Computers and Electronics in Agriculture 175 (2020) 105602

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

A computational model for soil fertility prediction in ubiquitous agriculture

Gilson Augusto Helfer[a,*], Jorge Luis Victória Barbosa[a], Ronaldo dos Santos[b], Adilson Ben da Costa[b]

[a] Universidade do Vale do Rio dos Sinos, Av.Unisinos, 950, São Leopoldo, RS, Brazil
[b] Universidade de Santa Cruz do Sul, Av.Independência, 2255, Santa Cruz do Sul, RS, Brazil

ARTICLE INFO

Keywords:
Ubiquitous computing
Context awareness
Soil fertility
Productivity
Precision agriculture

2010 MSC:
00-01
99-00

ABSTRACT

The application of sophisticated sensors to measure soil composition and plant needs are a tendence in precision agriculture. In any case, prediction models are built using machine learning algorithms. The goal is to make farming more efficient and productive with minimal impact on the environment. The present article proposes an architectural model that evaluates the soil's fertility and productivity through context history with Partial Least Squares Regression. Also productivity prediction of a wheat planted area was performed using climatic events between the years of 2001 and 2015 resulting a mean square error of calibration (RMSEC) of 0.20 T/ha, mean square errors of cross-validation of 0.54 T/ha with a Pearson coefficient (R^2) of 0.9189. For the prediction of organic matter and clay, the best results obtained were a R^2 of 0.9345, RMSECV of 0.54% and R^2 of 0.9239, RMSECV of 5.28%, respectively.

1. Introduction

The main goal of precision agriculture is to increase productivity and allow the rational use of inputs, thus reducing the environmental impacts caused by agricultural practices. Through it, the inputs can be used in a variable way, aiming to meet the specific demands of each location, which optimizes the production process. Therefore, it is necessary to characterize the variability of soil chemical and physical attributes through representative sampling (Costa et al., 2014).

Soils differ gradually in ion concentration, fluctuating in nutrient levels, pH and electrical conductivity. The chemical composition of the soil depends on its water content, sampled layer and targeted nutrient. Thus, one-time sampling does not reflect its variations in composition in different climatic seasons. During certain periods of time, the variation of the soil's composition becomes more evident, with higher levels of organic matter and texture compared to soils with low clay and carbon content, whereas in soils richer in clay texture, mineralized nutrient reserves are larger, increasing the soil's salt and ion content. Changes in pH depend on the characteristics of the studied soil, which is a key factor in the regulation of the present ions (Miranda et al., 2006).

Soil analysis is the simplest, most economical and efficient way to diagnose its fertility as well as the basis for recommending adequate amounts of correctives and fertilizers to increase crop productivity and, as a result, crop yield and profitability (Furtini Neto et al., 2001). Interlaboratory quality control programs such as the Official Network of Soil and Plant Tissue Analysis Laboratories of the States of Rio Grande do Sul and Santa Catarina (ROLAS) ensure the monitoring of the soil's measurement error analysis and provide suggestions on analysis methods (Griebeler et al., 2016).

Several procedures used to analyze the soil are expensive, generate polluting residues and demand a long time of sample preparation. The analysis of organic matter and clay – which generally represent the soil fertility – requires 21 h to determine its values. Furthermore, they generate extremely environmentally harmful waste. In contrast, modern techniques have been studied in recent years to replace official techniques. These include near-infrared molecular spectroscopy (NIR) (Muñoz and Kravchenko, 2011).

Differently, other analysis such as pH, electrical conductivity, temperature and soil moisture have been performed in the field for a long time, due to the existence of portable equipment and sensors. Soil pH is a predictor of various chemical activities as well as a useful tool in making plant type management decisions appropriate for a region. The evaluations of electrical conductivity demonstrate the presence of ions (salts) that are available to the crop. A higher presence of ions means a greater conduction of electricity (do Carmo et al., 2016).

Strategies must be adopted to perform data prediction when there are many variables involved, such as the generation of analytical models through Machine Learning methods. These algorithms eliminate

* Corresponding author.
E-mail addresses: ghelfer@gmail.com, ghelfer@unisc.br (G.A. Helfer), jbarbosa@unisinos.br (J.L. Victória Barbosa), rbastos@mx2.unisc.br (R.d. Santos), adilson@unisc.br (A.B. da Costa).

https://doi.org/10.1016/j.compag.2020.105602
Received 24 February 2020; Received in revised form 19 June 2020; Accepted 22 June 2020
Available online 11 July 2020
0168-1699/

Table 1
Comparison of related works.

Criterion	Nawar and Mouazen, 2019	Huong et al., 2018	Treboux and Genoud, 2018	Goap et al., 2018	dos Santos et al., 2019	de la Concepcion et al., 2014	Proposed Model
1. Architecture	No	No	No	Yes	Yes	Yes	Yes
2. Notifications	No	No	No	No	Yes	No	Yes
3. Analysis	Machine Learning	Markov Decision	Machine Learning	Machine Learning	ARIMA	N/A	Machine Learning
4. Contextual Histories	Yes	Yes	Yes	Yes	Yes	No	Yes
5. Agents	No	Yes	No	Yes	Yes	No	Yes
6. Ontology	No	No	No	No	No	No	Yes
7. Sensors	Infrared (IR)	Humidity	Satellite Image	Temperature, Humidity, UV	Temperature, Humidity	Camera, Temperature, Humidity	IR, pH, EC, Lux, Rain, Temperature, Humidity, Pressure
8. Ubiquity Support	No	No	No	Yes	Yes	No	Yes
9. Web Access	No	No	Yes	Yes	Yes	No	Yes
10. Mobile App	No	No	No	Yes	Yes	No	Yes

variables that do not correlate to the property of interest, such as those that add noise, nonlinearities, or irrelevant information (Ferreira, 2015).

After the variables are selected, Machine Learning methods such as Partial Least Squares Regression can be applied. This technique was developed in the 1970's and primarily used in near infrared region (NIR). Currently, PLS regression is used in many applied sciences and it is the most successful multivariate calibration method in the application of the combination of chemometrics with spectral data (Ferrão and Davanzo, 2005; Valderrama et al., 2009).

In aditional, PLS are generally applied in situations where process variables have high levels of correlation, noise, missing observations and imbalance in the proportion of variables and observation. So, a small number of linear combinations, independent of process variables, are generated. These new variables, called components, account for most of the divergence that is presented in the process of dimensionality reduction. Thus,a few factors or leverages are retained to represent tens or even hundreds of processing variables (Anzanello, 2013). A detailed mathematical and statistical structure of PLS regression was described by Höskuldsson in 1988 and Geladi in 2003 (Höskuldsson, 1988; Geladi, 2003).

A computational architecture based on a framework and middleware is commonly applied in ubiquitous computing. Techniques are used for remote communication, fault tolerance, high availability, remote information access and security, among others. However, for this concept to come true, context aware applications that can, for example, adjust their own operation or even act proactively must be developed (Dey et al., 2001).

In the ubiquitous computing environment, raw data may be structured, semi-structured and unstructured, for it is collected by heterogeneous sensor sources that contain noisy, missing or incomplete values. Therefore, the raw data should be well prepared before it is fed to the data stream mining process, while outliers must be removed, because it provides data points that are far from the mean of the corresponding normal variables. Information generated by these applications allows the creation of a historical database for further decision making (Armentano et al., 2017; Rosa et al., 2015; Barbosa et al., 2018).

This model was developed from the investigation of other models that use information technology in agriculture, aiming to better understand the area's problems and limitations, as well as to know the related technologies. Works developed in this segment do not use the generated information for fertility prediction and soil productivity. For this article, the proposed model was detailed, as well as its main components and their relationships. To evaluate the proposed work, performance tests were performed with soil fertility and productivity calibration models.

The contributions of this paper are: (i) the usage of context history to provide a robust data background for fertility and productivity trends, individually or correlated, based on a computational model, (ii) that the results of this model show an evolution of the infrared technique over the years, when compared to other research on soil organic matter, due to technological development, data processing capacity and, specifically, the precision of the equipment used, and (iii) that the employment of machine learning with information from meteorological sensors related to wheat productivity brought surprisingly good results, strengthening the idea of applying IoT to the precision agriculture.

Therefore, the present article proposes a computational architecture that uses the ubiquitous computing paradigm in agriculture to predict soil fertility and productivity through context-history.

This work is structured in five parts. The next section describes some related work. Section three presents the proposed model and architecture implementation. Section four shows the results of some experiments and quantitative analysis and, finally, the last section is intended for the article's conclusion.

2. Related works

The selected papers subjects are related to ubiquitous computing, precision agriculture and data prediction. Table 1 compares the articles found regarding the work purpose, type of sensors, architecture, data analysis and experiment location, including the proposed model.

An on-the-go spectrometer was employed by Nawar and Mouazen in 2019 to perform real-time ground measurements using near infrared reflectance spectroscopy. Calibration models for organic matter were generated to compare online results (connected to a tractor) with laboratory results. Although promising, the online technique performed less accurately in prediction due to the effects of other affective parameters, such as soil moisture and texture (Nawar and Mouazen, 2019).

A model based on the Markov Decision was proposed by Huong and others in 2018 to create an irrigation system that makes agriculture more energy and water efficient (Huong et al., 2018). The impact of machine learning on precision farming regarding color segregation in satellite imagery was presented by Treboux and Genoud (2018). An intelligent open-source programmed system for predicting soil irrigation requirements through various sensors was proposed by Goap and other authors also in 2018 (Goap et al., 2018).

In the same year, Santos and other authors presented a wireless sensor network architecture model to predict soil temperature and humidity. In it, the obtained results signaled the viability of the proposal, the limitations of the cultivation's real time monitoring and security mechanisms for data's transmission (dos Santos et al., 2019). It is also worth noting that a wireless sensor network to enable data collection from the agricultural environment such as temperature, humidity and light was designed and implemented in 2014 by Concepcion, Stefanelli and Trinchero (de la Concepcion et al., 2014).

To compare the proposed model and the aforementioned studies, a table of related works was created. These details can be seen in Table 1. And to compare them and identify relevant characteristics for evaluation in relation to the proposed model, the following criteria were created:

1. Architectural Presence: Checks if the work has any developed architecture;

2. Notifications: Types of automatic or demanded request notifications made available to the user, with informations regarding scenarios or decision making;

3. Form of Analysis: Identifies which tool is used for data analysis. In other words, it identifies how results were generated for decision making or information to users;

4. Contextual Usage: Identifies whether models store relevant information over time. Having any event history database for later use meets this criteria;

5. Agent Usage: Identifies whether the model has agents for generating results as well as performing the necessary actions within the model;

6. Ontology Usage: Identifies whether the model uses ontology for inferring results or representing model entities. Any use of ontology is considered relevant to this topic;

7. Sensors Used: List the sensors used in the related studies;

8. Ubiquity Support: Identifies whether the model has support for ubiquity in data collection. Any item that supports ubiquity in the model can fall under this topic;

9. Web Access: Identifies whether the model has any features available through web access;

10. Mobile Access: Identifies whether the model has any features available through application access.

No study directly related to soil fertility prediction in precision agriculture was found in the literature based on contextual history analysis. The researches that presented a computational model and sensor data for decision making, mostly made use of humidity and temperature sensors, that is, they used only physical aspects of the soil or the environment, and not chemical/organic, as suggested in the present article.

3. Proposed model

The model is intended to make predictions concerning soil fertility as well as crop yield from a historical basis. For this, the model stores and uses data from near infrared sensors, pH, electrical conductivity, luminosity, rainfall, pressure, temperature and humidity, as well as the period and location where the analysis was performed. Given the lack of a computational model that aggregates a wide range of sensors to perform the analysis, in a quick and noninvasive way, that does not result in polluting residues, the present research seeks to answer the following research question: “What would an architecture that uses contextual data look like? Which sensors for predicting yield, organic matter and soil clay in precision agriculture would be used?”.

3.1. Model architecture and context history

The architectural model was based on the Standard for Technical Architecture Modeling (TAM) pattern during the design phase (SAP AG, 2007). Fig. 1 presents the architecture composed by actors (A1, A2, A3 and A4), blocks (Mobile Assistant, Manager, Actuator and Server) and accesses. The inner part of the blocks consists of components and communication channels that are defined by the Cl, C2, C3 and C4 symbols.

The data is generated by the actors who, after performing, provide contextual information according to the provided demand. For this, the model has four blocks: a mobile assistant, a manager, a middleware for sensors connected to actuators and a server.

The Mobile Assistant receives the actions of the mobile client (A1) through a communication channel (C1) and communicates with the server using *webservices*. The manager is based on a website where it is possible to configure profiles, query all sensor information and perform productivity predictions. For the managing context histories, the information that is presented in Table 2 is stored.

The Actuator consists of a middleware that is responsible for servicing the client sensor (A3) through a C3 channel, with communication methods that send data to the server as well as receive updates and new configurations. The actuating agent collects soil and environmental data. Regarding soil fertility, the actuator also verifies if the location where the tractor driver is happens to be critical for treatment. This information is made available through notification to the mobile assistant soon after the prediction of organic matter and clay occurs. Based on these results it is decided if there is need for soil correction with fertilizers beyond the regular amount that is used. Regarding the physical aspects, the productivity prediction is performed by the managing agent. This agent is also responsible for inserting tags in the corresponding context histories.

The Server component communicates with the other components through internal channels and with the external data actor (A4) through the C4 channel. The Server includes mechanisms for treatment and prediction of data related to soil analysis (spectral data) and climate information (physical aspects), as well as location. Therefore, using

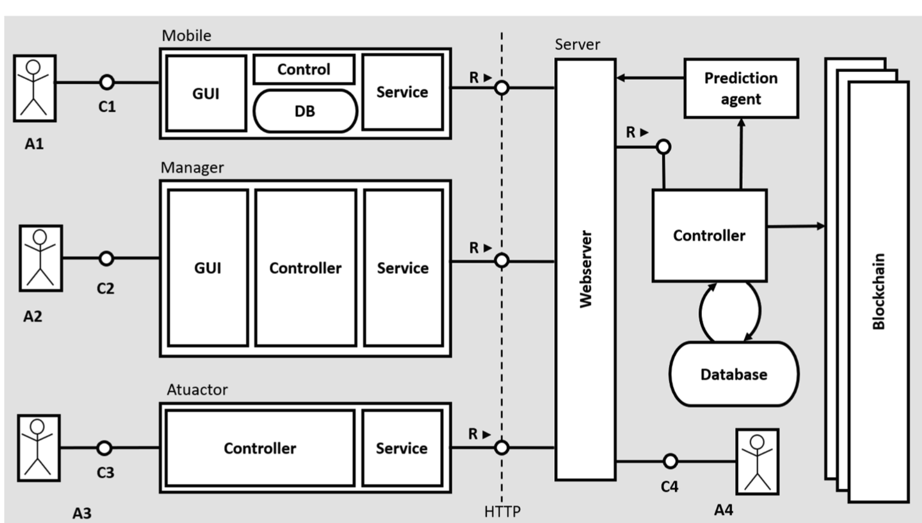


Fig. 1. Proposed model architecture.

Table 2
Property of context history.

Property	Format	Description
1. Event	Timestamp	Current Date and Time
2. Latitude	Decimal	Latitude in 17 digit decimal format
3. Longitude	Decimal	Longitude in 17 digit decimal format
4. Infrared	String	Spectral Data in CSV Format
5. Organic Matter	Decimal	Organic Matter Concentration
6. Clay	Decimal	Clay Concentration
7. pH	Decimal	Soil Hydrogen Potential
8. EC	Decimal	Electrical Soil Conductivity
9. Lux	Decimal	Luminosity
10. TempA	Decimal	Ambient Temperature
11. Humidity	Decimal	Relative Humidity Environment
12. TempS	Decimal	Soil Temperature
13. Moisture	Decimal	Soil Moisture
14. Rain	Decimal	Rain mm
15. Pressure	Decimal	Atmospheric Pressure
16. IP	String	Stores the request IP address

regression algorithms (*Machine Learning*), the results of organic matter, clay and productivity are predicted.

The Blockchain component serves as a decentralized real-time database for analysis results, which can be accessed by farmers and their supply chain partners, in order to increase product quality or alert provenance systems. This component was not developed into the model, its integration was only suggested.

3.2. Entity representation

The sensors data needs to be updated regularly and, therefore, abstract semantics are required for data acquisition to provide global accessibility to data through cloud-based services (Bhadoria and Chaudhari, 2019). An ontology has been proposed to facilitate the visualization and understanding of the entities that compose the model. It also aims to simplify and guide the development of all database modules, agents, and tables. Among those used in agriculture is AgroRDF, based on the AgroXML (Blank et al., 2013) standard, which has been developed since 2004 for data communication between farmers and

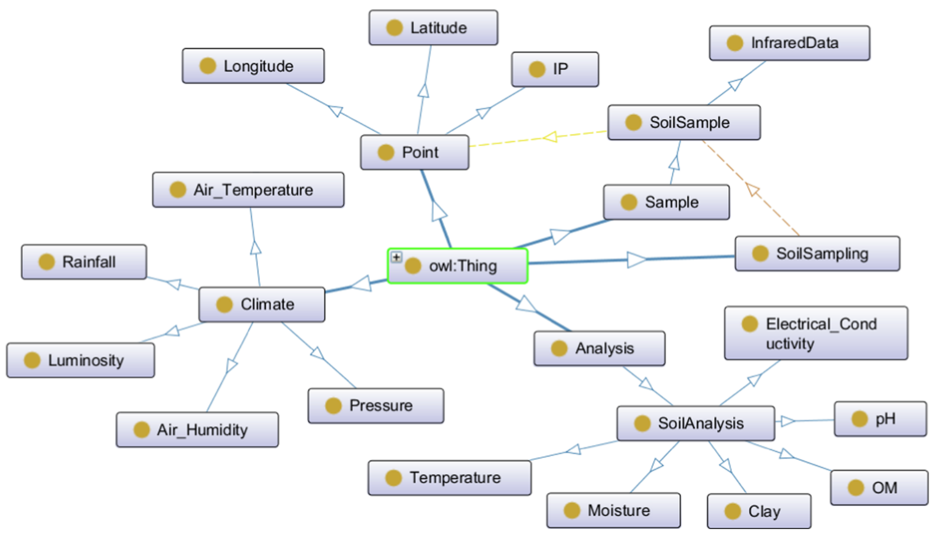


Fig. 2. Proposed ontology for the model.

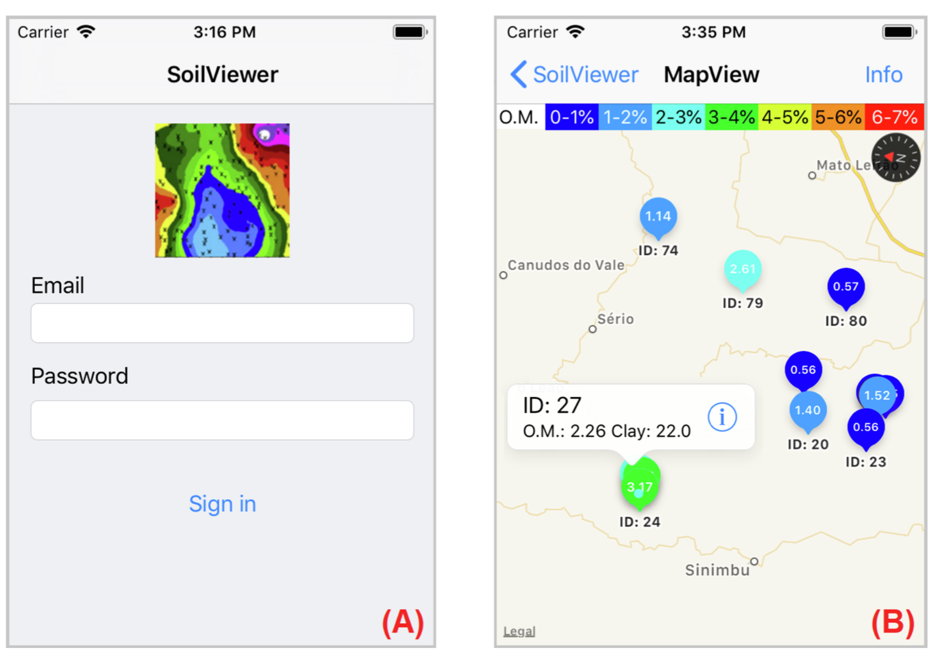


Fig. 3. Applications Interface: login screen (A) and visualization screen (B).

```
{
    "owner": 1,
    "name": "Falkland Farm",
    "address": "NC",
    "phone": "999887766",
    "email": "test@test.com",
    "sample": [
        {
            "id": 1
            "event": "2018-11-22 10:44:29",
            "latitude": "-29.8601376",
            "longitude": "-52.9074892",
            "infrared": "0.334,0.332,...,0.218",
            "om": "1.85045875108",
            "clay": "6.876870040",
            "pH": "7.5",
            "ec": "1.53",
            "temp": "21.22",
            "humidity": "66.4",
            "ip": "191.4.236.170",
        },
    ]
}
```


Fig. 4. Context history example for soil characterization.


```
{
    "owner": 1,
    "name": "Falkland Farm",
    "address": "NC",
    "phone": "999887766",
    "email": "test@test.com",
    "climate": [
        {
            "id": 1
            "event": "2018-11-22 11:02:39",
            "latitude": "-29.8601417",
            "longitude": "-52.9074821",
            "air_temp": "24.07",
            "air_hum": "69.21",
            "pressure": "1014.73",
            "luminosity": "165.18",
            "rainfall": "1.22",
            "ip": "191.4.236.170",
        },
    ]
}
```


Fig. 5. Context history example for weather events.

other supply chain partners (Schmitz et al., 2009). To represent the environments, results, resources and conditions of the sample, an extension of AgroRDF was performed, as shown in Fig. 2, but was not used for queries, only for general representation of entities and their relationships.

In the model, subclasses were also added, such as the *Infrared* in the SoilSample class, the *IP*, *Latitude* and *Longitude* in the Point class, *OM (Organic Matter)*, *Clay*, *Temperature*, *pH*, *Moisture* and *Electrical Conductivity* in the SoilAnalysis class, *Air Temperature*, *Air Humidity*, *Pressure*, *Rainfall* and *Luminosity* in the Climate class.

3.3. Implementation

The mobile assistant is designed for smartphones with Android and iOS operational systems. Thus, using the IDE Android Studio and Xcode, two interfaces were created: a login screen (A) and a screen where it was possible to view the infrared collection points with the respective prediction of organic matter and clay results (B), as shown in Fig. 3, in addition to receiving *Broadcast* messages that are related to notifications from the actuator, as reported in item 3.1.

The server was implemented in Python version 2.7 and uses Machine Learning algorithms (Scikit-Learn library) through prediction agents. The actuator consists of a mobile near-infrared sensor (Texas Instruments DLP NIRScan Nano), a conductivity meter, a pH meter, ground temperature and humidity sensors, and others that measure climate change, all coupled to a Raspberry Pi Model 3. A GPS module (model GY-GPS6MV2) was also connected to capture the latitude and longitude of the site and a GPRS board (model SIM800L) for data communication. An example context history database for the fertility model is illustrated in Fig. 4, while Fig. 5 provides an example of climate aspects.

For fertility prediction, a total of 450 soil samples were selected from different collection points of the Vale do Rio Pardo/RS, where the organic matter concentrations ranged from 0.41% to 6.15% and clay from 4% to 72%. These samples were supplied by the Central Analitica soil laboratory, where they were dried in a Marconi-MA037 model oven with air circulation for a period of at least 24 h at a temperature between 45 and 60 °C. Afterwards, the samples were ground in a Marconi-NI040 hammer mill, with 2 mm strainer, and stored in cardboard boxes (Tedesco et al., 1995; Bernardi et al., 2014).

A data set can be categorized by several features that affect the acceptability of input and parameters, such as distinctiveness, reliability, permanency, contestability and universality (Arya and Bhadoria, 2019). So, PLS was chosen because it is responsible for weighting these features according to their discriminative power for each different descriptor and also because of the minimal demands on measurement scales, sample size and residual distributions (Mehmood et al., 2012).

All PLS models were developed based on Daniel Pelliccia's website, which provides a step by step tutorial on how to build a NIR calibration

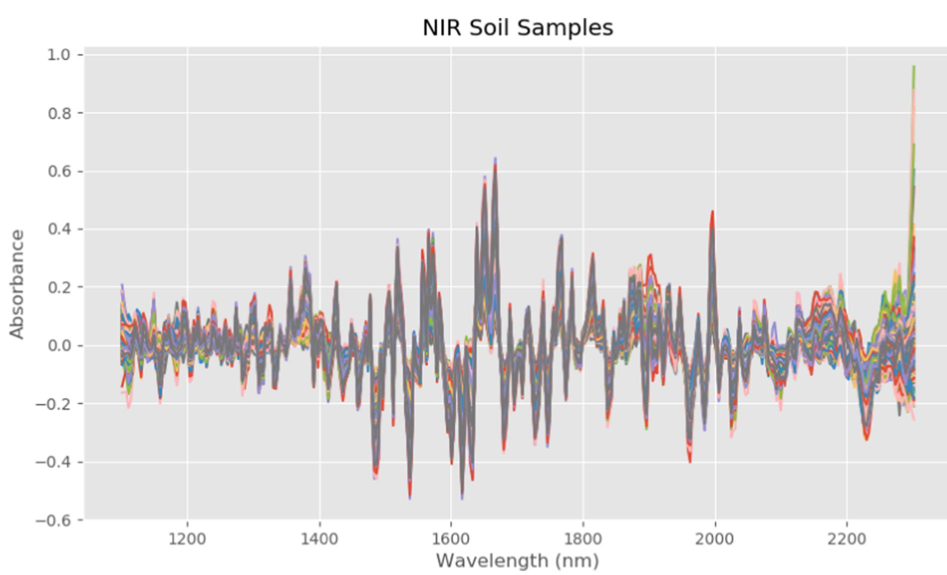
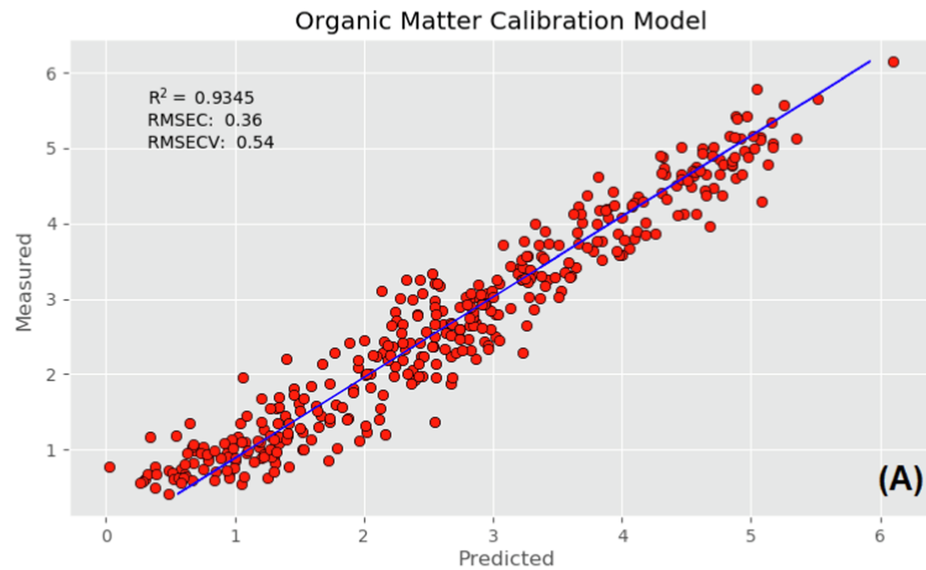


Fig. 6. Infrared data treated with SNV + Savitzky-Golay filters.



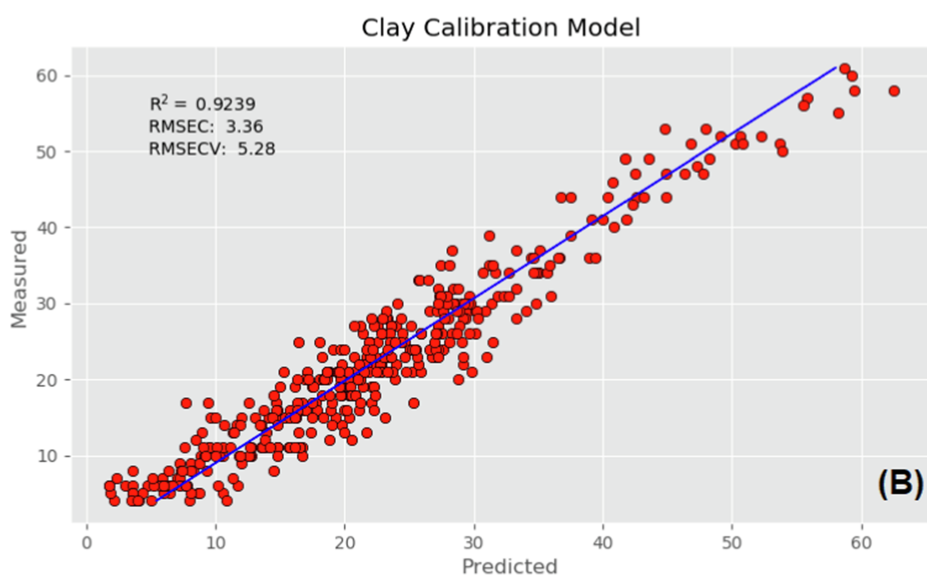


Fig. 7. Organic Matter (A) and Clay (B) calibration models.

model using Partial Least Squares Regression in Python (Pelliccia, 2019). In the proposed algorithm, the X variable represents the whole matrix data (sample vs. spectral absorbance/ sensor information) and the Y variable fits with concentrations (percentual of organic matter/ clay) or wheat productivity (tons per hectare), respectively.

3.4. Prediction models

Infrared data acquisition (spectra) was performed directly on the surface of the samples with the aid of a fiber optic spectrometer. The system operated at a 3.2 nm resolution in the region from 951 to 2450 nm, 4 cm^2 scan area (adjustable), with a speed of 6 samples per minute, according to the manufacturer's guidelines.

In the PLS model for soil fertility, a range between 1101.8 and 2301.6 nm of near infrared spectra was used, totaling 382 variables for organic matter and clay scenarios. To reduce baseline variation and spectral noise, the data was preprocessed by techniques such as Singular Normal Variate and Savitzky-Golay (1st derivative, 2nd polynomial order, 5 window points), available in the SciPy library. Finally, two PLS regression models were built for each fertility parameters after autoscaling the data. The organic matter model was setup with 17 factors, while the clay model used 16 of them. The optimum number of factors was chosen by the lower root mean squared error of prediction (RMSEP) computed in a range of at least 30 factors.

To predict productivity, in order to simulate the performance of the developed Machine Learning model and the lack of sensors installed in the field so far, the open data from the Siegfried Emanuel Heuser Economics and Statistics Foundation repository was used (Abertos, 2019). The extracted data presented the wheat yield values (T/ha) between 2001 and 2015 in the Santo Ângelo/RS region, which were correlated with climate information from the National Institute of Meteorology repository (de Meteorologia, 2019). The climatic values used in the model refer to the annual averages of the planting period until harvest (May to November) of insolation (hrs), precipitation (mm), atmospheric pressure (mbar), maximum temperature (Celcius), minimum temperature (Celcius), compensated temperature (Celcius), relative humidity (%), wind speed (m/s) and Piche evapotranspiration (mm) from meteorological station number 83907, located in the neighboring city of São Luiz Gonzaga/RS, totaling 9 variables. The productivity PLS model was built with mean centering data and 8

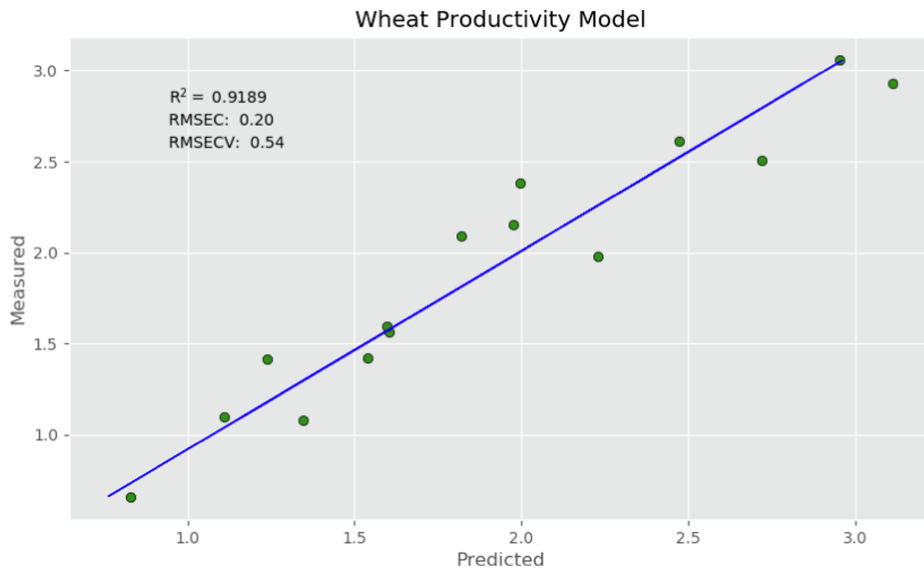


Fig. 8. Wheat Production calibration model.

Table 3
Productivity model performance – wheat.

Year	Productivity measured (T/ha)	Productivity predicted (T/ha)	Absolute error (T/ha)	Relative error (%)
2001	1.562	1.603	0.041	2.64
2002	1.418	1.236	0.182	12.83
2003	1.098	1.108	0.010	0.91
2004	2.093	1.819	0.274	13.09
2005	1.594	1.598	0.004	0.25
2006	1.419	1.539	0.1203	8.46
2007	0.662	0.829	0.167	25.22
2008	1.977	2.229	0.252	12.75
2009	2.611	2.473	0.138	5.29
2010	2.385	1.997	0.388	16.27
2011	2.504	2.721	0.217	8.67
2012	2.925	3.111	0.186	6.36
2013	2.150	1.976	0.174	8.09
2014	3.055	2.952	0.103	3.37
2015	1.082	1.345	0.263	24.31
Average	1.902	1.902	0.168	9.90

factors setup.

4. Results and discussions

To represent the calibration model for organic matter and clay, the average spectrum of 10 infrared analyzes each soil sample, composing its fertility database. Fig. 6 shows the changes in the profile of the near-infrared molecular spectroscopy, based on the 450 samples that were used in this modeling phase.

All soil samples employed were analyzed in triplicate by the official method (ROLAS) so that the necessary correlations could be made. Before calibration models was used Hotelling T^2 method inside a PCA model for remove outliers, a total of 382 spectra were used for Organic Matter and 397 spectra for Clay calibration.

The graphs in Fig. 7 shows the performance of these models of organic matter (A), whose Pearson correlation coefficient (R^2) was 0.9345, mean square error of calibration (RMSEC) of 0.36% and mean square error of cross-validation (RMSECV) of 0.54%, and clay (B), with R^2 of 0.9239, RMSEC of 3.36% and RMSECV of 5.284%.

Comparing the results of the calibration models generated by the context histories with works by different authors, better predictive evaluation indicators were obtained. For organic matter, Nawar and Mouazem in 2019 had an R^2 of up to 0.84, but a smaller RMSEC of 0.14%, as well as clay, Wetterlind and others in 2015 obtained an R^2 of 0.76 and an RMSECV of 6.4% (Nawar and Mouazen, 2019; Wetterlind et al., 2015). The smaller RMSEC errors of these authors are due to the low sample representativeness. Both focused their experiments on batches of approximately 0.5 km^2 while the context histories of the proposed model represented several collection points of the Vale do Rio Pardo/RS (around 13,000 km^2 of rural and urban area).

Furthermore, other factors such as sensitivity, reproducibility and interference of equipment, intrinsic to the method, could also be discussed. It is noteworthy that the model tends to improve its predictive capacity as new contextual data builds on the historical basis.

For wheat yield prediction, the results obtained were an R^2 of 0.8725, RMSEC of 0.25 T/ha and RMSECV of 0.88 T/ha.

The graph in Fig. 8 shows the performance of the wheat production yield model whose (R^2) was 0.9189, RMSEC of 0.20 T/ha and RMSECV of 0.54 T/ha. In comparison, Aleksandra Wolanin in 2019 presented an R^2 of 0.89 and RMSEC of 1.58 T/ha using optical sensor data installed on Sentinel-2 and Landsat 8 satellites using Machine Learning methods (Wolanin et al., 2019).

Table 3 compares the measure wheat yield values with those predicted through the calibration model. The relative error values are computed between absolute error and measure data. Excluding the years 2007 and 2015 which had a rainy climate resulting in low productivity, in general the proposed prediction methodology is perfectly adequate for implementation as a precision farming tool. The same calculation was proposed for the prediction models of organic matter and clay and presented an average relative error of 15.55% and 16.37% for all data samples, respectively.

Table 3 compares the measured wheat yield values with those predicted through the calibration model. The relative error values are computed between absolute error and the measured data. Excluding the years of 2007 and 2015 which had a rainy climate that resulted in low productivity, in general terms, the proposed prediction methodology presents itself as adequate to be implemented as a precision farming tool. The same calculation was proposed for the prediction models of organic matter and clay and presented an average relative error of 15.55% and 16.37% for all data samples, respectively.

5. Conclusion

The present article proposed a computational model for predicting soil fertility and productivity through context-history. Based on the information used, it is possible to build a historical database for later use. Regarding fertility, infrared data was analyzed so that organic matter and clay concentrations could be predicted. For productivity, data from open repositories supported the model built from *Machine Learning*.

Thus, the following work strategies were highlighted: context prediction based on linear regression models, validation of the obtained results and architecture development. The authors recognize that the model determination coefficient for clay prediction could be better. Due to this, it is suggested the use of a spectral region selection technique, using similarity and parallel processing models, as well as a further study to eliminate *outliers*, both aiming at improving predictive capacity.

It is noteworthy that the generated data served to evaluate the performance of the model. The real variables to be used in the context histories will come from the sensors that were installed in the field, that is, the temperature, humidity, atmospheric pressure, pH, electrical conductivity, luminosity and rainfall sensors, as well as infrared, which was the basis for the information of soil fertility.

All models used Partial Least Squares Regression as the multivariate technique due the linear relation between data and its property of interest. The use of PLS in spectral data (soil, food, medicine, fuel, etc.) is known within the scientific community but, for weather sensor data and wheat productivity, no references were currently found.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Gilson Augusto Helfer: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Jorge Luis Victória Barbosa:** Conceptualization, Visualization, Supervision, Writing - review & editing, Project administration. **Ronaldo dos Santos:** Data curation, Methodology, Investigation. **Adilson Ben da Costa:** Conceptualization, Visualization, Validation, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgement

The authors acknowledge the Rio Grande do Sul State Research Support Foundation (FAPERGS), the Higher Education Personnel Improvement Coordination – Brazil (CAPES) – Financing Code 001, the National Council for Scientific and Technological Development (CNPq), the University of Santa Cruz do Sul and the Central Analytical Sector, for providing soil samples, as well as the University of the Vale do Rio dos Sinos (Unisinos) for supporting the development of this work. The authors would like to especially acknowledge the Unisinos Graduate Program in Applied Computing (PPGCA) and the Mobile Computing Laboratory (Mobilab) support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.compag.2020.105602.

References

Abertos, F.D., 2019. Fundação de economia e estatística siegfried emanuel heuser — rs. URL:https://dados.fee.tche.br.

Anzanello, M.J., 2013. Seleção de variáveis para classificação de bateladas produtivas com base em múltiplos critérios. Production 23 (4), 858–865. https://doi.org/10.1590/S0103-65132013005000001.

Armentano, R., Bhadoria, R.S., Chatterjee, P., Deka, G.C. (Eds.), The Internet of Things, Chapman and Hall/CRC, 2017. doi: 10.1201/9781315156026.

Arya, K.V., Bhadoria, R.S., 2019. The Biometric Computing. Chapman and Hall/CRC. https://doi.org/10.1201/9781351013437.

Barbosa, J., Tavares, J., Cardoso, I., Alves, B., Martini, B., 2018. Trailcare: An indoor and outdoor context-aware system to assist wheelchair users. International Journal of Human-Computer Studies 116, 1–14. https://doi.org/10.1016/j.ijhcs.2018.04.001.

Bernardi, A., Naime, J., Resende, A., Inamasu, R., Bassoi, L., 2014. Agricultura de precisão: resultados de um novo olhar. Manual de métodos de análise de solos, second ed., EMBRAPA, São Carlos, SP. URL:http://ainfo.cnptia.embrapa.br/digital/bitstream/item/113993/1/Agricultura-de-precisao-2014.pdf.

Bhadoria, R.S., Chaudhari, N.S., 2019. Pragmatic sensory data semantics with service-oriented computing. Journal of Organizational and End User Computing 31 (2), 22–36. https://doi.org/10.4018/joeuc.2019040102.

Blank, S., Bartolein, C., Meyer, A., Ostermeier, R., Rostanin, O., 2013. IGreen: A ubiquitous dynamic network to enable manufacturer independent data exchange in future precision farming. Computers and Electronics in Agriculture 98, 109–116. https://doi.org/10.1016/j.compag.2013.08.001.

Costa, N.R., Carvalho, M.d.P.e., Dal Bem, E.A., Dalchiavon, F.C., Caldas, R.R., 2014. Orange yield correlated with soil chemical attributes aiming specific management zones. Pesquisa Agropecuária Tropical 44 (4), 391–398. doi:10.1590/S1983-40632014000400001.

de la Concepcion, A.R., Stefanelli, R., Trinchero, D., 2014. A wireless sensor network platform optimized for assisted sustainable agriculture. In: IEEE Global Humanitarian Technology Conference (GHTC 2014), IEEE, pp. 159–165. doi: 10.1109/GHTC.2014.6970276.

de Meteorologia, I.I.N., 2019. Ministério da Agricultura, Pecuária e Abastecimento. URL:http://www.inmet.gov.br.

Dey, A.K., Abowd, G.D., Salber, D., 2001. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction 16 (2–4), 97–166. https://doi.org/10.1207/S15327051HCI16234_02.

do Carmo, D.L., Silva, C.A., de Lima, J.M., Pinheiro, G.L., 2016. Electrical conductivity and chemical composition of soil solution: comparison of solution samplers in tropical soils. Revista Brasileira de Ciência do Solo 40. https://doi.org/10.1590/18069657rbcs20140795.

dos Santos, U.J.L., Pessin, G., da Costa, C.A., da Rosa Righi, R., 2019. Agriprediction: A proactive internet of things model to anticipate problems and improve production in agricultural crops. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2018.10.010.

Ferrão, M.F., Davanzo, C.U., 2005. Horizontal attenuated total reflection applied to simultaneous determination of ash and protein contents in commercial wheat flour. Analytica Chimica Acta 540 (2), 411–415. https://doi.org/10.1016/j.aca.2005.03.038.

Ferreira, M.M.C., 2015. Quimiometria: conceitos, métodos e aplicações, Editora da Unicamp. Campinas-SP. https://doi.org/10.7476/9788526814714.

Furtini Neto, A.E., Vale, F.R., Resende, A.V., Guilherme, L.R.G., Guedes, G.A.A., 2001. Fertilidade do solo, first ed. UFLA/FAEPE, Lavras.

Geladi, P., 2003. Chemometrics in spectroscopy. part 1. classical chemometrics. Spectrochimica Acta Part B: Atomic Spectroscopy 58 (5), 767–782. https://doi.org/10.1016/s0584-8547(03)00037-5.

Goap, A., Sharma, D., Shukla, A., Rama Krishna, C., 2018. An iot based smart irrigation management system using machine learning and open source technologies. Computers and Electronics in Agriculture 155, 41–49. doi: 10.1016/j.compag.2018.09.040.

Griebeler, G., da Silva, L.S., Cargnelutti Filho, A., Santos, L.d.S., 2016. Avaliação de um programa interlaboratorial de controle de qualidade de resultados de análise de solo. Revista Ceres 63 (3), 371–379. https://doi.org/10.1590/0034-737X201663030014.

Höskuldsson, A., 1988. PLS regression methods. Journal of Chemometrics 2 (3), 211–228. https://doi.org/10.1002/cem.1180020306.

Huong, T.T., Thanh, N.H., Van, N.T., Dat, N.T., Long, N.V., Marshall, A.J., 2018. Water and energy-efficient irrigation based on markov decision model for precision agriculture. In: 2018 IEEE Seventh International Conference on Communications and Electronics (ICCE), pp. 51–56. https://doi.org/10.1109/cce.2018.8465723.

Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in partial least squares regression. Chemometrics and Intelligent Laboratory Systems 118, 62–69. https://doi.org/10.1016/j.chemolab.2012.07.010.

Miranda, J., da Costa, L.M., Ruiz, H.A., Einloft, R., 2006. Composição química da solução de solo sob diferentes coberturas vegetais e análise de carbono orgânico solúvel no deflúvio de pequenos cursos de água. Revista Brasileira de Ciência do Solo (4), 633–647. https://doi.org/10.1590/S0100-06832006000400004.

Muñoz, J.D., Kravchenko, A., 2011. Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. Geoderma 166 (1), 102–110. https://doi.org/10.1016/j.geoderma.2011.07.017.

Nawar, S., Mouazen, A., 2019. On-line vis-nir spectroscopy prediction of soil organic carbon using machine learning. Soil and Tillage Research 190, 120–127. https://doi.org/10.1016/j.still.2019.03.006.

Pelliccia, D., 2019. Partial least squares regression in python. URL:https://nirpyresearch.com/partial-least-squares-regression-python/.

Rosa, J.H., Barbosa, J.L.V., Kich, M., Brito, L., 2015. A Multi-Temporal Context-aware System for Competences Management. International Journal of Artificial Intelligence in Education 4, 455–492. https://doi.org/10.1007/s40593-015-0047-y.

SAP AG, Standardized technical architecture modeling – conceptual and design level. URL:http://www.fmc-modeling.org/download/fmc-and-tam/SAP-TAM_Standard.pdf (accessed 23-Jan-2019) (mar 2007).

Schmitz, M., Martini, D., Kunisch, M., Mösinger, H.-J., 2009. agroXML enabling standardized, platform-independent internet data exchange in farm management information systems. In: Metadata and Semantics, Springer US, Boston, MA, pp. 463–468. doi: 10.1007/978-0-387-77745-0_45.

Tedesco, M., Gianello, C., Bissani, C., Bohnen, H., Volkweiss, S., 1995. Análise de solo, plantas e outros materiais, second ed. UFRGS/Departamento de Solos, Porto Alegre.

Treboux, J., Genoud, D., 2018. Improved machine learning methodology for high precision agriculture. 2018 Global Internet of Things Summit (GIoTS), 1–6 doi: 10.1109/GIOTS.2018.8534558.

Valderrama, P., Braga, J.W.B., Poppi, R.J., 2009. Estado da arte de figuras de mérito em calibração multivariada. Química Nova 32 (5), 1278–1287. https://doi.org/10.1590/s0100-40422009000500034.

Wetterlind, J., Piikki, K., Stenberg, B., Söderström, M., 2015. Exploring the predictability of soil texture and organic matter content with a commercial integrated soil profiling tool. European Journal of Soil Science 66 (4), 631–638. https://doi.org/10.1111/ejss.12228.

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., Guanter, L., 2019. Estimating crop primary productivity with sentinel-2 and landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sensing of Environment 225, 441–457. https://doi.org/10.1016/j.rse.2019.03.002.