



# Computational Model to Predict Soil Density Using Machine Learning Methods

## *Modelo computacional para la estimación de la densidad del suelo a través del sensoramiento continuo*

Dr.C. Darina Lara Coba<sup>I</sup>; Dr.C. Miguel Herrera Suárez<sup>II</sup>; Dr.C. María Matilde García Lorenzo<sup>III</sup>;  
M.Sc. Roberto Beltran<sup>IV</sup>

<sup>I</sup> Universidad Tecnológica de la Habana José Antonio Echeverría, Facultad de Ingeniería Mecánica,  
Departamento de Mecánica Aplicada, Marianao, La Habana, Cuba.

<sup>II</sup> Universidad Técnica de Manabí (UTM), Facultad de Ciencias Matemáticas Físicas y Químicas, Escuela de Mecánica,  
Portoviejo, Manabí, Ecuador.

<sup>III</sup> Universidad Central “Marta Abreu” de Las Villas, Grupo de Inteligencia Artificial,  
Centro de Investigaciones Informáticas, Santa Clara, Cuba.

<sup>IV</sup> Universidad de las Fuerzas Armadas, Dpto. Energía y Mecánica., Extensión Latacunga, Ecuador.

**ABSTRACT.** The machine learning methods have been used successfully in the calculation of parameters of various problems of engineering, in which the complicated variables have a relation nonlinear among themselves and the modelation does not enable representing the intervening problem through a mathematical function of easy deduction. For the estimation of soil properties several variables are analyzed that make their estimation by means of mathematical models is a complex process transferring the problem solution to artificial intelligence field. The present work aims at developing a mathematical model for the estimation of soil density through the on-the-go soil sensing, a method of automatized learning. The computational learning automated tool used was WEKA, by means of which three procedures of automatized learning applied (multilayer perceptron neuronal artificial nets and K-nearest neighbor). The validation of the model came true by means of the crossed and experimental validation. Results evidence that the best method is the K-nearest neighbor with absolute mean error of 0.06 and a correlation coefficient of 0.89; variables of bigger weight in prediction were moisture content followed by work speed, power, width of the working tool and the depth.

**Keywords:** Bulk density, artificial intelligence, soil compaction, prediction of soil compaction

**RESUMEN.** Los métodos de aprendizaje automatizado han sido usados con éxito en el cálculo de parámetros de diversos problemas de ingeniería, en que las variables involucradas tienen una relación no lineal entre sí y la modelación no permite representar el problema mediante una función matemática de fácil deducción. Para la estimación de las propiedades del suelo se involucran diversas variables, que hacen que su estimación por medio de modelos matemáticos sea un proceso complejo trasladando la solución del problema al campo de la inteligencia artificial. El presente trabajo tiene como objetivo desarrollar un modelo matemático para la estimación de la densidad del suelo a través del sensoramiento continuo utilizando un método de aprendizaje automatizado. La herramienta computacional de aprendizaje automatizado que se utilizó fue WEKA, mediante la cual se aplicaron tres métodos de predicción numérica (redes neuronales artificiales del tipo multicapa, K de los vecinos más cercanos y regresión lineal). La validación del modelo se realizó mediante la validación cruzada y experimental. Los resultados muestran que el mejor método es el K de los vecinos más cercanos con un error medio absoluto de 0,06 y un coeficiente de correlación de 0,89; las variables de mayor peso en la predicción fueron: el contenido de humedad seguido de la velocidad de trabajo, la fuerza, ancho de la herramienta de trabajo y la profundidad.

**Palabras clave:** densidad volumétrica, inteligencia artificial, compactación del suelo, estimación de la densidad del suelo.

## INTRODUCTION

Soil density can be considered one of the most important factors that indicate soil compaction. Soil compaction is decreased in volume by applying a mechanical load, this causes changes in the moisture content and soil aeration, preventing root development of the plants (Gill & Vanden Berg, 1968; Soane y van Ouwerkerk, 1994; Mouazen & Ramon, 2006; RPACAPV, 2008; Herrera *et al.*, 2011; Naderi-Boldaji *et al.*, 2013; Quraishi y Mouazen, 2013). It is determined through direct or indirect measurement of the bulk density or the soil penetration resistance. Quantitative determination of soil compaction is necessary to determine its severity and to identify the most appropriate way to fight it. Alternatively, increasing the compaction can be detected using indirect measurements of soil stresses (mechanical resistance to penetrate objects), or reducing interconnected pore spaces (fluid permeability).

Given the importance of knowledge of soil compaction as one of the most significant variables when developing crop management strategies, sensors have been developed for its continuous measurement and in specific site (Mouazen & Ramon, 2006; Herrera *et al.*, 2011; Naderi-Boldaji *et al.*, 2013; Quraishi & Mouazen, 2013). As working principle measuring stresses, soil permeability fluids, or water content in the soil (Adamchuk *et al.*, 2008) are used. The most accepted are those that base their operating principle in measuring the soil strength either in whole soil profile or in a given depth. These readings of strength are then used to estimate soil compaction.

Regarding this problem, several researchers have developed equations that allow correlating the sensor readings with the magnitude of soil compaction measured from the penetration resistance or the bulk density, which include both sensor-operating variables as characterizing their physical condition, being a determining factor sensor characteristics and its operating principle.

The operating parameters that are taken into account during the sensing compaction are depth, width of the working organ and speed. The variables most commonly used to characterize the physical condition of the soil are moisture, texture and density, being moisture content, the variable with the highest significance in the accuracy of predictions (Andrade *et al.*, 2001; Mouazen *et al.*, 2003; Hall & Raper, 2005).

With the advancement of information technology, methods based on machine learning have been fostered, which have focused on the development of models for the estimation of different variables. Some of them are Neural Networks and K nearest neighbors.

Machine Learning (AA) is a branch of Artificial Intelligence (AI), which aims to develop techniques for creating programs that can learn similar to humans, i.e. learn by themselves. Finally, the purpose is to create programs able to generalize behaviors from an unstructured information supplied as examples, which is a knowledge induction process. In many cases,

## INTRODUCCIÓN

La densidad del suelo puede ser considerada como uno de los factores de mayor importancia que indican la compactación del suelo. La compactación del suelo, es la disminución de su volumen por la aplicación de una carga mecánica, esta causa cambios en el contenido de humedad y en la aireación del suelo, impidiendo el desarrollo radicular de las plantas (Gill y Vanden Berg, 1968; Soane y van Ouwerkerk, 1994; Mouazen y Ramon, 2006; RPACAPV, 2008; Herrera *et al.*, 2011; Naderi-Boldaji *et al.*, 2013; Quraishi y Mouazen, 2013), es determinada a través de la medición directa o indirecta de la densidad volumétrica o la resistencia a la penetración del suelo. La determinación cuantitativa de la compactación del suelo es necesaria para determinar su severidad e identificar el modo más adecuado de combatirla. Como alternativa, el incremento de la compactación puede ser detectado mediante el uso de mediciones indirectas de las tensiones del suelo (resistencia mecánica para penetrar objetos), o la reducción de espacios porosos interconectados (permeabilidad de fluidos).

Dada la importancia que tiene el conocimiento de la compactación del suelo como una de las variables de mayor significación a la hora de elaborar estrategias de manejo de los cultivos, se han desarrollado sensores para su medición continua y en sitio específico (Mouazen y Ramon, 2006; Herrera *et al.*, 2011; Naderi-Boldaji *et al.*, 2013; Quraishi y Mouazen, 2013). Como principio de funcionamiento emplean la medición de los esfuerzos del suelo, la permeabilidad de los fluidos, o el contenido de agua en el suelo (Adamchuk *et al.*, 2008), los más aceptados son los que basan su principio de funcionamiento en la medición de la resistencia mecánica del suelo ya sea en todo el perfil del suelo o en una profundidad dada. Estas lecturas de la resistencia mecánica son utilizadas posteriormente para estimar la compactación del suelo.

A partir de esta problemática varios investigadores han desarrollado ecuaciones que le permiten correlacionar las lecturas del sensor con la magnitud de la compactación del suelo medida a partir de la resistencia a la penetración o de la densidad volumétrica, en las que incluyen tanto las variables de operación del sensor como las que caracterizan su estado físico, siendo un factor determinante las características del sensor y su principio de funcionamiento.

Los parámetros de operación que se tienen en cuenta durante el sensado de la compactación, son: la profundidad, el ancho del órgano de trabajo y la velocidad de trabajo. Las variables más utilizadas para caracterizar el estado físico del suelo son: humedad, textura y densidad del suelo, siendo el contenido de humedad la variable que mayor significación tiene en la exactitud de las predicciones (Andrade *et al.*, 2001; Mouazen *et al.*, 2003; Hall y Raper, 2005).

Con el avance de las técnicas informáticas, se han desarrollado métodos basados en el aprendizaje automatizado, los cuales han sido enfocados fundamentalmente al desarrollo de modelos para la estimación de diferentes variables (entre los que están Redes neuronales y K de los vecinos más cercanos).

El Aprendizaje Automatizado (AA) es una rama de la Inteligencia Artificial (IA) cuyo objetivo es desarrollar técnicas que permitan crear programas que puedan aprender de forma similar a lo realizado por los humanos, es decir, aprender por sí mismos. En fin, crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento. En muchas ocasiones el campo de actuación del mismo se solapa con el de la estadística, ya que las dos disciplinas se basan en el análisis de datos. Sin embargo, el Aprendizaje Automatizado se centra más en el estudio de la complejidad computacional de los problemas (Bello, 2012).

its field of action overlaps the statistics one, since the two disciplines are based on data analysis. However, machine learning is more focused on studying the computational complexity of the problems (Bello, 2012).

For the preliminary data analysis, various methods, such as selection of examples (experiments) and attributes (variables), and normalization of attributes, are used, allowing improving the quality of the learning model. They summarize the description of objects, thus the information is preserved as compact and light as possible. That favors learning task by distinguishing relevant attributes in the database, summarizing the information and increasing effectiveness of the learning process.

There are several methods of machine learning for data processing, including neural networks of multilayer type (MLP) and K nearest neighbor (K-NN) for the development of models to solve various problems and for their application, various tools, like Orange, Keel and Weka, can be used.

Given this problem, the objective of this work is to develop a computational model for estimating soil density through continuous sensing using an automated method of learning.

## METHODS

### Methods Used for Model Development

To develop the model, two numerical prediction methods were used: neural networks of multilayer type (MLP) and K nearest neighbor (K-NN).

Multilayer type neural networks are computational models that aim to simulate human brain from the development of architectures that take features of its functioning, without developing replicas of it (Burr, 1988). They are excellent classifiers trained in discrimination (Rynkiewicz, 2012), using learning with back propagation of errors (BP). This procedure belongs to the category of supervised; it requires knowing the correct outputs for each sample input. The learning process backward error propagation is for networks with a layer of input units, any number of intermediate layers and a layer of output units (Figure 1), (Carpenter, 1989)

Nearest neighbor systems are based on the assumption that the neighboring examples belong to the same class. Their learning phase is very simple, as it is limited to store the examples of the training set. The classifier looks for the k examples nearest to the data that are to be classified and assigns them the most frequent class among them. They are fast in training, but produce non-discriminatory classifiers, which means that the training process is based on maximizing the membership of the data to a class, ignoring information between classes (Araújo, 2006).

The K-NN constitutes a classical algorithm of the known form of solution to a similar problem and has been employed in classification and regression problems. The method consists in comparing the new instance to be classified with the existing data or examples of the problem in question, retrieving the k nearest examples, which depends on the similarity between the

Para el análisis previo de los datos se utilizan varios métodos como la selección de ejemplos (experimentos) y de atributos (variables), y la normalización de los atributos, permitiendo mejorar la calidad del modelo de aprendizaje. Sintetizando la descripción de los objetos, de manera tal que la información sea preservada de la forma más compacta y clara posible, favoreciendo la labor de aprendizaje, es decir, distingue los atributos realmente relevantes en la base de datos, sintetizando la información e incrementando la efectividad del proceso de aprendizaje.

Existen varios métodos de aprendizaje automatizado para el procesamiento de los datos entre los que se encuentran Redes neuronales del tipo multicapa (MLP) y K de los vecinos más cercanos (K-NN) para el desarrollo de modelos para la solución de diversos problemas y para su aplicación se pueden utilizar varias herramientas, entre ellas están Orange, Keel, Weka.

Tomando en cuenta esta problemática, el objetivo del trabajo es desarrollar un modelo matemático para la estimación de la densidad del suelo a través del sensoramiento continuo empleando un método de aprendizaje automatizado.

## MÉTODOS

### Métodos empleados para el desarrollo del modelo

Para el desarrollo del modelo se emplearon tres métodos de predicción numérica: redes neuronales del tipo multicapa (MLP), K de los vecinos más cercanos (K-NN).

Las redes neuronales del tipo multicapa, es un modelo computacional que pretende simular el funcionamiento del cerebro humano a partir del desarrollo de una arquitectura que toma rasgos del funcionamiento de este órgano, sin llegar a desarrollar una réplica del mismo (Burr, 1988). Son excelentes clasificadores, entrenados de forma discriminativa (Rynkiewicz, 2012), utilizando el aprendizaje con propagación hacia atrás de los errores BP (back propagation), este procedimiento pertenece a la categoría de supervisado, pues requiere conocer las salidas correctas para cada ejemplo de entrada. El procedimiento de aprendizaje con propagación del error hacia atrás es para redes con una capa de unidades de entrada, cualquier cantidad de capas intermedias y una capa de unidades de salida (Fig. 1), (Carpenter, 1989).

Los sistemas de vecinos más cercanos se basan en la suposición de que los ejemplos cercanos pertenecen a la misma clase. Su fase de aprendizaje es muy simple, pues se limita a almacenar los ejemplos del conjunto de entrenamiento. El clasificador busca los k ejemplos más cercanos al dato que se quiere clasificar y le asigna la clase más frecuente entre ellos. Son rápidos en el entrenamiento pero producen clasificadores no discriminatorios, lo que significa que el proceso de entrenamiento está basado en la maximización de la pertenencia de los datos a una clase, ignorando la información entre clases (Araújo, 2006).

El k-NN constituye un algoritmo clásico de la forma de solución conocida a un problema similar y ha sido empleado en problemas de clasificación y regresión. El método básicamente consiste en comparar la nueva instancia a clasificar con los datos o ejemplos existentes del problema en cuestión, recuperando los k ejemplos más cercanos, lo cual depende del parecido entre los atributos del



attributes of the new example with the examples of the learning sample or training. That brings about that the majority class of those k examples closest to it is returned.

nuevo ejemplo con los ejemplos de la muestra de aprendizaje o entrenamiento. Como resultado del mismo se devuelve la clase mayoritaria de aquellos k ejemplos más cercanos a él.

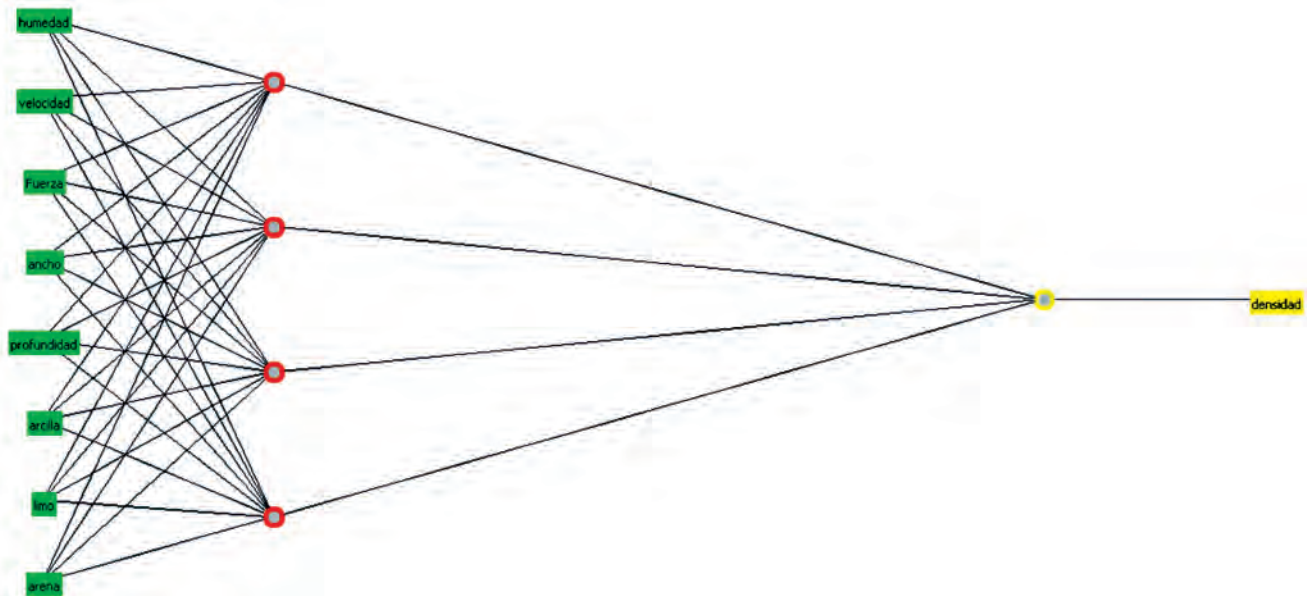


FIGURE 1. Diagram of a neural network of the type MLP.  
 FIGURA 1. Esquema de una red neuronal del tipo MLP.

For the application of the method and for the development of the model, Weka was the tool used, which constitutes a work environment that integrates a wide collection of algorithms. It was developed in Java language by a team of researchers from the University of Waikato (New Zealand) under GNU (General Public License), and it is characterized by the independence of its architecture, since it works on any platform on which there is a Virtual Java machine available. It allows applying, analyzing and evaluating some of the most relevant techniques of data analysis, among them data pre-processing, classification, regression, grouping, rules of association and visualization. It is free software. It is geared towards extensibility, so it is possible to add new functionalities. All this justifies that it is one of the most used tools in data mining (Witten & Frank, 2005).

Para la aplicación del método y su vez para el desarrollo del modelo la herramienta que se utilizó fue Weka, la cual constituye un entorno de trabajo que integra una amplia colección de algoritmos. Fue desarrollado en lenguaje Java por un equipo de investigadores de la universidad de Waikato (Nueva Zelanda) bajo licencia GNU (General Public License), y se caracteriza por la independencia de su arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible. Permite aplicar, analizar y evaluar algunas de las técnicas más relevantes del análisis de datos, dentro de las que se enmarcan: el pre-procesamiento de datos, clasificación, regresión, agrupamiento, reglas de asociación y visualización, es un software libre. Está orientado a la extensibilidad, por lo que es posible añadir nuevas funcionalidades. Todo lo anterior justifica que sea una de las herramientas más utilizadas en la minería de datos (Witten y Frank, 2005).

### Development of a Computational Model for Soil Density Estimation

The development of the model, which consisted of several stages, initially a database was created in a text editor, which included measurements of the horizontal draft force (D) of the soil during the experimentation under controlled conditions of the soil channel of CEMA, in addition to experiments performed by other researchers in different soil conditions and types (Table 1). The variables required by the models for the estimation of soil density, were also included. They were: moisture content (MC); force (D); working speed of the tool (V); working width (W); depth of work (d); content of clay (clay); silt content; sand content.

### Desarrollo de un modelo computacional para la estimación de la densidad del suelo

El desarrollo del modelo, contó de varias etapas, inicialmente se creó una base de datos en un editor de texto, donde se incluyeron mediciones de la fuerza de tiro horizontal (D) del suelo durante la experimentación en las condiciones controladas del canal de suelos del CEMA, además de experimentos realizados por otros investigadores en diferentes condiciones y tipos de suelo (Tabla 1). También se incluyeron las variables requeridas por los modelos para la estimación de la densidad del suelo, siendo: el contenido de humedad (MC); la fuerza (D); la velocidad de trabajo de la herramienta (V); el ancho de trabajo (W); la profundidad de trabajo (d); el contenido de arcilla (arcilla); el contenido de limo (limo); el contenido de arena (arena).

**TABLE 1. Soil Types Contained in the Database**  
**TABLA 1. Tipos de suelo contenidos en la base de datos**

Type of soil	Texture			Classification
	Clay	Silt	sand	
Oxisol	42	36.66	21.3	Loamy clay
Regesol	5.6	14	80.4	Sandy loam
Alfisol	25	21	54	Loamy clay loam

Subsequently a pre-processing of the database was performed, consisting of normalizing it, with the aim of transforming the range of values to a given interval (usually [0; 1]), so that all attributes (variables) are in the same rank. This procedure was performed once the database was imported into the automated learning tool Weka 3.7.5, using equation (1).

$$ValorNorm = \frac{Valor - Min}{Max - Min} \quad (1)$$

Once the database was normalized, the models were searched for using two numerical prediction methods: multilayer neural networks (MLP) and K nearest neighbor (K-NN) through the selected computational tool, a data division technique was applied for teaching training validation, called K-Fold Cross Validations, with K = 10, for each method.

**Result of the Application of Numerical Prediction Methods**

**Multilayer Neural Networks.** A neural network was designed with eight input attributes (independent variables) and one class (dependent variable), which has a hidden layer with four neurons. It was trained from the normalized data using back propagation methodology.

The training of the neural network, allowed obtaining a model for the prediction of soil density. The cross-validation results of this model showed (Table 2) that there is a correlation between the predicted and observed variables (r = 0.75), with a root mean square error of 0.16 g cm<sup>-3</sup> and an absolute average error of 13%.

**K- Nearest Neighbors.** This method was used to search for a model that allows the estimation of soil density. With this, the K examples closest to the problem are determined (K = 1). The results of the application of this method showed (Table 2), that there is correlation between the predicted and observed results (r = 0.89). The mean square error is 0.11 g cm<sup>-3</sup>.

Posteriormente se realizó un pre procesamiento de la base de datos, consistente en normalizar la misma, con el objetivo de transformar el rango de valores a un intervalo determinado (normalmente [0;1]), para que todos los atributos (variables) estén en el mismo rango. Este procedimiento se realizó una vez importada la base de datos en la herramienta de aprendizaje automatizado Weka 3.7.5, mediante la ecuación (1).

$$ValorNorm = \frac{Valor - Min}{Max - Min} \quad (1)$$

Una vez normalizada la base de datos, se pasó a la búsqueda de los modelos para lo cual se recurrió al empleo de tres métodos de predicción numérica: regresión lineal (RL); redes neuronales del tipo multicapa (MLP); y K- vecinos más cercanos (K-NN) a través de la herramienta computacional seleccionada, se aplicó una técnica de división de los datos para la enseñanza – entrenamiento – validación, denominada validación cruzada (K-Fold Cross Validations) con K=10, para cada método.

**Resultado de la aplicación de los métodos de predicción numérica**

**Redes neuronales del tipo multicapa.** Se diseñó una red neuronal con ocho atributos de entrada (variables independientes) y una clase (variable dependiente), la misma posee una capa oculta con cuatro neuronas. Fue entrenada a partir de los datos normalizados empleando la metodología de aprendizaje propagación hacia atrás (*backpropagation*).

El entrenamiento de la red neuronal, permitió obtener un modelo para la predicción de la densidad del suelo. Los resultados de la validación cruzada de este modelo mostraron (Tabla 2), que existe correlación entre las variables predichas y observadas (r=0,75), con una raíz del error cuadrático medio de 0,16 g·cm<sup>-3</sup>. Y un error medio absoluto de un 13%.

**K- Vecinos más Cercanos.** Este método se empleó para la búsqueda de un modelo que permita la estimación de la densidad del suelo. Con este se determina los K ejemplos más cercanos al problema (K=1).

Los resultados de la aplicación de este método mostraron (Tabla 2), que existe correlación entre los resultados predichos y observados (r=0,89), el error cuadrático medio es de 0,11 g·cm<sup>-3</sup>.

**TABLE 2. Results of the Methods Used**  
**TABLA 2. Resultado de los métodos utilizados**

Parameters	MLP	K NN
Correlation coefficient	0.75	0.89
Absolute average error	0.13	0.06
Mean square error (g cm <sup>-3</sup> )	0.16	0.11
Absolute relative error (%)	68.67	32.59
Standardized mean square error (%)	67.39	47.15
Number of instances	75	75

**Experimental Validation of the Model**

For the validation of the model developed by the method of the nearest neighboring K, 10 experiments were carried out on the soil channel. The horizontal draft force was determined for a wide range of moisture and density conditions (Table 3).

**Validación experimental del modelo**

Para la validación del modelo desarrollado por el método de los K vecinos más cercanos se realizaron 10 experimentos en el canal de suelos, se determinó la fuerza de tiro horizontal, para un amplio rango de condiciones de humedad y densidad (Tabla 3).

**TABLE 3. Experimental Results**  
**TABLA 3. Resultados experimentales**

Experiments	MC, kg·kg <sup>-1</sup>	BD, g·cm <sup>-3</sup>	V, m·s <sup>-1</sup>	D, kN
1	0.3938	1.06	1.56	0.566
2	0.3527	1.11	1.05	0.602
3	0.3327	1.12	0.82	0.602
4	0.2051	1.33	1.62	0.459
5	0.2306	1.31	1.56	0.459
6	0.22	0.95	1.41	0.456
7	0.3576	1.00	0.47	0.205
8	0.3477	1.00	0.47	0.205
9	0.3616	1.01	0.46	0.329
10	0.2306	1.31	0.46	0.488

From the experimental results (Table 3), and solving the obtained model, the values of the soil density were predicted, to be used in the validation of the model. Once the predicted values were obtained, the errors were determined by the statistics: mean square error (equation 2), standardized mean square error (equation 3), absolute mean error (equation 4), relative absolute error (equation 5), mean square error and correlation coefficient (equation 6).

Mean Squared Error

$$ECM = \frac{\sum_{i=1}^n (V_{obs} - V_{pred})^2}{n} \tag{2}$$

Where: Vobs, observed values;

Vpred, predicted values;

N, sample number.

Standardized Mean Square Error

$$ECME = \sqrt{\frac{\sum_{i=1}^n (V_{obs} - V_{pred})^2}{n}} \tag{3}$$

Absolute Average Error

$$EMA = \frac{\sum_{i=1}^n \left| \frac{V_{obs} - V_{pred}}{V_{obs}} \right|}{n} \tag{4}$$

Absolute Relative Error

$$EAR = \frac{\sum_{i=1}^n |V_{obs} - V_{pred}|}{\sum_{i=1}^n |V_{obs} - \bar{V}|} \tag{5}$$

Correlation Coefficient

$$r_{V_{obs}V_{pred}} = \frac{\sum_{i=1}^n (V_{obs} - \bar{V}_{obs})(V_{pred} - \bar{V}_{pred})}{(n-1)\sigma_{V_{obs}}\sigma_{V_{pred}}} \tag{6}$$

A partir de los resultados experimentales (Tabla 3), y resolviendo el modelo obtenido, se predijeron los valores de la densidad del suelo, a emplear en la validación del modelo. Una vez obtenidos los valores predichos los errores se determinaron mediante los estadígrafos: error cuadrático medio (ecuación 2), error cuadrático medio estandarizado (ecuación 3), error medio absoluto (ecuación 4), error absoluto relativo (ecuación 5), error cuadrático medio y el coeficiente de correlación (ecuación 6).

Error cuadrático medio

$$ECM = \frac{\sum_{i=1}^n (V_{obs} - V_{pred})^2}{n} \tag{2}$$

donde: Vobs, valores observados;

Vpred, valores predichos;

n, número de muestra.

Error cuadrático medio estandarizado

$$ECME = \sqrt{\frac{\sum_{i=1}^n (V_{obs} - V_{pred})^2}{n}} \tag{3}$$

Error medio absoluto

$$EMA = \frac{\sum_{i=1}^n \left| \frac{V_{obs} - V_{pred}}{V_{obs}} \right|}{n} \tag{4}$$

Error absoluto relativo

$$EAR = \frac{\sum_{i=1}^n |V_{obs} - V_{pred}|}{\sum_{i=1}^n |V_{obs} - \bar{V}|} \tag{5}$$

Coefficiente de correlación

$$r_{V_{obs}V_{pred}} = \frac{\sum_{i=1}^n (V_{obs} - \bar{V}_{obs})(V_{pred} - \bar{V}_{pred})}{(n-1)\sigma_{V_{obs}}\sigma_{V_{pred}}} \tag{6}$$

$\sigma_{Vobs}$ , absolute deviation of the observed values;  
 $\sigma_{pred}$ , absolute deviation of the predicted values.

The results of a correlation analysis show (Table 4) that there is a high correlation between the observed and predicted values  $r = 0.957$  (Figure 2). The standardized mean square error is  $0.0017 \text{ g cm}^{-3}$ , which is an indicator that the model developed is valid for the estimation of soil density in the limits analyzed.

$\sigma_{Vobs}$ , desviación absoluta de los valores observados;  
 $\sigma_{pred}$ , desviación absoluta de los valores predichos.

Los resultados de un análisis de correlación muestran (Tabla 4), que existe una alta correlación entre los valores observados y predichos  $r=0,957$  (Figura 2). El error cuadrático medio estandarizado es de  $0,0017 \text{ g}\cdot\text{cm}^{-3}$ , lo cual es un indicador de que el modelo desarrollado es válido para la estimación de la densidad del suelo en los límites analizados.

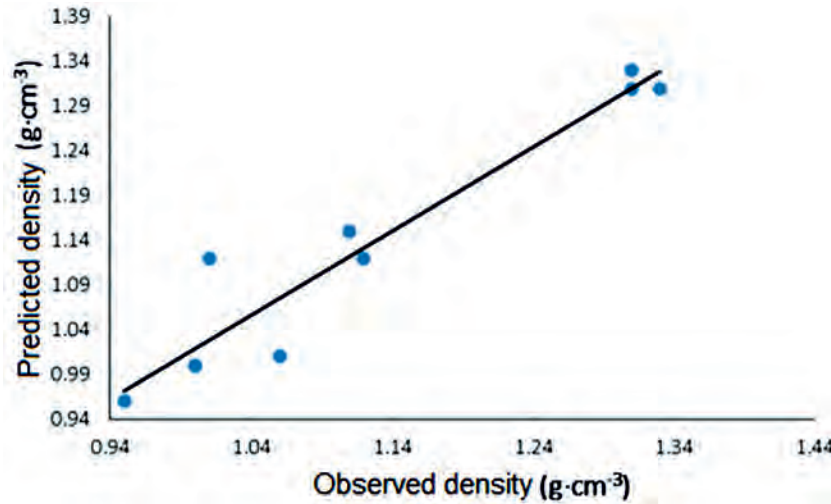


FIGURE 2. Scatter plot of observed and predicted soil density.  
 FIGURA 2. Gráfico de dispersión de la densidad del suelo observada y la predicha.

TABLE 4. Statistics of the Experimental Validation of the Model  
 TABLA 4. Estadígrafos de la validación experimental del modelo

Parameters	
Correlation coefficient	0.957
Absolute average error	0.027
Mean square error ( $\text{g cm}^{-3}$ )	0.0017
Absolute relative error (%)	21.18
Standardized mean square error (%)	46.02
Number of instances	10

## CONCLUSIONS

- The model developed by the method of K nearest neighbors is valid for the estimation of the soil density in the limits of the independent variables.
- The results of the validation of the new model showed that it is adequate to estimate the soil density by on-the-go soil sensing, since it exhibited a high coefficient of correlation ( $r = 0.957$ ) between the values predicted and observed in addition to a mean error absolute of 2.7%.
- The variables of greatest weight in the developed model are moisture content followed by speed, draft force, width of the working tool and depth.

## CONCLUSIONES

- El modelo desarrollado por el método de K vecinos más cercanos es válido para la estimación de la densidad del suelo en los límites de las variables independientes;
- Los resultados de la validación del nuevo modelo desarrollado evidenciaron que el mismo es adecuado para estimar la densidad del suelo mediante el sensoramiento continuo, pues este exhibió un elevado coeficiente de correlación  $r=0,957$  entre los valores predichos y observados además de un error medio absoluto de 2,7%;
- Las variables de mayor peso en el modelo desarrollado son: el contenido de humedad seguido de la velocidad, la fuerza de tiro, el ancho de la herramienta de trabajo y la profundidad.

## REFERENCES/REFERENCIAS BIBLIOGRÁFICAS

ADAMCHUK, V.I.; HEMMAT, A.; MOUAZEN, A.M.: "Soil Compaction Sensor Systems - Current Developments", [en línea], En: *ASAE Annual Meeting*, Ed. American Society of Agricultural and Biological Engineers (ASABE), Providence, Rhode Island, p. 10, 2008, DOI:



- 10.13031/2013.24814, Disponible en: <http://elibrary.asabe.org/abstract.asp?JID=5&AID=24814&CID=prov2008&T=1>, [Consulta: 30 de enero de 2017].
- ANDRADE, P.; ROSA, U.; UPADHYAYA, S.; JENKINS, B.; AGUERA, J.; JOSIAH, M.: "Soil profile force measurements using an instrumented tine", [en línea], En: *ASAE Annual Meeting*, Ed. American Society of Agricultural and Biological Engineers (ASABE), St. Joseph, Michigan, p. 11, 2001, DOI: 10.13031/2013.15697, Disponible en: <http://elibrary.asabe.org/abstract.asp?JID=5&AID=15697&CID=s-ca2001&T=1>, [Consulta: 30 de enero de 2017].
- ARAÚZO, A.: *Un sistema inteligente para selección de características en clasificación*, Universidad de Granada, Tesis de Doctorado, España, 238 p., 2006.
- BELLO, M.: *Un método de aproximación de funciones basado en el enfoque de los prototipos más cercanos utilizando relaciones de similitud*, Universidad Central "Marta Abreu" de Las Villas, Tesis de Grado, Santa Clara, Cuba, 94 p., 2012.
- BURR, D.J.: "Experiments on neural net recognition of spoken and written text", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7): 1162-1168, 1988, ISSN: 0096-3518, DOI: 10.1109/29.1643, Disponible en: <http://ieeexplore.ieee.org/document/1643/>, [Consulta: 30 de enero de 2017].
- CARPENTER, G.A.: "Neural network models for pattern recognition and associative memory", *Neural Networks*, 2(4): 243-257, 1989, ISSN: 0893-6080, DOI: 10.1016/0893-6080(89)90035-X, Disponible en: <http://www.sciencedirect.com/science/article/pii/089360808990035X>, [Consulta: 30 de enero de 2017].
- GILL, W.R.; VANDEN BERG, G.E.: *Soil dynamics in tillage and traction*, Ed. United States Department of Agriculture, Washington, D.C., 511 p., 1968.
- HALL, H.E.; RAPER, R.L.: "Development and concept evaluation of an on-the-go soil strength measurement system", *Transactions of the ASAE*, 48(2): 469-477, 2005, ISSN: 2151-0059, DOI: 10.13031/2013.18311, Disponible en: <http://elibrary.asabe.org/abstract.asp?JID=3&AID=18311&CID=t2005&v=48&i=2&T=1>, [Consulta: 30 de enero de 2017].
- HERRERA, S.M.; IGLESIAS, C.C.; LARA, C.D.; GONZÁLEZ, C.O.; LÓPEZ, B.E.: "Desarrollo de un sensor para la medición continua de la compactación del suelo", *Revista Ciencias Técnicas Agropecuarias*, 20(1): 06-11, 2011, ISSN: 2071-0054, Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S2071-00542011000100001&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2071-00542011000100001&lng=es&nrm=iso&tlng=es), [Consulta: 30 de enero de 2017].
- MOUAZEN, A.M.; RAMON, H.: "Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content", *Soil and Tillage Research*, 86(2): 218-229, 2006, ISSN: 0167-1987, DOI: 10.1016/j.still.2005.02.026, Disponible en: <http://www.sciencedirect.com/science/article/pii/S0167198705000711>, [Consulta: 30 de enero de 2017].
- MOUAZEN, A.M.; RAMON, H.; BAERDEMAEKER, J.D.: "Modelling Compaction from On-line Measurement of Soil Properties and Sensor Draught", *Precision Agriculture*, 4(2): 203-212, 2003, ISSN: 1385-2256, 1573-1618, DOI: 10.1023/A:1024513523618, Disponible en: <https://link.springer.com/article/10.1023/A:1024513523618>, [Consulta: 30 de enero de 2017].
- NADERI-BOLDAJI, M.; SHARIFI, A.; ALIMARDANI, R.; HEMMAT, A.; KEYHANI, A.; LOONSTRA, E.H.; WEISSKOPF, P.; STETTLER, M.; KELLER, T.: "Use of a triple-sensor fusion system for on-the-go measurement of soil compaction", *Soil and Tillage Research*, 128: 44-53, 2013, ISSN: 0167-1987, DOI: 10.1016/j.still.2012.10.002, Disponible en: <http://www.sciencedirect.com/science/article/pii/S016719871200222X>, [Consulta: 30 de enero de 2017].
- QURAIISHI, M.Z.; MOUAZEN, A.M.: "Calibration of an on-line sensor for measurement of topsoil bulk density in all soil textures", *Soil and Tillage Research*, 126: 219-228, 2013, ISSN: 0167-1987, DOI: 10.1016/j.still.2012.08.005, Disponible en: <http://www.sciencedirect.com/science/article/pii/S0167198712001821>, [Consulta: 30 de enero de 2017].
- RPACAPV: "Compactación del suelo", [en línea], 26 de mayo de 2008, Disponible en: [http://www.euskadi.net/r33-2288/es/contenidos/informacion/suelo/es\\_1044/compactacion.html](http://www.euskadi.net/r33-2288/es/contenidos/informacion/suelo/es_1044/compactacion.html), [Consulta: 14 de octubre de 2017].
- RYNKIEWICZ, J.: "General bound of overfitting for MLP regression models", *Neurocomputing*, ser. Advances in artificial neural networks, machine learning, and computational intelligence (ESANN 2011), 90: 106-110, 2012, ISSN: 0925-2312, DOI: 10.1016/j.neucom.2011.11.028, Disponible en: <http://www.sciencedirect.com/science/article/pii/S0925231212001865>, [Consulta: 30 de enero de 2017].
- SOANE, B.D.; VAN OUWERKERK, C.: *Soil Compaction in Crop Production*, [en línea], Ed. Elsevier, 695 p., Google-Books-ID: DP3KBA-AAQBAJ, 1994, ISBN: 978-0-08-093400-6, Disponible en: <https://books.google.com/cu/books?id=DP3KBA-AAQBAJ>, [Consulta: 30 de enero de 2017].
- WITTEN, I.H.; FRANK, E.: *Data Mining: Practical Machine Learning Tools and Techniques*, [en línea], Ed. Morgan Kaufmann, 2.ª ed., Amsterdam - Boston, MA, 560 p., 2005, ISBN: 978-0-12-088407-0, Disponible en: <https://www.amazon.es/Data-Mining-Practical-Techniques-Management/dp/0120884070>, [Consulta: 30 de enero de 2017].

Received: 14/06/2017.

Approved: 22/12/2017.

Darina Lara Coba, Prof. Auxiliar, Universidad Tecnológica de la Habana José Antonio Echeverría, Facultad de Ingeniería Mecánica, Departamento de Mecánica Aplicada, Marianao, La Habana, Cuba.

e-mail: [dlarac@mecanica.cujae.edu.cu](mailto:dlarac@mecanica.cujae.edu.cu)

Miguel Herrera Suárez, e-mail: [miguelhs2000@yahoo.com](mailto:miguelhs2000@yahoo.com)

María Matilde García Lorenzo, e-mail: [miguelhs2000@yahoo.com](mailto:miguelhs2000@yahoo.com)

Roberto Félix Beltrán Reyna, e-mail: [rfbeltrnr@espe.edu.ec](mailto:rfbeltrnr@espe.edu.ec)

Note: the mention of commercial equipment marks, instruments or specific materials obeys identification purposes, there is not any promotional commitment related to them, neither for the authors nor for the editor.