OPEN ACCESS

ORIGINAL ARTICLE

UNAH UNIVERSIDAD AGRARIA DE LA HABANA

Revista Ciencias Técnicas Agropecuarias 34
January-December 2025
ISSN 1010-2760, eISSN 2071-0054
Cu-ID: https://cu-id.com/2177/v34e10

Estimation of Soil Properties by Remote Sensing and Machine Learning

Estimación de propiedades de un suelo mediante teledetección y aprendizaje automático

Roberto Alejandro García-Reyes[I*], María Elena Ruíz-Pérez[II], Sergio Rodríguez-Rodríguez[III] and Juan Alejandro Villazón-Gómez[IV]

[I]*Ministerio de la Agricultura, Departamento Provincial de Suelos y Fertilizantes, Holguín, Cuba.*
[II]*Universidad Agraria de La Habana, Facultad de Ciencias Técnicas, San José de las Lajas, Mayabeque, Cuba.*
[III]*Universidad de Granma, Facultad de Ciencias Agrícolas, Centro de Estudios de Biotecnología Vegetal, Bayamo, Granma, Cuba.*
[IV]*Universidad de Holguín. Departamento de Ciencias Agropecuarias, Holguín, Cuba.*

*Author for correspondence: Roberto Alejandro García-Reyes, e- mail: ralejandro9409@gmail.com

ABSTRACT: Soil suitability for a crop can be determined by quality indicators, which can be obtained by different methods. The objective of this research was to estimate the properties of a vertisol dedicated to rice cultivation by means of remote sensing and automatic learning. Chemical and physical properties of the soil that affect rice physiology were determined “in situ”, with a completely randomized sampling of 100 points at a depth of 0 m to 0.20 m. Properties were estimated from the Normalized Difference Vegetation Index (NDVI) using machine learning tools. Organic matter, calcium, magnesium and phosphorus can be estimated by linear regression with NDVI, considering that they had Standard Errors of 0.27, 3.41, 4.12 and 1.68, respectively, and coefficients of determination close to 1. The random forest technique showed the best performance, with values in its determination close to 1 and an error in its estimation and validation close to 0.
Keywords: Normalized Difference Vegetation Index, Vertisol.

RESUMEN: La aptitud del suelo para un cultivo, se puede determinar mediante indicadores de calidad, que pueden ser obtenidos por diferentes métodos. La investigación tuvo como objetivo la estimación de propiedades de un Vertisuelo dedicado al cultivo del arroz mediante teledetección y aprendizaje automático. Se determinaron “*in situ*” propiedades químicas y físicas del suelo que inciden en la fisiología del arroz, con un muestreo completamente aleatorizado de 100 puntos a una profundidad de 0 m a 0,20 m. Las propiedades se estimaron a partir del Índice de Vegetación por Diferencia Normalizada (NDVI) mediante herramientas de aprendizaje automático. La materia orgánica, el calcio, el magnesio y el fósforo, se pueden estimar mediante regresión lineal con el NDVI, teniendo en cuenta que tuvieron Errores Estándares de 0,27; 3,41; 4,12 y 1,68, respectivamente, y los coeficientes de determinación cercanos a 1. La técnica de bosques aleatorios mostró el mejor rendimiento, con valores en su determinación cercanos a 1 y un error en su estimación y validación cercanos a 0.
Palabras clave: índice de vegetación por diferencia normalizada, vertisol.

INTRODUCTION

Rice cultivation is in high demand for world food and its production reaches more than 700 million tons. Cuba is one of the nations with a high consumption of this cereal, which amounts to 80.38 kg person^{-1} year^{-1} (Del Valle *et al.*, 2022). In the country, a total of 16 847 hectares are dedicated to the crop for a production of 266 596 t. The production of the crop is 3.35 t ha^{-1} and is found in productions of state enterprises, cooperatives and the private sector (Casanovas *et al.*, 2022).

In Holguín province, Cuba, due to the construction of the East-West water transfer, sufficient water is available for rice (*Oryza sativa* L.) cultivation. In this context, it is important to consider that the growth of rice plants depends on the physical and chemical conditions of the soil, which affect the capacity of the crop's root system to grow efficiently. Agricultural operations such as land preparation for cultivation, tillage, fertilization, irrigation management and planting methods alter soil properties in the short and long term, impacting the sustainability and yield of the crop (Baroudy *et al.*, 2020).

Received: 31/07/2024
Accepted: 24/02/2025
The authors of this work declare no conflict of interests.
AUTHOR'S CONTRIBUTIONS: Conceptualization: Roberto Alejandro García-Reyes. **Data curation:** Roberto Alejandro García-Reyes, María Elena Ruíz Pérez and Sergio Rodríguez Rodríguez. **Investigation:** Roberto Alejandro García-Reyes, María Elena Ruíz Pérez, Juan Alejandro Villazón Gómez. **Methodology:** Roberto Alejandro García-Reyes, María Elena Ruíz Pérez and Sergio Rodríguez Rodríguez. **Supervision:** María Elena Ruíz Pérez and Sergio Rodríguez Rodríguez. **Validation:** María Elena Ruíz Pérez and Sergio Rodríguez Rodríguez Rodríguez, Juan Alejandro Villazón Gómez. **Papers/Editing, original project:** Roberto Alejandro García-Reyes. **Writing, proofreading and editing:** Roberto Alejandro García-Reyes, María Elena Ruíz Pérez and Sergio Rodríguez Rodríguez, Juan Alejandro Villazón Gómez.

Among the most widely used methods to evaluate the condition of soils dedicated to this crop is the one proposed by Baroudy *et al.* (2020), where satellite images are used to determine soil and crop spectral indices, which are correlated with in situ information. Singh *et al.* (2024) also propose machine learning algorithms based on mathematical models to estimate properties using in situ and remote sensing information. Both investigations have used different regression methods that can be used to build models to estimate properties at a lower cost and provide rapid information over time (Siqueira *et al.*, 2024).

The models derived from the use of machine learning most commonly used in the literature, highlight neural networks, random forests and regression vectors among others (Ließ *et al.*, 2016). However, in Cuba, no studies have been published using these tools to evaluate the suitability of soils for rice cultivation. Hence the reasons that make possible the use of spatial remote sensing Luque (2023), such as its ability to discriminate large areas that have different characteristics in their physical-chemical composition of the surface exposed on the ground. Therefore, the objective of this research is to estimate the properties of a soil dedicated to rice cultivation using remote sensing and remote sensing.

MATERIALS AND METHODS

The selected area belongs to the Agriculture Enterprise Guatemala, CCS "Tomás Machado" of Cosme Herrera village, located at 20°44'54.601"N and 75°50'43.743"W of Mayarí municipality, Holguín province (Figure 1).

In the study area, according to data from the Guaro Meteorological Station, located at 20º40'21 "N and 75º46'57" W in the municipality of Mayarí at 20.96 masl, the annual precipitation is 1 067.6 mm and the average temperature is 25.6 °C according to studies carried out by Villazón *et al.* (2023).

Regarding meteorological data from the beginning of April until May 26, 2022, the date of sampling, the total precipitation was 168.7 mm, with an average temperature of 25.4 °C and an average relative humidity of 73.8 %. The characteristic soil of the area is of the Chromic Vertisol type

FIGURE 1. Location of the area where the research was carried out, belonging to the CCS "Tomás Machado", Cosme Herrera village, Mayarí, Holguín, Cuba. Source: SAS PlanetNightly 200718.10081 (http://www.geojamal.com)

Hernández *et al.* (2015) with a slope < 2 % so it can be considered flat. In the 100 ha area, a systematic sampling was performed in 100 points georeferenced with a GPS with 3 m appreciation, at a distance between points of 100 m.

The samples were taken with an auger for agrochemical analysis in the depth range between 0 m to 0.20 m because this is the depth where the highest content of rice roots is found, capable of absorbing water and the nutrients necessary for their growth and development (Angladette, 1969).

The soil properties selected for this work are shown in Table 1. The methods followed for their selection are given in García *et al.* (2025). All properties were determined according to the Cuban standards for the determination of chemical properties of soils in the National Soil Laboratory network of the country.

For the estimation of the properties selected, the Normalized Difference Vegetation Index (NDVI) of the study area was determined from the image of April 26, 2022, belonging to the Landsat 9 OLI/TIRS 2 satellite (LC09_L2SP_011046_20220426_20220428_02_T1) of the United States Geological Survey at path 011 row 046 and were projected in the WGS 84 UTM Zone 18 North System in QGIS 3 software. 10 "A Coruña".

TABLE 1. Determination of the most influential soil properties on soil quality for rice cultivation

Name of the soil property	Properties analyzed	Unit of measurement	Analytical technique used
pH in water	pH_{H2O}	unit	(NC 2001.2015)
Assimilable phosphorus	P_2O_5	mg kg^{-1}	(NC 52.1999)
Assimilable potassium	K_2O	mg kg^{-1}	
Total nitrogen	Nt	%	(NC 11261: 2009)
Organic matter	MO	%	(NC 1043.2014)
Calcium	Ca	cmol kg^{-1}	(NC 209:2002)
Assimilable magnesium	Mg	cmol kg^{-1}	(NC 209:2002)
Assimilable sodium	Na	cmol kg^{-1}	
Electrical conductivity	CE	dS m $^{-1}$	(NC 776: 2010)

NC: Cuban standard

For the determination of the NDVI, Equation 1 according to Rouse Jr *et al.* (1974), was used, after performing the atmospheric correction to eliminate the effect of clouds on the image.

$$NDVI \quad \frac{(B_{NIR} - B_{red})}{(B_{NIR} + B_{red})} \qquad (1)$$

where: B_{NIR} is the infrared band and B_{red} is the red band of the sensor.

Given the objective of the research, an approach proposed by Choudhury & Mandal, (2021) is used, which consists of building models to estimate soil properties from NDVI maps.

Machine learning tools were used to estimate soil properties, which were: simple linear regression Khanal *et al.* (2018), Random Forest (RF) Park *et al.* (2024), support vector machine for regression (SVM) Shrestha & Shukla (2015), stochastic gradient Decent (SGD) Nisbet *et al*. (2009) and k-nearest neighbors (kNN) (Taghizadeh *et al.*, 2022).

To evaluate the effectiveness of the capabilities of the regression and machine learning models used to predict soil properties from NDVI, the following statistics were determined: Root Mean Square Error (RMSE) Montgomery *et al.* (2003) Mean Absolute Error (MAE) Panigrahi *et al.* (2023), the coefficient of determination (R^2) (Gatera et al., 2023), correlation coefficient (r) Montgomery *et al.* (2003) and the Durbin-Watson (DW) statistic (de Smith et al., 2013). In all cases, 70 % of the data was used to perform the estimation and the remaining 30 % to validate the model according to the methodology proposed by Whetton *et al.* (2017).

A hierarchical classification was performed based on the performance metrics of the machine learning models (random forests, regression vector, k-nearest neighbor and stochastic gradient) using the Nash-Sutcliffe efficiency index (EF) proposed by Nash & Sutcliffe (1970) and the concordance index according to Willmott (1982).

RESULTS AND DISCUSSION

The descriptive analysis of NDVI calculated from a Landsat 9 satellite image provides a complete overview of the data, where the average NDVI value found was 0.26 as shown in (Table 2).

It is important to note that the NDVI scale ranges from -1 to 1 and, in this particular case, the observed values are within this range. This agrees with Rawashdeh (2012) proposition that, for NDVI, values between 0 and 0.5 indicate a limited presence of vegetation, in accordance with the current conditions of the study area. These insights shed light on the vegetation landscape and contribute to a deeper understanding of the research results.

Table 3 shows the linear regression analysis statistics of the models generated between NDVI and soil properties used as a quality index. With the use of the simple regression technique between NDVI values and soil properties used as quality indicator, it can be evidenced that there is a high correlation of 0.98 between soil organic matter content and NDVI.

The coefficient of determination was 0.94, so it can be affirmed that the NDVI can predict the organic matter content, with an error in its prediction in all cases within the permissible ranges in which the determined variables are measured Ayoubi *et al.* (2011), both NDVI with values that can range from -1 to 1 and organic matter with maximum values of 6.5 %.

The results highlight remarkable positive associations between magnesium concentration present in the soil and NDVI, which stands at 0.93 % correlation coefficient. These results align with the results reported by Mazur *et al.* (2022), who also observed a correlation of 0.95 between magnesium content and NDVI in a soil specifically intended for cereal cultivation.

TABLE 2. Descriptive analysis of NDVI in the study area

Spectral Index	Mean	SD	SE	CV (%)	Minimum	Maximum	Median
NDVI	0,26	0,06	0,01	2,74	0,11	0,43	0,25

SD: Standard deviation; SE: standard error; CV: variation coefficient

TABLE 3. Linear regression analysis statisticians

Linear regression models	R^2	r^2	SE	MAE	DW
NDVI vs. MO (%)	0,94	0,98	0,27	0,20	0,33
NDVI vs. Mg (cmol kg^{-1})	0,88	0,93	4,12	3,55	0,10
NDVI vs. Ca (cmol kg^{-1})	0,90	0,95	3,41	2,84	0,14
NDVI vs. P_2O_5 (mg kg^{-1})	0,90	0,82	1,68	1,33	0,13
NDVI vs. Nt (%)	0,84	0,70	0,01	0,01	0,12
NDVI vs K_2O (mg kg^{-1})	0,31	0,56	12,26	9,67	0,23
NDVI vs. CE (dS m^{-1})	0,58	0,76	0,21	0,17	0,11
NVDI vs Na (cmol kg-1)	0,47	0,63	0,11	0,27	0,16
NVDI vs pH	0,34	0,31	0,04	0,17	0,15

R^2: coefficient of determination; r^2: correlation coefficient; SE: standard error; MAE: Mean Absolute Error; DW: Durwin-Watson.

The relationship between NDVI and calcium content of 0.94 is significantly associated with the findings presented by Abdalkarim *et al.* (2023), in which he suggests that there is a simultaneous occurrence of an increase in calcium content in soils and a reduction in vegetation cover.

In the simple linear regression analysis between NDVI and potassium. The regression coefficient indicates that the fitted model explains 0.31 of the potassium variability. The correlation coefficient is equal to 0.56 which shows a moderate relationship between the variables. The standard error of the estimation shows that the standard deviation of the residuals is 12.26 mg kg^{-1} which can be used to construct prediction limits. The mean absolute error is 9.67 mg kg^{-1} which is the average value of the residuals.

In all the simple linear regression models it is observed that the Durbin-Watson (DW) statistic values range from 0.06 to 0.33 which indicates that there is a positive autocorrelation between the residuals. These values of the DW statistic refer that the spatial autocorrelation that exists is due to the fact that the properties have a tendency to be clustered with areas where spatial behavior tends to be the main source of errors.

This research suggests that when low values of the coefficient of determination (0.50) are obtained it does not mean that the models are of poor quality; rather it points to the presence of a number of essential factors, not taken into account by the model and qualitative characteristics that are difficult to determine from Landsat satellite data (Gopp *et al.*, 2019).

Not all models provided by linear regression analysis are adequate, because of the lack of a robust model structure for the above-mentioned properties, to make unbiased inferences regarding the functional dependence between NDVI and soil properties. The linear regression model is the most widely used approach to estimate soil properties with the use of remotely sensed derived data (Vergopolan *et al.*, 2021). However, it has limitations in handling nonlinear relationships between response variables and predictors that generally exist across different agricultural land uses (Khanal *et al.*, 2018).

As it was seen not all relationships between NVDI and soil properties that result in quality indicators satisfy a linear relationship, therefore, it becomes necessary to explore other types of machine learning models. The parameters of calibration and validation of the models yielded by the random forest tool are shown in Table 4. It was observed that the RF model estimated, adequately, the potassium content by presenting low values of RMSE (2.34) and high values of R^2 (0.98).

The RF model performed well during the validation process, especially in the estimation of pH, supported by the R^2 value (0.93), which indicates a good fit between the measured and estimated data, and explained 93% of the variability in the data. Therefore, it can be assumed that RF is reliable and accurate for estimating this variable. Unlike the model that yields simple linear regression for estimating total nitrogen, this property is not predicted. In the case of using RF, it determines 75% of the total variation of NDVI values, which affirms a high determination of this property with errors close to 0.

Several studies have confirmed that the RF model predicts soil properties significantly better than linear regression methods. Studies illustrate the use of the original sensor bands and the determination of spectral indices to estimate soil properties (Zhang *et al.*, 2018). As in this research, there are reference of works with NDVI values in bare soils which have estimated soil properties with RF models (Jiang *et al.*, 2018).

The SVM model, shown in Table 5, has a varied performance, with some very positive and some extremely negative results. In the case of Na it shows an R^2 of 0.79 in the validation, suggesting that SVM is quite effective for this variable, capturing most of the variability in the data. However, for Nt it presents an R^2 of 0.41 in the validation, a very low value according to the range in which this type of coefficient should perform best.

The kNN proves to be a robust model according to Table 6, with consistent performance in most of the variables. For K2O, an R^2 of 0.84 is obtained in the validation, indicating an excellent predictive capacity. Meanwhile, for pH it shows an R^2 of 0.67, indicating that the model is adequate, although not as accurate as RF for this variable.

TABLE 4. Calibration and validation of models generated by random forests (RF)

Models	Calibration			Validation		
	RMSE	MAE	R^2	RMSE	MAE	R^2
RF_Nt	0,02	0,01	0,38	0,01	0,01	0,75
RF_P_2O_5	3,41	2,02	0,35	2,08	1,17	0,76
RF_ K_2O	6,56	3,11	0,84	2,34	1,67	0,98
RF_Ca	10,32	7,92	0,15	7,94	5,97	0,48
RF_Mg	11,96	8,98	0,01	9,12	6,96	0,39
RF_MO	0,97	0,77	0,45	0,73	0,57	0,69
RF_CE	0,31	0,13	0,33	0,20	0,07	0,73
RF_Na	0,27	0,17	0,41	0,15	0,11	0,80
RF_pH	0,20	0,09	0,39	0,06	0,04	0,93

RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; R^2: Coefficient of Determination

TABLE 5. Calibration and validation of Support Vector Machine (SVM) generated models

Models	Calibration			Validation		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SVM_Nt	0,03	0,03	0,72	0,03	0,03	0,41
SVM_P_2O_5	3,61	2,64	0,42	3,2	2,16	0,53
SVM_K_2O	15,19	9,57	0,12	13,65	8,53	0,28
SVM_Ca	10,6	8,67	0,11	9,49	7,62	0,26
SVM_Mg	11,86	9,17	0,01	11,07	8,35	0,1
SVM_MO	0,99	0,82	0,42	0,89	0,71	0,53
SVM_CE	0,29	0,18	0,42	0,25	0,15	0,56
SVM_Na	0,22	0,15	0,6	0,16	0,12	0,79
SVM_pH	0,2	0,12	0,37	0,16	0,09	0,56

RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; R^2: Coefficient of Determination

TABLE 6. Calibration and validation of the models generated by k-nearest neighbors (kNN)

Models	Calibration			Validation		
	RMSE	MAE	R^2	RMSE	MAE	R^2
kNN_Nt	0,01	0,01	0,57	0,01	0,01	0,69
kNN_P_2O_5	2,89	1,97	0,53	2,15	1,37	0,74
kNN_K_2O	5,38	3,29	0,89	3,67	2,18	0,95
kNN_Ca	9,09	7,01	0,34	6,8	5,03	0,62
kNN_Mg	10,05	7,48	0,29	8,21	5,98	0,51
kNN_MO	0,8	0,63	0,62	0,65	0,5	0,75
kNN_CE	0,27	0,15	0,48	0,18	0,1	0.77
kNN_Na	0,24	0,15	0,55	0,17	0,11	0,75
kNN_pH	0,19	0,1	0,4	0,14	0,06	0,67

RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; R^2: Coefficient of Determination

The results of the models derived from the SGD are presented in Table 7. According to the values obtained from the statisticians evaluated, the results obtained are moderate in all the variables, without standing out in any one in particular. The MO during validation achieves an R^2 of 0.13, which suggests a limited performance, while the K2O achieves an R^2 of 0.41, which indicates that the model can capture some relationships, but is outperformed by RF and kNN.

In relation to the results obtained, Forkuor *et al.* (2017) posit that obtaining low R^2 values can generally be attributed to a complex interaction and high variability of environmental factors and high variability in agricultural practices such as soil management, nutrient application and vegetation cover. In agreement with the opinion of Chai y Draxler (2014), RMSE is unlikely to provide a robust evaluation of the models. It is especially worth noting that the efficiency and concordance index can provide more effective information on model performance. These differences in estimators for evaluation are due to the fact that between observed and estimated values when squared, they may overestimate larger values in the estimated data while smaller values may be neglected (Willmott, 1982).

Table 8 presents the efficiency and agreement indices of each soil property estimate and quality indicator from the machine learning models (random forests, regression vector, k-nearest neighbor and stochastic gradient). It is evident that the RF model demonstrates the highest level of efficiency, as indicated by the prevalence of values consistently ranging from 0.77 to 1.0 for the efficiency index, along with an equally impressive range of values from 0.93 to 1.0 for the concordance index. In a broader interpretation, this suggests that the RF model possesses a commendable ability to make accurate predictions, particularly in the context of aligning the values of various soil properties with observed average conditions, while the high degree of agreement means that the predictive results are remarkably congruent with actual observed data in real scenarios, presenting a very favorable outlook for the application of machine learning methodologies in this domain.

Following the RF model in the hierarchical ranking based on the efficiency and concordance index performance metrics, the k-NN model also shows favorable performance, yielding results that fall within a range of 0.78 to 1.0 for the efficiency index, and a high validity concordance index, ranging from 0.92 to 1.0, indicating its effectiveness in making reliable predictions. In this context of comparative performance metrics, it is noteworthy that the SGD model shows superior results when juxtaposed with the SVM model, highlighting the relative effectiveness of these different predictive modeling approaches in analyzing soil property data.

TABLE 7. Calibration and validation of stochastic gradient descent (SGD) generated models

Models	Calibration			Validation		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SGD_SQI	0,15	0,11	0,26	0,13	0,11	0,38
SGD_Nt	0,02	0,01	0,14	0,02	0,01	0,23
SGD_P_2O_5	4,28	3,64	0,01	4,01	3,44	0,09
SGD_K_2O	13,39	11,31	0,61	12,43	10,57	0,73
SGD_Ca	11,7	9,46	0,09	10,82	8,87	0,03
SGD_Mg	12,2	10,02	0,05	11,37	9,52	0,06
SGD_MO	1,26	1,04	0,06	1,21	0,99	0.13
SGD_CE	0,36	0,29	0,08	0,33	0,27	0,22
SGD _Na	0,33	0,24	0,09	0,29	0,22	0,28
SGD_pH	0,21	0,16	0,27	0,19	0,14	0,35

RMSE: Root Mean Squared Error; MAE: Mean Absolute Error; R^2: Coefficient of Determination

TABLA 8. Efficiency and concordance index in soil property estimation

Soil properties	Efficiency index				Concordance index			
	RF	SVM	KNN	SGD	RF	SVM	KNN	SGD
Nt	0,98	0,21	0,98	0,94	1,00	0,69	1,00	0,99
P	0,97	0,70	0,91	0,88	0,99	0,99	0,95	0,98
K	1,00	0,41	0,99	0,39	1,00	0,92	1,00	0,95
Ca	0,77	0,68	0,84	0,52	0,97	0,15	0,92	0,43
Mg	0,60	0,02	0,57	0,76	0,93	0,91	0,94	0,88
Na	0,99	0,99	0,99	0,97	1,00	1,00	1,00	0,99
MO	0,86	0,68	0,78	0,79	0,98	0,96	0,80	0,94
CE	0,44	0,83	0,68	0,22	0,98	0,95	0,97	0,86
pH	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Ko *et al.* (2024) state that a high Efficiency Index suggests that the model captures well the variability of the observed data, being higher the values in this study than those reported by this research in the soybean crop. They show that the efficiency of machine learning models may be further influenced by geographic area or soil type.

Dharumarajan *et al.* (2017) obtained in the prediction of soil organic carbon and pH from vegetation spectral indices obtained from satellite images, Concordance Index values of 0.37 and 0.38 respectively, and only the model yielded in the prediction of electrical conductivity was acceptable with 0.70 for this Index, which in this study was higher with 0.86.

CONCLUSIONS

It is possible to estimate organic matter, magnesium, calcium and phosphorus from NVDI with simple linear regression given that when using the Durbin-Watson statistic the values range from 0.06 to 0.33 which indicates that there is a positive autocorrelation between the residuals, existing properties having a tendency to be clustered with areas where spatial behavior is the main source of errors.

The random forest model is the most adequate for estimating potassium and pH, supported by values of R^2 (close to 1), RMSE (close to 0) and the Efficiency and Concordance Indices (close to 1).

The remaining algorithms achieved best fit in the following decreasing order according to the values of the calculated efficiency and concordance indexes, kNN, SGD and SVM.

REFERENCES

ABDALKARIM, K.; GAZNAYEE, H.A.A.; AL-QURAISHI, A.M.F.; ABDALLA, Z.O.: "Predictive Digital Mapping of Surface Soil Properties using Remote Sensing and Multivariate Statistical Analysis.", *Zanco Journal of Pure and Applied Sciences*, 35(6): 189-203, 2023, ISSN: 2412-3986.

ANGLADETTE, A.: *El arroz. Agricultura Tropical, Colección Agricultura Tropical*, Ed. Editorial Blume, 867 p., 1969, ISBN: 84-7313-835-X.

AYOUBI, S.; SHAHRI, A.P.; KARCHEGANI, P.M.; SAHRAWAT, K.L.: "Application of artificial neural network (ANN) to predict soil organic matter using remote sensing data in two ecosystems", *Biomass and remote sensing of biomass*, 10: 181-196, 2011.

BAROUDY, A.A.L.; ALI, A.M.; MOHAMED, E.S.; MOGHANM, F.S.; SHOKR, M.S.; SAVIN, I.; PODDUBSKY, A.M.; DING, Z.; KHEIR, A.; ALDOSARI, A.A.: "Modeling land suitability for rice crop using remote sensing and soil quality indicators: The case study of the nile delta", *Sustainability*, 12(22): 9653, 2020, ISSN: 2071-1050.

CASANOVAS, C.E.; SUÁREZ DEL VILLAR, L.E.; ÁLVAREZ, S.A.; AVILLEIRA, C.I.: "Valoración de la seguridad alimentaria cubana a partir de la superficie agrícola explotada y los rendimientos agrícolas", *Revista Universidad y Sociedad*, 14(5): 304-314, 2022, ISSN: 2218-3620.

CHAI, T.; DRAXLER, R.R.: "Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature", *Geoscientific model development*, 7(3): 1247-1250, 2014, ISSN: 1991-9603.

CHOUDHURY, B.U.; MANDAL, S.: "Indexing soil properties through constructing minimum datasets for soil quality assessment of surface and profile soils of intermontane valley (Barak, North East India)", *Ecological Indicators*, 123: 107369, 2021, ISSN: 1470-160X.

DE SMITH, M.; GOODCHILD, M.; LONGLEY, P.: *Geoespacial analysis. A comprehensive guide to principles, techniques and software tolos*, Ed. The Winchelsea Press, Winchelsea, UK, Winchelsea, UK, 2013.

DEL VALLE, M.J.; GONZÁLEZ, V.; RAFAEL, P.L.; SÁNCHEZ, A.O.R.; DELGADO, T.C.: "Efecto de las variables climáticas sobre el rendimiento agrícola del arroz (Oryza sativa L.)", *Ingeniería Agrícola*, 12(1), 2022, ISSN: 2227-8761.

DHARUMARAJAN, S.; HEGDE, R.; SINGH, S.: "Spatial prediction of major soil properties using Random Forest techniques-A case study in semi-arid tropics of South India, Geoderma Reg., 10, 154-162", 2017.

FORKUOR, G.; HOUNKPATIN, O.K.; WELP, G.; THIEL, M.: "High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models", *PloS one*, 12(1): e0170478, 2017, ISSN: 1932-6203.

GARCÍA, R.R.A.; RUÍZ, P.M.E.; SERGIO, R.R.: "Indicador de calidad de un Vertisol dedicado al arroz en la provincia Holguín, Cuba", *Revista Ciencias Técnicas Agropecuarias*, 34, 2025.

GATERA, A.; KURADUSENGE, M.; BAJPAI, G.; MIKEKA, C.; SHRIVASTAVA, S.: "Comparison of random forest and support vector machine regression models for forecasting road accidents", *Scientific African*, 21: e01739, 2023, ISSN: 2468-2276.

GOPP, N.; SAVENKOV, O.; NECHAEVA, T.; SMIRNOVA, N.; SMIRNOV, A.: "Application of NDVI in digital mapping of phosphorus content in soils and phosphorus supply assessment in plants", *Izvestiya, Atmospheric and Oceanic Physics*, 55: 1322-1328, 2019, ISSN: 0001-4338.

HERNÁNDEZ, J.A.; PÉREZ, P.; BOSCH, I.D.; CASTRO, S.N.: *Clasificación de los suelos de Cuba.*, Ed. Instituto Nacional de Ciencias Agrícolas, Instituto Nacional de Ciencias Agrícolas, ed., San José de las Lajas, Mayabeque, Cuba, 91 p., 2015, ISBN: 959-7023-77-6.

JIANG, Y.; RAO, L.; SUN, K.; HAN, Y.; GUO, X.: "Spatio-temporal distribution of soil nitrogen in Poyang lake ecological economic zone (South-China)", *Science of the total environment*, 626: 235-243, 2018, ISSN: 0048-9697.

KHANAL, S.; FULTON, J.; KLOPFENSTEIN, A.; DOURIDAS, N.; SHEARER, S.: "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield", *Computers and electronics in agriculture*, 153: 213-225, 2018, ISSN: 0168-1699.

KO, J.; SHIN, T.; KANG, J.; BAEK, J.; SANG, W.G.: "Combining machine learning and remote sensing-integrated crop modeling for rice and soybean crop simulation", *Frontiers in Plant Science*, 15: 1320969, 2024, ISSN: 1664-462X.

LIESS, M.; SCHMIDT, J.; GLASER, B.: "Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches", *PLoS One*, 11(4): e0153673, 2016, ISSN: 1932-6203.

LUQUE, R.L.: "Revisión Sistemática de Literatura de Imágenes Satelitales en Hidrología y Agricultura", *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E55): 264-278, 2023, ISSN: 1646-9895.

MAZUR, P.; GOZDOWSKI, D.; WNUK, A.: "Relationships between soil electrical conductivity and sentinel-2-derived NDVI with pH and content of selected nutrients", *Agronomy*, 12(2): 354, 2022, ISSN: 2073-4395.

MONTGOMERY, D.C.; SKINNER, K.R.; RUNGER, G.C.: "Process monitoring for multiple count data using generalized linear model-based control charts", *International Journal of Production Research*, 41(6): 1167-1180, 2003, ISSN: 0020-7543.

NASH, J.E.; SUTCLIFFE, J.V.: "River flow forecasting through conceptual models part I-A discussion of principles", *Journal of hydrology*, 10(3): 282-290, 1970, ISSN: 0022-1694.

NORMA CUBANA (NC): "*Calidad del suelo-determinación de los componentes orgánicos",* NC: 1043.2014, Oficina Nacional de Normalización, Cuba. 2014.

NORMA CUBANA (NC): "*Calidad del suelo-determinación de los componentes orgánicos",* NC: 1043.2014, Oficina Nacional de Normalización, Cuba. 2014.

NORMA CUBANA (NC): "*Determinación de las formas móviles de Fósforo y Potasio",*. NC: 52.1999, Oficina Nacional de Normalización, Cuba. 1999.

NORMA CUBANA (NC): "*Calidad del Suelo. Determinación de la capacidad de intercambio catiónico y de los cationes intercambiables del suelo*", NC: 209.2002, Oficina Nacional de Normalización, Cuba. 2002.

NORMA CUBANA (NC): "*Calidad del Suelo. Determinación del Nitrógeno total Método* Kjeldahl", NC: 11261.2009, Oficina Nacional de Normalización, Cuba. 2009.

NORMA CUBANA (NC): "*Calidad del Suelo. Evaluación de la afectación por salinidad*", NC: 776.2010, Oficina Nacional de Normalización, Cuba. 2010.

NORMA CUBANA (NC): "*Calidad del suelo-determinación de pH*". NC: 10390.2014, Oficina Nacional de Normalización, Cuba. 2014.

PANIGRAHI, B.; KATHALA, K.C.R.; SUJATHA, M.: "A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models", *Procedia Computer Science*, 218: 2684-2693, 2023, ISSN: 1877-0509.

PARK, H.J.; BAEK, N.; SEO, B.S.; JEONG, Y.J.; YANG, H.I.; LEE, S.I.; YOON, K.S.; KIM, H.Y.; CHOI, W.J.: "Estimation of the electrical conductivity of saturated paste from soil-water extracts of coastal saline paddy soils using random forest and multiple regression models", *Journal of Soils and Sediments*, 24(3): 1250-1259, 2024, ISSN: 1439-0108.

ROUSE JR, J.W.; HAAS, R.H.; DEERING, D.; SCHELL, J.; HARLAN, J.C.: *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*, 1974.

SHRESTHA, N.; SHUKLA, S.: "Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment", *Agricultural and forest meteorology*, 200: 172-184, 2015, ISSN: 0168-1923.

SINGH, G.; SINGH, J.; WANI, O.A.; EGBUERI, J.; AGBASI, J.C.: "Assessment of groundwater suitability for sustainable irrigation: a comprehensive study using indexical, statistical, and machine learning approaches", *Groundwater for Sustainable Development*, 24: 101059, 2024, ISSN: 2352-801X.

SIQUEIRA, R.G.; MOQUEDACE, C.M.; FERNANDES-FILHO, E.I.; SCHAEFER, C.E.R.; FRANCELINO, M.R.; SACRAMENTO, I.F.; MICHEL, R.F.: "Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica", *Catena*, 235: 107677, 2024, ISSN: 0341-8162.

TAGHIZADEH, M.R.; KHADEMI, H.; KHAYAMIM, F.; ZERAATPISHEH, M.; HEUNG, B.; SCHOLTEN, T.: "A comparison of model averaging techniques to predict the spatial distribution of soil properties", *Remote S*, 14(3): 472, 2022, ISSN: 2072-4292.

VERGOPOLAN, N.; CHANEY, N.W.; PAN, M.; SHEFFIELD, J.; BECK, H.E.; FERGUSON, C.; TORRES-, R.L.; SADRI, S.; WOOD, E.F.: "SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US", *Scientific data*, 8(1): 264, 2021, ISSN: 2052-4463.

VILLAZÓN, J.A.; NORIS, P.; GARCÍA, R.A.; CRUZ, M.: "Análisis temporal de la agresividad y concentración de las precipitaciones en áreas agropecuarias de la provincia de Holguín, Cuba", *Idesia (Arica)*, 41(3): 77-86, 2023.

WHETTON, R.; ZHAO, Y.; SHADDAD, S.; MOUAZEN, A.M.: "Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI", *Computers and electronics in agriculture*, 138: 127-136, 2017, ISSN: 0168-1699.

WILLMOTT, C.J.: "Some comments on the evaluation of model performance", *Bulletin of the American Meteorological Society*, 63(11): 1309-1313, 1982, ISSN: 0003-0007.

ZHANG, Y.; SUI, B.; SHEN, H.; WANG, Z.: "Estimating temporal changes in soil pH in the black soil region of Northeast China using remote sensing", *Computers and Electronics in Agriculture*, 154: 204-212, 2018, ISSN: 0168-1699.

Roberto Alejandro García-Reyes, Ing., Inv., Ministerio de la Agricultura, Departamento Provincial de Suelos y Fertilizantes, provincia Holguín, Cuba. e-mail: ralejandro9409@gmail.com.
María Elena Ruiz-Pérez, Dr.C., Profesora Titular, Universidad Agraria de La Habana "Fructuoso Rodríguez Pérez". Carretera Tapaste y Autopista Nacional km 231/2, San José de Las Lajas, Mayabeque, Cuba. CP 32700. e-mail: mruiz@unah.edu.cu.
Sergio Rodríguez-Rodríguez, Dr.C., Universidad de Granma, Facultad de Ciencias Agrícolas, Centro de Estudios de Biotecnología Vegetal, Bayamo, Granma, Cuba. e-mail: ralejandro9409@gmail.com
Juan Alejandro Villazón-Gómez, MSc., Profesor, Universidad de Holguín. Departamento de Ciencias Agropecuarias, Holguín, Cuba. e-mail: ralejandro9409@gmail.com