Intelligent spectroscopy system used for physicochemical variables estimation in sugar cane soils

Ofelia Landeta-Escamilla¹, Oscar Sandoval-Gonzalez¹, Albino Martínez-Sibaja¹*, J. J. Agustín Flores-Cuautle², Rubén Posada-Gómez² and Alejandro Alvarado-Lassman¹

¹ Tecnológico Nacional de Mexico/ I.T.Orizaba, Orizaba, VZ, 94320, MX; ² CONACYT-TecNM/ I.T.Orizaba, Orizaba, VZ, 94320, MX; * Correspondence: cei.depi.i to@gmail.com; Tel.: +52 2727033425

Abstract: Soil conditions is a major aspect of interest for farmers due to the knowing of the physicochemical properties of the same can help with any necessary restoration of soil that guarantees the quality and the production of their crop. However, technology and analysis of the soil become of difficult access mainly in developing countries, by which the present paper shows the development of a system thought to estimate physicochemical variables of soils growing sugar cane through studies of spectroscopy. Its characteristic is that it is a portable system, with low cost, easy to use and can estimate physicochemical variables in-situ with the objective of knowing the degree of degradation present in the soil and through this help the farmers define possible strategies to restore it. The device uses the frequency response of the soil determining values of magnitude and phase, which are used by algorithms of artificial intelligence capable of getting an estimation of the physicochemical properties. The obtained results show errors below 8% in the estimation of the variables compared to the analysis results of the soil at laboratories.

Keywords: Artificial intelligence, electronics, psychochemical properties, frequency response of soil.

1. Introduction

A global issue to attend due to its relevancy is soil degradation, firstly, because it is a natural resource, formed and regenerated slowly, not renewable, and very fragile in its degradation [1]. Therefore, it becomes one of the most vulnerable groups, because it is not possible to take actions to avoid or mitigate negative impacts produced by environmental detriment [2], and its crucial role in ecosystems functioning [3] and due to the impact on agronematics and food security [4] by decreasing its current and potential capability to produce goods and services. Land management is far from being sustainable, there has been an inadequate operation of soils in the conventional production systems [5], the excessive use of fertilizers has led to nitrogen depositions affecting not only surface and underground water, but also soils with acidification, because an incorrect application of modern agrotechnologies [6] generates unstable ecosystems which creates external dependence of energy and materials for their continuity in time [7]. Limiting factors in the production systems are increased, with the nutrients and water, interfering in the biogeochemical cycles [8]. On the opposite site, there are places where the lack of fertilizers is degrading soils and decreasing yield, which is important because they also help with carbon fixation to mitigate climate change. Hence, it is necessary to monitor soils health [9] to determine current soils status and to monitor nutrients levels to apply sustainable land management practices.

Through this perspective, the soil is not only a support where crops grow, but also an active and dynamic component with clear repercussions in the development and production of plants.
Organizations like the Food and Agriculture Organization (FAO), the United Nations Convention to Combat Desertification (UNCCD), and other participants in the Collaborative Partnership on Forest have been involved in determining the level of land and soil degradation and are promoting management practices and restoration of degraded areas urgently. It has been determined, using NDVI (Normalized Difference Vegetation Index) analysis and deforestation statistics, that the biggest biomass decline and forest loss have taken place in Eastern Africa, Southeast Asia and Central and South America [10][11].

To classify global land degradation a scale based on levels ranging from 1 to 4 has been proposed by FAO-SOLAW (State of the World’s Land and Water Resources), being Type 1 - high degradation trend or highly degraded lands and Type 4 - Improving lands. The status for 2011 was 25% for Type 1, 8% for type 2, 36% for type 3, 10% for type 4 and the remaining 20% for bare areas and water. Additionally, the level of degradation is associated with poverty levels [11]. A different form to classify it, could be described as 56% is due to water erosion, 28% related to wind erosion, 12% to chemical degradation and 4% to physical degradation, all of them leading to a loss of organic matter, changes in nutrient content and toxic substances, salinization and compaction [12].

There are many approaches to determine the level of soil degradation, some of them treat the problem rather as a single one instead of a complex process integrating variables such as: vegetation growth, flow of water, infiltration, land use and land management, some variables could be partially hidden by management practices, the use of fertilizers[13] and land resilience [14]. Some others analyzed information from global datasets, which may mask local changes and most relationships are based on local studies and should not be used to infer global conclusions.

A more detailed physicochemical analysis by methodologies carried out at laboratories require to have glassware, apparatus, chemicals, and trained personnel. Although these methodologies have been validated, to assure quality control it is necessary to have controlled samples to evaluate method precision and uncertainty. That makes the access difficult for farmers, not all the samples taken are representative of the real condition of soils and some locations are impossible to access.

Thus, investigations have been carried out to obtain correct results in field with the use of different sensors and technology reducing the time and cost of the regular analysis. Sensors operate by measuring physical or chemical properties via transducers, while other technology has focused on linking sensors with artificial intelligence or satellite images to infer soils situation in general.

Methods to determine chemical properties include: (1) electrical / electromagnetic sensors to obtain organic matter or total carbon content, salinity, CEC, and residual nitrate or total nitrogen content, (2) optical and radiometric sensors to determine OM or OC, pH, CEC, residual nitrogen or total nitrogen content, and (3) electrochemical to determine salinity, pH, residual nitrate or total nitrogen content, and other macronutrients [15]. The property used in electrical and electromagnetic sensors is the ECa. Accordingly, ECa data (EMI sensors) was linked to global positioning system (GPS). Results collected for 4 years were related to soil properties and pasture yield, finding important variations in macronutrients (50 – 110%) and 20 to 26% of variability in organic matter and clay content and stability (less than 10%) for pH and ECa throughout the area evaluated. Significant correlations were found between ECa and altitude and pH and between ECa and grasses, other species, and yield [16]. Other two studies using this property, while Serrano et al. concluded no correlation was found, Kweon et al. by including a second chemical variable (CEC) determined a good correlation, this highlights the importance of using a matrix of variables to obtain better results.

Other methodology implemented to determine chemical properties is the optics, by using spectroscopy which is very sensitive [17] and operates by comparing amplitude and phase of signals incident on and reflected from soils [18]. It has been implemented in several studies to determine many variables; such as moisture, OM, pH, EC, CEC, TC, ammonium nitrogen, hot water extractable nitrogen, nitrates, total Nitrogen (TN), available phosphorus, Phosphorus absorptive Coefficient (PAC), where twelve spectroscopic models were developed and correlations obtained were in the range of 0.45 to 0.93 [19], one more only included OC, moisture and TN finding excellent precision for the online determinations [20].
A related study aiming to determine a wide range of variables included: OM, Water content (WC), Bulk Density (BD), CEC, Calcium (Ca), Magnesium (Mg), pH, Iron (Fe), Copper (Cu), Phosphorus (P), Nitrates- Nitrogen (NO3 - N), Potassium (K), and Sodium (Na) showed a correlation (R2) of 0.75 was found between OM and N, Fe, Ca, Mg, CEC, WC, and BD in a soil and a correlation of OM with Ca, Mg, CEC, WC, and BD [21].

A related study aiming to determine organic matter and moisture concluded that although the models were robust, specific calibrations at locations were needed because of the interfering nutrients. [22]. On the other hand, a study evaluating OM and texture found a correlation for the prediction of sandy soils was found with an R2 of 0.63 and for OM of 0.83 [23].

Mouazen et al. in 2007 published a research providing results for the variables: Carbon, moisture content, pH and phosphorous, and generating general calibration models. Predictions results were good for moisture with an R2 = 0.89 and only quantitative approximations for predictions related to pH, carbon and phosphorous (R2 = 0.71, 0.73, 0.96 respectively) whereas for the measurements, a large variability was observed within small fields [24].

A later research by the same authors, (2014) joins Artificial intelligence to spectroscopy improving the precision of moisture and OC, concluding that there is a strong correlation between moisture, clay content and plasticity index, meanwhile, the study carried out by [25], showed no difference with the fusion of elements. Another fusion, combining Vis-NIR spectrometry and an ECa sensor (EM38) [26], indicated that more studies and statistical analysis using sensor fusion are necessary to determine soil properties networks and increase the accuracy of determination.

Spectroscopy has been used along with WSN to improve the precision of the measurements, to monitor variables such as irrigation, fertilization, pesticides control, animal and pastures monitoring, horticulture, greenhouse and viticulture in large areas [27], the normalized soil moisture index (NMSI) obtained using reflectance values and exploratory analysis of the raw spectra using an artificial intelligence tool, the principal component analyses R(PCA) [28], an impedance spectroscopic sensor capable of measuring multiple frequencies to determine soil moisture and ionic content and has a built-in self-calibration system. An advantage of these systems is the capability of calibrating the system to different types of soils and whether their main disadvantage is the noise they have that could decrease accuracy. Results report having a 10% error range for real and imaginary impedance and 12% for soil saline water content measure which means mixed models allow to determine specific ions such as nitrates, sulphates and phosphates [18].

From previous lines, it stands out that grouping technologies must be done to find more suitable solutions to determine the level of degradation in soils accurately and thus, propose local solutions based on the causes found at the specific locations. Additionally, the relationship between poverty and the percentage of land degradation is also a distinguished issue because it highlights the need to propose new low-cost technology to determine needs at these places and promote sustainable land management practices. Therefore, the Intelligent Multimodal Soil Sensors as the group of sensors that determine physicochemical and microbiological properties through Artificial Intelligence techniques. The developed system seeks to create a low-cost technology and easy implementation, so that the farmers could handle it, being able to boost and create technology which can know the current state of soils so that strategies of restoration could be implemented.

2. Materials and Methods

2.1 Soil’s Samplings

The region of so-called high mountains in Veracruz, Mexico is formed by 57 municipalities with an approximate area of 6053 km2. The 0.5% of the total surface is planted by sugar cane and the vertisol soil represents the 18.5%. Therefore, an essential point to develop the presented system as the sampling of soils with sugar cane production made in the region of high mountains. One hundred soil samples along this mentioned area were taken; those samples were collected from the mentioned region using a method described in (UDA, 2014). All samples were sent to a specialized soil laboratory to obtain the following physicochemical properties: pH, tampon pH, organic matter (MO), phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), sulfur (S), Boro (B), Copper (Cu),
iron (Fe), Manganese (Mn), Zinc (Zn), Sodium (Na), Conductivity, Nitrogen-Nitrate (N\text{2–NO}_3^-), Ionic Exchange Capacity (CIC), cationic saturation (SC) for K, Ca, Mg, and Na, Hydrogen (H), K/Mg ratio, Ca/Mg ratio, texture (sand, lime and clay), apparent density, field capacity 1/3 bar and permanent wilting point 15 Bar. These analyses were carried out with the purpose of having a database that could be used to train the artificial intelligence system, which can estimate such variables in one side and on the other, have a collection of data that could allow the validation and the comparison of the system created with respect to the conventional lab studies. Figure 1 shows the pH, P, K, Ca, Mg, Na, Conductivity and N\text{2–NO}_3^- laboratory results from the analyzed samples.

![Figure 1. Variables statistics of the 100 soil samples (pH, P, K, Ca, Mg, Na, Conductivity and N\text{2–NO}_3^-).](image)

### 2.2 Development of the system

The system of spectroscopy to estimate physicochemical variables in sugar cane soils is integrated by 4 sections: (a) the data signal processing spectroscopy module, (b) the capacitive sensor, (c) the controlled temperature camera and the estimation of physicochemical properties system using artificial intelligence.

The module of data signal processing spectroscopy module was developed to obtain frequency response in terms of magnitude and phase of a soil sample. This is achieved through a frequency sweep from 0 to 100 KHz of a sinusoidal signal applied to the capacitive sensor. The soil sample under study in the middle of two parallel plates which has the purpose of applying a controlled electric field of Magnitude and frequency. Between the capacitor plates, which were made with stainless steel measuring 10 cm long, 10 cm width, and 3 mm thick, the soil sample to be analyzed is placed as a dielectric (the soil sample is restrained in a block of 10 cm x 10 cm x 1 cm). By applying a difference of potential (V) between both terminals of the capacitor, an electric field is established (E) between them. This electric field is created as a direct result of the charges separation according to their charge type. The frequency response allows to know the content of salts and minerals which present a resonance to a specific frequency. The differences in magnitude and phase between the entrance signal and the obtained one in the capacitor are the variables used to correlate the data to make an estimate of the physicochemical properties in soils.

Figure 2 shows the configuration of the electric circuit. The sensor is part of a voltage divisor and acts as an RC filter, which is required to know the frequency response of the sensor. The excitation signal consists of a sinusoidal wave of 10 volts peak to peak (10VPP) which is generated by the control module. This control module is formed by a MyRIO platform from National Instruments that contains an FPGA XILIX Zynq-7010 and a 667 MHz dual-cortex ARM Cortex-A9 processor, which was programmed to generate a sinusoidal signal with a frequency sweep from 0-50KHz in steps of 100Hz. The signal is sent to the RC circuit shaped by the resistance and capacitive sensor plates, the soil sample is placed as the dielectric of the capacitor. In each step of the excitation signal in the frequency sweep the frequency response in magnitude and phase is obtained from the capacitor. At the end of the study it is possible to obtain the behavior of the sensor in Magnitude and phase throughout the band of frequencies. This graphic represents a distinctive pattern which allows
the algorithms of artificial intelligence to estimate and quantify the content of the physicochemical variables.

**Figure 2.** General diagram of the functioning of the capacitive sensor using the spectroscopy technique

Figure 3 shows the block diagram for the generation of the sine wave, frequency modulation and the signal acquisition of the capacitor. For the generation of the sine wave, a FPGA signal generator block was used, it is possible to modify the frequency and the amplitude is modified using a conventional division. This signal is passed through the DAC Port C, pin A01 and an RMS value is obtained using the FPGA RMS block. Running in the same loop, the Capacitor signal is acquired using the ADC Port C, pin AI1 an RMS value is also computed.

**Figure 3.** LabVIEW block diagram of the sine wave generation, frequency modulation and the signal acquisition of the Capacitor.

A Graphical User Interface (GUI) was designed and programmed in LabVIEW which is connected to the FPGA and the artificial intelligence block as it is shown in Figure 5. The user is able of modifying different parameters like the frequency steps, initial and final frequency and path locations. Moreover, the interface plots the results of the frequency response of the capacitor in magnitude and phase in real time. The results obtained during the analysis are sent to the artificial intelligence block in Python via TCP/IP. The artificial intelligence block receives the magnitude and
phase data previously acquired during the experiments and compute the results of each IA algorithm
sending back the result to the PC to be visualized in the GUI.

Figure 4. Integration of the PC GUI with the FPGA and the Artificial Intelligence Block

Figure 5 shows the tests of the system functioning performing in real time. The signal of
sinusoidal excitation can be observed in white and in red the signal of the response of the capacitor.
The changes in the magnitude and phase of the signal can be appreciated when the frequency of the
excitation signal changes.

Figure 5. Tests of the functioning of the system in real time.

Within the system, it is important to highlight that a camera temperature controlled is necessary, due
to the temperature generates deviations in the graphics of Magnitude of the signal. A temperature
sensor (LM35) was used for monitoring and controlling the temperature in the measuring camera,
this sensor with an incandescent filament generates a thermic radiation inside the camera. The
voltage supplied to the filament was generated by an electronic dimmer. This dimmer requires a control voltage which is supplied by the FPGA. The programmed PID temperature control, Inside the FPGA was establishing 35°C as the reference point of the controller. The temperature was chosen because this is the average temperature at the samples region and the system is intended to work at the field. In figure 6 is shown the integration of the FPGA, the power electronics and the controlled temperature camera.

![Figure 6. Integration of the controlled temperature camera](image)

The integration of the whole components for the controlled temperature camera are shown in Figure 7. A PID control is running to reach the objective of 35°C inside the camera.

![Figure 7. PID control of the controlled temperature camera](image)

3. Results

3.1 Magnitude and phase results of the frequency response of the sensor

Measurements were carried out using each of the soil samples at a 35°C temperature. Figures 8 and 9 show the frequency response in magnitude and phase obtained by the corresponding sensor to 30% of the total analyzed samples.
By doing a detailed analysis of the obtained studies in the laboratory to determine the physicochemical properties of the soils and comparing them with the magnitude graphics of magnitude and phase obtained with the sensor, it is possible to appreciate that to a higher difference in the physicochemical properties of soils there is also higher differences in the graphics of magnitude and phase of the sensor. Figure 10 shows a matrix of correlation between the variables of magnitude and phase and the physicochemical variables of the soil samples.
Figure 10. Correlation Matrix between the features

The frequency response of the soil itself can neither determine nor quantify the content of physicochemical properties, it is for this reason that the section of artificial intelligence has important relevance to do this determination.

3.2 Repeatability

A required test to know entirely the reliability of the sensor is the parameter of repeatability. 8 sub-samples were separated to carry out this test to each sample of soil, which means that the analysis was effected to each sub-sample of soil, resulting in a mean quadratic error of 4.96%. Figure 11 and figure 12 shows the analysis of the frequency response in magnitude and phase of only 4 samples with their corresponding 8 sub-samples. It is identified that the sensor frequency response with soils that have the same physicochemical conditions (the sub-samples of each sample of soil) present a repeatable behavior. After analyzing the 100% of the sub-samples, a range of error of 5% approximately was presented; which means the repeatability of the sensor is above the 95%.
3.3 Sensor response as function of temperature

The evaluation of the behavior when there are temperature variations is also relevant. Therefore, figures 13 and 14 show how by varying environmental temperature, the measurements have variations mainly in magnitude. It was determined to use the temperature controlled at 35°C due to it is proposed for the present device to be used in-situ in the fields where sugar cane is cultivated, and temperatures are in the rage of 35°C.
Figure 13. Amplitude frequency response as a function of the deepness all the substrates

Figure 14. Phase frequency response as a function of the deepness all the substrates

A sweep frequency from 1 Hz-100 KHz and from 100 KHz to 1 Hz is performed to each soil sample. According to the analysis performed it is noticed that the hysteresis in the frequency 0 to 1 KHz, which is the offset of the sensor presents 8.3% of difference and from 1 KHz to 100 KHz presents a difference of 3.61%.
3.4 Artificial Intelligence for determining physicochemical variables

The current section poses the application of diverse algorithms of artificial intelligence with the purpose of obtaining the validation of a system capable of determining physicochemical properties of soils from the results obtained by the capacitive sensor with spectroscopy.

In the topic of supervised learning and based on the characteristics of the algorithms and the type of data obtained in the present study, the algorithms of Artificial Neural Networks, Logistic Regression, Linear SVM, Gradient Boosting Classifier, Decision Tree, Random Forest, Naive Bayes and Nearest Neighbors were selected. The Scikit-learn library was used in Python to train, analyze and compute in real time all the information acquired by the sensor.

All of these algorithms are used in classification so the data of the training results, the results with the test data, and the time of training were obtained for each one of them with the purpose of validating the performance and the degree of recognition for each algorithm. Figure 15 indicates the procedure to determine the psychochemical properties of soil using the artificial intelligent algorithms. The input data acquired by the sensor is sent via TCP/IP communication protocol to Python in which the artificial intelligence block is programmed to determine according to the input and training data the corresponding output related the psychochemical properties of soil.

![Figure 15. Determining psychochemical properties of soil using IA algorithms and the Sensor data](image)

The following step was focused on building and training a classifier, which can classify the data accurately. For this section, it is necessary to divide the data in two, one section to train and the other to test. The first group of data was used to train the classifier and the validation of the accuracy of the classifier was implemented with the second group. A ratio of 70%/30% is usually implemented to divide the data base due to it contains enough data to accomplish the analysis.

Another important factor is that the distribution of different classes in the data of training and the ones of test should equal the distribution in the current data base. Regularly, an algorithm is used to separate the data randomly for the training group and the validation one.

A total of 100 data of soil samples were used, they correspond to the data of magnitude and phase of the study of soils carried out by the capacitive sensor with spectroscopy. The input array used is an array of 200 elements (corresponding to the magnitude y frequency response of soil from 0 Hz to 100 KHz). The data used in the output, the ones with which the classifier was trained are the physicochemical data obtained at the laboratory; pH, P, K, Ca, Mg, Na, Conductivity and N₂-NO₃.

Figure 16 shows a graphic worth the punctuation with the data of training using the variables of magnitude and phase of the study with spectroscopy and the physicochemical data of the laboratory for: pH, P, K, Ca, Mg, Na, Conductivity and N₂-NO₃.
Figure 16. Train Score

Figure 17 shows a graphic with the punctuation of classification with test data using the variables of magnitude and phase of the study of spectroscopy and the physicochemical data of the laboratory for: pH, P, K, Ca, Mg, Na, Conductivity and N2-NO3.

Figure 17. Test Score comparison for all classifiers

Figure 18 shows a graphic written the punctuation of classification written the data of test using the variables of magnitude and phase of the study of spectroscopy and the physicochemical data of laboratory for: pH, P, K, Ca, Mg, Na, Conductivity and N2-NO3.

Figure 18. Training times (seconds) comparison among classifiers
4. Discussion

The artificial intelligence techniques make use of the most available data possible for a better performance. By using the frequency response of soil, a complex set of information is generated. The information related to correlation data between the magnitude-frequency and the physicochemical properties of soil can be submitted to several A.I. algorithms. Therefore, different techniques of artificial intelligence were used to validate the system with the objective of evaluating which was the most convenient to implement so the system offers an appropriate performance in the task of estimating the physicochemical variables. Through all the experiments developed or was possible to confirm that the use of classifiers such as logistics regression, Nearest Neighbors and Linear SVM were reliable algorithms for the recognition and estimation of the desired variables. The correlation coefficient reached by using the logistics regression and the nearest neighbors is more than 0.8. The obtained results from A.I. algorithms can give a rough estimation about the physicochemical state of the soil under study. Considering the technique used in this work the accuracy obtained is good enough for serving as an indicator of the soil state.

5. Conclusions

The development of a new system capable of estimating some of the most relevant physicochemical properties in soils implementing low frequency spectroscopy techniques and artificial intelligence was presented. This system is a useful alternative and at a low cost for people with little technical experience who could do an analysis of the soil and which could allow the analysis to be in field and make decisions for its restoration. The obtained results indicate the instrument presents good repeatability during the analysis carried out. Similarly, results show that artificial intelligence algorithms such as logistics regression and nearest neighborhood can obtain results with an accuracy above 90% in its recognition. Which makes the instrument appropriate to obtain an adequate estimation of the most relevant physicochemical properties in soils, in a prompt, simple and low-cost means.

For future improvement of the proposed system a cross validation using more than one classifier can be implemented. For serving as universal soil properties, further studies are needed using different soil samples. Even though the presented results show the suitableness of the proposed system as soil physicochemical indicator.

Funding: This research was funded by Consejo Nacional de Ciencia y Tecnología (CONACYT), Sectoral fund of environmental, grant number 26282.

References


10. FAO, “Global Forest Resources Assessment 2015 - How are the world’s forests changing?,” 2015.


