agropecu. Bras. 38 (4) (2003) 445–451.
[27] A.M. da Costa, J.M.L. Ferreira, J.H.M. Viana, A.R. de Oliveira, Indicadores de sustentabilidade em agroecossistemas (ISA), XXXIV Congresso Brasileiro de Ciência do Solo, (2013), pp. 1–4.
[28] E. Caldeira, E. Fonseca, L.B. Oliveira, A.C.M. Pereira, M. Vilela, P.S. Vilela, Is@ digital: uma ferramenta para gestão de sustentabilidade em agroecossistemas, Anais do XXII Simpósio Brasileiro de Sistemas Multimídia e Web (Vol. 2): Workshops e Sessão de Pôsteres, (2016), pp. 130–131.
[29] A. Jović, K. Brkić, N. Bogunović, A review of feature selection methods with applications, 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), (2015), pp. 1200–1205, https://doi.org/10.1109/MIPRO.2015.7160458.
[30] M.A. Hall, Correlation-based Feature Subset Selection for Machine Learning, University of Waikato, Hamilton, New Zealand, 1998 Ph.D. thesis.
[31] L.M. Joyce, M. James, Kullback-Leibler Divergence, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 720–722. 10.1007/978-3-642-04898-2_327.
[32] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.
[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. 11 (2009) 10–18, https://doi.org/10.1145/1656274.1656278.
[34] A. Kamilaris, A. Kartakoullis, F.X. Prenafeta-Bold, A review on the practice of big data analysis in agriculture, Comput. Electron. Agric. 143 (2017) 23–37, https://doi.org/10.1016/j.compag.2017.09.037.
[35] M. Chi, A. Plaza, J.A. Benediktsson, Z. Sun, J. Shen, Y. Zhu, Big data for remote sensing: Challenges and opportunities, Proc. IEEE 104 (11) (2016) 2207–2219, https://doi.org/10.1109/JPROC.2016.2598228.
[36] K.E. Giller, P. Tittonell, M.C. Rufino, M.T. Wijk, S. Zingore, P. Mapfumo, S. Adjei-Nsiah, M. Herrero, R. Chikowo, M. Corbeels, E.C. Rowe, F. Baijukya, A. Mwijage, J. Smith, E. Yeboah, W.J. Burg, O.M. Sanogo, M. Misiko, N. de Ridder, S. Karanja, C. Kaizzi, J. K’ungu, M. Mwale, D. Nwaga, C. Pacini, B. Vanlauwe, Communicating complexity: integrated assessment of trade-offs concerning soil fertility management within african farming systems to support innovation and development, Agric. Syst. 104 (2) (2011) 191–203, https://doi.org/10.1016/j.agsy.2010.07.002.
[37] M.S. Xuan Pham, How data analytics is transforming agriculture, Bus. Horiz. 61 (1) (2018) 125–133, https://doi.org/10.1016/j.bushor.2017.09.011.
[38] M. Muntean, Business intelligence issues for sustainability projects, Sustainability 10 (2018) 335, https://doi.org/10.3390/su10020335.
[39] A.J. Chen, R. Watson, M.-c. Boudreau, E. Karahanna, Organizational adoption of green is and it: an institutional perspective, ICIS 2009 Proceedings - Thirtieth International Conference on Information Systems, (2009), p. 142.
[40] A. Malhotra, N. Melville, R.T. Watson, Spurring impactful research on information systems for environmental sustainability, Manag. Inf. Syst. Q. 37 (4) (2013) 1265–1274.
[41] S.A. Nikolidakis, D. Kandris, D.D. Vergados, C. Douligeris, Energy efficient automated control of irrigation in agriculture by using wireless sensor networks, Comput. Electron. Agric. 113 (2015) 154–163.
[42] E.M. De Olde, F.W. Oudshoorn, C.A. Sørensen, E.A. Bokkers, I.J. De Boer, Assessing sustainability at farm-level: Lessons learned from a comparison of tools in practice, Ecol. Indic. 66 (2016) 391–404.
[43] F. Häni, F. Braga, A. Stämpfli, T. Keller, M. Fischer, H. Porsche, et al., Rise, a tool for holistic sustainability assessment at the farm level, Int. Food Agribus. Manag. Rev. 6 (4) (2003) 78–90.
[44] G.C. Pacini, G. Lazzerini, C. Vazzana, Aesis: a support tool for the evaluation of sustainability of agroecosystems. example of applications to organic and integrated farming systems in tuscany, Italy, Ital. J. Agron. 6 (1) (2011) 3.
[45] N. Van Cauwenbergh, K. Biala, C. Bielders, V. Brouckaert, L. Franchois, V.G. Cidad, M. Hermy, E. Mathijs, B. Muys, J. Reijnders, et al., Safe-a hierarchical framework for assessing the sustainability of agricultural systems, Agric. Ecosyst. Environ. 120 (2-4) (2007) 229–242.
[46] M. Kropff, J. Bouma, J. Jones, Systems approaches for the design of sustainable agro-ecosystems, Agric. Syst. 70 (2) (2001) 369–393.

[47] R. Shaw, R. Lark, A. Williams, D. Chadwick, D. Jones, Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture, Agric. Ecosyst. Environ. 230 (2016) 294–306.

[48] C.C. de Resende, A.C. Pereira, R.T. Cardoso, A.B. de Magalhães, Investigating market efficiency through a forecasting model based on differential equations, Phys. A 474 (2017) 199–212.

[49] S. Mittal, M. Mehar, Socio-economic factors affecting adoption of modern information and communication technology by farmers in India: analysis using multivariate Probit model, J. Agric. Educ. Ext. 22 (2) (2016) 199–212.

[50] I. Thysen, Agriculture in the information society, J. Agric. Eng. Res. 76 (3) (2000) 297–303.

[51] R. Nikkilä, I. Seilonen, K. Koskinen, Software architecture for farm management information systems in precision agriculture, Comput. Electron. Agric. 70 (2) (2010) 328–336.

[52] R. Bongiovanni, J. Lowenberg-DeBoer, Precision agriculture and sustainability, Precis. Agric. 5 (4) (2004) 359–387.

[53] S. Fountas, G. Carli, C.G. Sørensen, Z. Tsiropoulos, C. Cavalaris, A. Vatsanidou, B. Liakos, M. Canavari, J. Wiebensohn, B. Tisserye, Farm management information systems: Current situation and future perspectives, Comput. Electron. Agric. 115 (2015) 40–50.

[54] A.Z. Abbasi, N. Islam, Z.A. Shaikh, et al., A review of wireless sensors and networks' applications in agriculture, Comput. Stand. Interfaces 36 (2) (2014) 263–270.

[55] Y. Kim, R.G. Evans, W.M. Iversen, Remote sensing and control of an irrigation system using a distributed wireless sensor network, IEEE Trans. Instrum. Meas. 57 (7) (2008) 1379–1387.

[56] B.A. Aubert, A. Schroeder, J. Grimaudo, It as enabler of sustainable farming: an empirical analysis of farmers' adoption decision of precision agriculture technology, Decis. Support Syst. 54 (1) (2012) 510–520.

[57] X. Wang, Z. Du, Y. Chen, M. Yang, A green-aware virtual machine migration strategy for sustainable datacenter powered by renewable energy, Simul. Model. Pract. Theory (2015), https://doi.org/10.1016/j.simpat.2015.01.005.

[58] G.E. Dumitran, L.I. Vuta, Study on lake izvorul muntelui rehabilitation, Simul. Model. Pract. Theory (2010), https://doi.org/10.1016/j.simpat.2010.05.008.

[59] N. Bessis, S. Sotiriadis, F. Pop, V. Cristea, Using a novel message-exchanging optimization (MEO) model to reduce energy consumption in distributed systems, Simul. Model. Pract. Theory (2013), https://doi.org/10.1016/j.simpat.2013.02.003.

[60] B.S. Onggo, J. Panadero, C.G. Corlu, A.A. Juan, Agri-food supply chains with stochastic demands: a multi-period inventory routing problem with perishable products, Simul. Model. Pract. Theory (2019), https://doi.org/10.1016/j.simpat.2019.101970.

[61] M. Huang, Z. Liu, Y. Tao, Mechanical fault diagnosis and prediction in iot based on multi-source sensing data fusion, Simul. Model. Pract. Theory (2019) 101981, https://doi.org/10.1016/j.simpat.2019.101981.

[62] W.W. Cohen, Fast effective rule induction, Machine Learning Proceedings 1995, Elsevier, 1995, pp. 115–123.

[63] A.M. Novo, M. Slingerland, K. Jansen, A. Kanellopoulos, K.E. Giller, Feasibility and competitiveness of intensive smallholder dairy farming in brazil in comparison with soya and sugarcane: Case study of the balde cheio programme, Agric. Syst. 121 (2013) 63–72.

[64] M.S. Borges, C.A.M. Guedes, R.L. de Assis, Um estudo do "projeto balde cheio" como vetor de desenvolvimento sustentável do pequeno produtor de leite, Revista Brasileira de Agropecuária Sustentável (RBAS) 1 (1) (2011) 151–161.

[65] M. da Agricultura Pecuária e Abastecimento Brasil, Interpretação de resultadosde análise de solo, 2003. URL https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/487554/1/Com82.pdf.

[66] C. Brasil, Resolução conama no 357, de 17 de março de 2005, 2005, URL http://www.icmbio.gov.br/cepsul/images/stories/legislacao/Resolucao/2005/res_conama_357_2005_classificacao_corpos_agua_rtfcda_altrd_res_393_2007_397_2008_410_2009_430_2011.pdf.

[67] M. da Saúde Brasil, Portaria ms no 518/2004, 2004, URL http://bvsms.saude.gov.br/bvs/publicacoes/portaria_518_2004.pdf.