PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

PREDICTING SOIL BULK DENSITY USING HIERARCHICAL PEDOTRANSFER FUNCTIONS AND ARTIFICIAL NEURAL NETWORKS

JORGE SEBASTIÁN SILVA ORELLANA

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Advisor:

CARLOS BONILLA MELÉNDEZ

Santiago de Chile, April, 2017

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

PREDICTING SOIL BULK DENSITY USING HIERARCHICAL PEDOTRANSFER FUNCTIONS AND ARTIFICIAL NEURAL NETWORKS

JORGE SEBASTIÁN SILVA ORELLANA

Members of the Committee:

CARLOS BONILLA MELÉNDEZ

FRANCISCO SUÁREZ POCH

FELIPE ABURTO GUERRERO

CÉSAR SÁEZ NAVARRETE

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Santiago de Chile, April, 2017

To my family: parents, grandparents, wife and daughter for their unconditional support.

ACKNOWLEDGEMENTS

First I want to thank my family who supported me all this time. This includes my parents Cecilia Orellana and Enrique Silva, grandparents Jorge Orellana and Cecilia Alfaro, wife Antonieta Ortiz-Arrieta and daughter Marina Silva.

I express my greatest gratitude to the advising of Professor Carlos Bonilla for his deep empathy, which leaves me a profound life teaching. Without their guidance, support and patience this study would not have been possible. I would also like to thank to the members of the research team of Professor Bonilla, because of their suggestions and collaboration help me to grow as a researcher.

GENERAL INDEX

LIST OF TABLES

LIST OF FIGURES

RESUMEN

La densidad aparente del suelo (ρ_b) es una propiedad clave en física e hidrología de suelos. En la ausencia de mediciones en terreno, las Funciones de PedoTransferencia (FPTs) son típicamente utilizadas para estimar la ρ_b. Recientemente, y como resultado del avance de las Redes Neuronales Artificiales (RNAs), un nuevo método para formular ecuaciones para estimar la ρ_b ha sido desarrollado. Esta técnica es particularmente adecuada para formular un acercamiento jerárquico y para condiciones de cantidad limitada de datos de suelo. Es por esto que el objetivo de este estudio fue utilizar un acercamiento jerárquico para desarrollar una serie de FPTs utilizando RNAs para predecir la ρ_b, y comparar estas estimaciones con las obtenidas con 10 FPTs existentes reportadas en la literatura. Esta evaluación fue hecha con un set de datos independiente de 1.007 mediciones de ρ_b provenientes de un amplio rango de suelos de Chile Central. Los resultados probaron que el desempeño (en términos de exactitud y error global) es función de la cantidad de variables de entrada con un rango de r^2 de 0,22 a 0,72 y un rango de RMSE de 0,32 a 0,17 Mg m^{-3}. Los resultados demuestran que usando sólo información de la distribución de tamaño de partícula como variable de entrada para estimar la ρ_b no tiene una precisión adecuada y que un modelo basado en el contenido de carbono orgánico (CO) es capaz de estimar con más precisión que uno basado en la distribución de tamaño de partícula, incluso en suelos con bajo contenido de CO. Además, la inclusión del pH y los cationes básicos como variables de entrada mejora la exactitud de las estimaciones. Los resultados muestran que la mejor ecuación se generó con las variables de entrada de: distribución de tamaño de partícula, contenido de CO, profundidad de suelo y contenido de humedad a punto de marchitez permanente. Finalmente, con las mismas variables de entrada, las ecuaciones desarrolladas con RNAs incrementan la calidad de las estimaciones en comparación con las clásicas regresiones multi-variables, mejorando el r^2 entre 0,02 y 0,14 y el RMSE entre 0,01 y 0,04 Mg m^{-3}.

Palabras Claves: Chile Central, densidad aparente del suelo, funciones de pedotransferencia, redes neuronales artificiales.

ABSTRACT

Soil bulk density (ρ_b) is a key soil property in soil physics and hydrology. In the absence of field measurements, the ρ_b values are typically estimated using PedoTransfer Functions (PTFs). Recently, because of the progress of Artificial Neural Network (ANN) technique, a new method has been used to develop equations for predicting ρ_b. This technique is particularly suitable for developing a hierarchical approach for a range of conditions from very limited soil input data to a more extended set of predictors. Therefore, the objective of this study was to use a hierarchical approach to develop a series of PTFs using the ANN technique for predicting ρ_b and compare the estimates with those obtained using 10 existing PTFs. This study was conducted using an independent sample set of 1,007 measured ρ_b values from a wide range of soils from Central Chile. The results proved that the performance of the developed equations improved (in accuracy and overall error) with the addition of input parameters (r^2 improved from 0.22 to 0.72 and RMSE from 0.32 to 0.17 Mg m^{-3}). The lowest performance was observed when using sand, silt, and clay contents as inputs. The results demonstrated that an equation based on the organic carbon (OC) content predicted the ρ_b values more effectively than those based on soil particle size distribution, even when used to predict ρ_b for soils with a low OC. Moreover, adding the pH and basic cations as inputs increased the accuracy of the estimates. The highest performance was found when using sand, silt, clay, OC, soil depth, and water content at wilting point as inputs. With the same input parameters, the equations developed with the ANN technique enhanced the quality of estimates compared with the 10 PTFs evaluated in this study, improved the r^2 between 0.02 and 0.14, and reduced the overall error between 0.01 and 0.04 Mg m^{-3}.

Keywords: Artificial Neural Networks, Central Chile, pedotransfer functions, soil bulk density.

1. INTRODUCTION

1.1. Overwiew

Soil bulk density (ρ_b) is one of the key soil properties in soil physics and hydrology and is computed as the mass of an oven-dry sample of undisturbed soil per unit bulk volume (Mg m^{-3}). The ρ_b has been incorporated as an essential input parameter in water, biophysical, sediment and nutrient transport models (Suuster et al., 2011). These models are very sensitive to ρ_b because it directly affects soil porosity and site productivity by controlling soil compaction, infiltration, and runoff rate (Brown and Heuvelink, 2005). Therefore, the ρ_b is considered as a major soil property that governs soil functioning in ecosystems and soil management. Data on ρ_b is used in research and applications in hydrology, agronomy, meteorology, ecology, environmental protection, and many other soil-related fields (Rawls et al., 2004). Moreover, the ρ_b is related to several soil parameters, such as soil organic fraction, soil texture, soil structure, water retention, carbon sequestration, nutrient mass of a soil layer and hydraulic conductivity (Dam et al., 2005). Furthermore, the ρ_b is used for soil unit conversions (from weight to area and volume), such as carbon reservoirs and stock of nutrients (Nanko et al., 2014).

Two general ways to measure ρ_b are used. One is the core method, which uses a cylinder that is inserted in the soil layer. Later, the dry mass of the soil is determined by oven drying at 105°C in laboratory. Once the value of the dry mass is obtained is divided by the known value of the cylinder volume and the ρ_b is computed (Klute, 1986).

The other way to measure the ρ_b is the clod method. This method extracts an undisturbed clod of soil and the sample is weighted and as the clod method, the dry mass of the soil is determined by oven drying at 105°C. After this, a length of thread is tied around the clod and is introduced in a container with melted paraffin. When the adhering paraffin solidifies, the clod and paraffin are weighted together and the clod is introduced in a graduated cylinder with water with known volume. The displaced volume corresponds to the soil volume of the clod. Finally, the ρ_b value is achieved with a formula that uses

the paraffin and water densities. It is recognized that the clod method gives slightly higher values of ρ_b because it does not incorporate the inter-pores in the soil (Blake and Hartge, 1986).

For large-scale areas, measuring ρ_b is a labor-intensive and time-consuming task (Kaur et al., 2002; Benites et al., 2007); as a result, soil surveys usually do not report the ρ_b value and it is typically necessary to estimate it based on other soil properties (Hollis et al., 2012). A series of regression equations has been developed to compute ρ_b from more simple soil properties. These types of equations are called PedoTransfer Functions (PTFs) (Rawls, 1983; Bonilla and Cancino, 2001; Brahim et al., 2012). Without adequate PTFs many modeling tools would not have been improved simply because there are no resources to be able to measure all the parameters required by them. Therefore, it is imperative to use an accurate ρ_b value or a reliable estimate.

Several studies have been done to identify and understand the factors controlling the ρ_b. Therefore, PTFs for estimating ρ_b have been developed worldwide and use different types of functions such as linear, polynomial, logarithmic and exponential (Nanko et al., 2014). These studies have shown that ρ_b mainly depends on the organic matter (OM) content (Adams, 1973; Prévost, 2004; Hollis et al., 2012); however, when the OM content is low, soil particle size distribution becomes more significant for estimating ρ_b. Additionally, other studies have concluded that ρ_b changes with some chemical soil properties such as pH (Bernoux et al., 1998; Brahim et al., 2012) or basic cations (sum of Ca^{+2}, Mg^{+2}, and K^+) (Benites et al., 2007) and some physical soil properties such as the soil water content at wilting point (θ_{1500}) (Heuscher et al., 2005) and at field capacity (θ_{33}) (Patil and Chaturvedi, 2012). Other functions have used qualitative factors such as morphological data, parent material (Calhoun et al., 2001; Jalabert et al., 2010), soil horizons (Alexander, 1980), vegetation (Jalabert et al., 2010), and soil type (Suuster et al., 2011).

Recently, new methods have been developed to estimate the ρ_b, including the boosted regression trees model (Martin et al., 2009; Jalabert et al., 2010). Although this new technique generates an improvement in the calibration (Martin et al., 2009), it does not necessarily predict the ρ_b with more accuracy (Tranter et al., 2007; Martin et al., 2009).

Additionally, the Artificial Neural Networks (ANN) technique, which considers the complex relationships between input variables, are being used in order to build PTFs for predicting soil properties (Koekkoek and Booltink, 1999; Wösten et al., 2001; Merdun et al., 2006; Baker and Ellison, 2008; Lagos-Avid and Bonilla, 2017) and specifically for predicting ρ_b (Al-Qinna and Jaber 2013; Xiangsheng et al., 2016).

The ANN technique use an iterative calibration procedure to find a relationship between soil properties, including the capability to detect complex nonlinear relationships between dependent and independent variables. Basically, an ANN consists of many interconnected simple computational elements called nodes, and the outputs of nodes are used as input to other nodes in the network (Fig. 1). When the number of inputs is larger than three, ANN usually do better than regression techniques, so they are a good alternative for the development of empirical models (Wösten et al., 2001).

However, the performance of a PTF, independently of the method from which is generated, greatly depend on the sample set used for calibration (accuracy) and the sample set used for evaluation (reliability) (Pachepsky et al., 1999). Therefore, testing a PTF using an independent sample set including a wide range of soil types is essential to evaluate their performance and applicability for other site conditions. The ANN technique considers the above, by accounting a portion of the dataset for independent testing. Additionally, a hierarchical design with an increasing number of predictors to accommodate different levels of soil data that can be considered as potential predictors for ρ_b allows the use of limited and more extended sets of predictors that facilitate the practical use of the PTFs (Schaap et al., 2001).
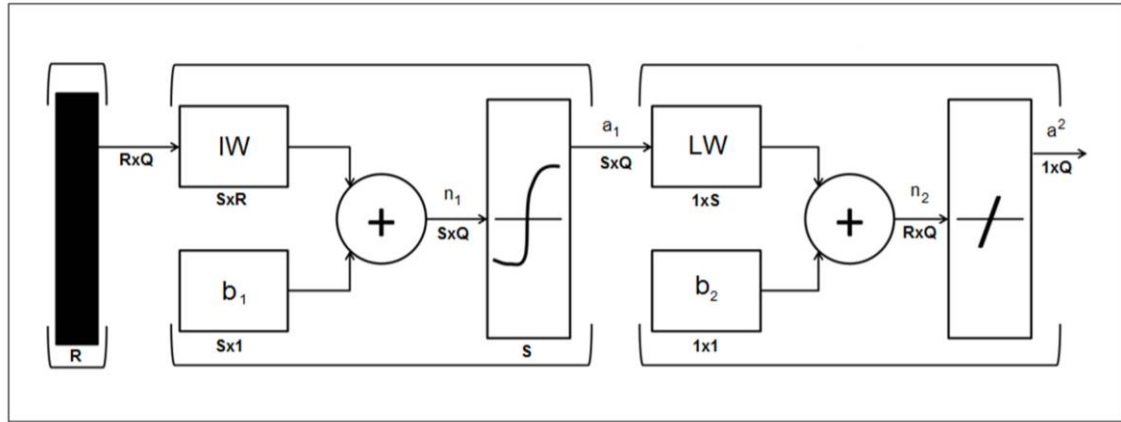
Figure 1: Basic structure of an Artificial Neural Network.

There are not yet equations for predicting ρ_b based on ANN and hierarchical approach with more than OC content, soil particle size distribution and depth as inputs. Moreover, the PTFs for estimating ρ_b have been developed worldwide but no previous studies have developed functions for estimating ρ_b using the soils of Chile. Therefore, the objective of this study was to use a hierarchical approach to develop a series of PTFs using the ANN technique for predicting ρ_b and to compare the estimates with those obtained using 10 already existing PTFs for predicting ρ_b. This study used an independent sample set of 1,007 measured ρ_b values from a wide range of soil conditions from Central Chile. Additionally, an external validation for these 10 PTFs was generated.

1.2.   Hypothesis

This research is based on the hypothesis that it is possible to develop a set of equations hierarchically established for estimating the soil bulk density using the artificial neural network techniques based on different measured soil parameters and achieve high accuracy in the estimates.

1.3.   Objective

The overall objective of this study was to develop a set of equations hierarchically based to estimate the soil bulk density based on artificial neural network techniques. For this purpose, the following specific objectives were established:

a) Review and select the equations already available for estimating the soil bulk density.
b) Identify key variables that control and explain the soil bulk density.
c) Building a soil dataset based on the Natural Resources Information Center soil surveys (CIREN, in spanish).
d) Formulate and validate a hierarchical set of equations to estimate the soil bulk density using the artificial neural network techniques and based on the CIREN soil surveys.
e) Compare the reliability of the equations developed with the 10 already published pedotransfer functions for estimating the soil bulk density.
f) Generate an external validation for these 10 selected published pedotransfer functions by testing these functions with and without calibration with the soil dataset of CIREN.

1.4.   Methodology

The CIREN soil dataset was digitized and filtered. The resulting dataset consists of 1,007 samples with soil bulk density data from 243 soil series collected in six different regions along Central Chile (from V to IX Region). Correlation and multiple regression analysis were used to identify the main variables controlling the soil bulk density.

The main parameters influencing the soil bulk density were used for developing a hierarchical set of pedotransfer functions for estimating the soil bulk density with

artificial neural network, and the equations performance were compared with 10 already existing pedotransfer functions. The performance was evaluated with the coefficient of determination, the root mean square error, and the Nash-Sutcliffe model efficiency. Additionally, the equations collected from literature were tested with and without calibration as external validation for these equations.

1.5.   Results

A hierarchical set of new equations developed with artificial neural network was built for estimating the soil bulk density. The input variables were: sand, silt and clay content; organic carbon content; pH; basic cations (sum of Mg^{+2}, K^{+2}, Ca^{+2}); soil depth; and soil water content at wilting point. These inputs were settled in six equation types. These equations estimated the soil bulk density with r^2 of 0.22, 0.38, 0.44, 0.49, 0.55, and 0.72, respectively. The equations were compared with 10 empirical published pedotransfer functions collected from literature. The generated equations showed better performance when predicting soil bulk density than classical multivariable regressions.

1.6.   Conclusions

The main conclusions of this study were:

1)   The interaction between the soil properties is a complex relationship and soil bulk density cannot be accurately estimated with only one input parameter.

2)   The use of artificial neural network techniques proved to be a suitable method for building six equations for predicting ρ_b across a wide range of soil types and inputs, and their capability for predicting ρ_b was related to the number of parameters used in the network.

3) Depending on the field and soil survey data, these ANN are meant to be selected and used hierarchically to predict ρ_b values.

4) Sand, silt, clay, organic carbon, soil depth and soil water content at wilting point can be used to explain with high accuracy ($r^2=0.72$) the soil bulk density.

5) The results demonstrate that ANN technique is an option for building a robust relationship for predicting variable soil properties as bulk density. Therefore, it is highly recommended the use of this new set of equations.

6) The inputs required for the networks are usually available in soil surveys and the networks can be easily coupled with other soil and hydrology models to provide an alternative for estimating ρ_b when its measured value is not available.

2.   LITERATURE REVIEW

Soil bulk density (ρ_b) is one of the key soil properties in soil physics and hydrology and is computed as the mass of an oven-dry sample of undisturbed soil per unit bulk volume (Mg m^{-3}). The ρ_b has been incorporated as an input parameter in water, biophysical, sediment and nutrient transport models (Suuster et al., 2011). These models are very sensitive to ρ_b because it directly affects soil porosity and site productivity by controlling soil compaction, infiltration, and runoff rate (Brown and Heuvelink, 2005). Therefore, it is imperative to use an accurate ρ_b value or a reliable estimate. Moreover, ρ_b is related to other important soil properties such as soil structure, water retention, nutrient content, and hydraulic conductivity (Dam et al., 2005; Nanko et al., 2014).

For large-scale areas, measuring ρ_b is a labor-intensive and time-consuming task (Kaur et al., 2002; Benites et al., 2007); as a result, soil surveys usually do not report the ρ_b value and it is typically necessary to estimate it based on other soil properties (Hollis et al., 2012). A series of regression equations has been developed to compute ρ_b from more simple soil properties. These types of equations are called PedoTransfer Functions (PTFs) (Rawls, 1983; Bonilla and Cancino, 2001; Brahim et al., 2012).

Either a physically or empirically based approach has been used for building the PTFs for predicting ρ_b. For example, the physically based PTF developed by Adams (1973) uses the organic matter content (OM, %), the bulk density of organic matter fraction ($\rho_{b,o}$, Mg m^{-3}), and the bulk density of mineral fraction ($\rho_{b,m}$, Mg m^{-3}) for predicting the ρ_b value (Table 2). Several studies have used this equation and calibrated the $\rho_{b,o}$ and $\rho_{b,m}$ values for specific sites (Tremblay et al., 2002; De Vos et al., 2005; Han et al., 2012). These studies used soils with high OM content (from 3% to 30%) and concluded that this property is essential for predicting ρ_b.

On the other hand, the empirically based PTFs use different types of functions such as linear, polynomial, logarithmic, and exponential (Nanko et al., 2014). Previous studies have shown that ρ_b mainly depends on the OM content for these type of PTFs (Adams, 1973; Prévost, 2004; Hollis et al., 2012); however, when the OM content is low (1% or

less), Bernoux et al. (1998) suggested that soil particle size distribution becomes more significant for estimating ρ_b. Additionally, other studies have concluded that ρ_b changes with some chemical soil properties such as pH (Bernoux et al., 1998; Brahim et al., 2012) or basic cations (sum of Ca^{+2}, Mg^{+2}, and K^+) (Benites et al., 2007) and some physical soil properties such as the soil water content at wilting point (θ_{1500}) (Heuscher et al., 2005) and at field capacity (θ_{33}) (Patil and Chaturvedi, 2012). Other functions have used qualitative factors such as morphological data, parent material (Calhoun et al., 2001; Jalabert et al., 2010), soil horizons (Alexander, 1980), vegetation (Jalabert et al., 2010), and soil type (Suuster et al., 2011).

The performance of a PTF greatly depends on the sample set used for calibration (accuracy) and the sample set used for evaluation (reliability) (Pachepsky et al., 1999). Therefore, testing a PTF using an independent sample set including a wide range of soil types is essential to evaluate their performance and applicability for other site conditions. Nonetheless, few studies have tested PTFs using independent soil datasets (Boucneau et al., 1998; Kaur et al., 2002; De Vos et al., 2005; Han et al., 2012; Nanko et al. 2014; Vasiliniuc and Patriche, 2015). These studies showed significant differences among the PTFs when they were tested on different soils than those used for calibration, which established that PTFs depend on the specific soils from which they were developed.

Recently, considering the complex relationships between soil properties, different studies have been conducted using the Artificial Neural Network (ANN) technique to build PTFs for predicting soil properties (Koekkoek and Booltink, 1999; Wösten et al., 2001; Merdun et al., 2006; Baker and Ellison, 2008; Lagos-Avid and Bonilla, 2017) and specifically for predicting ρ_b (Al-Qinna and Jaber 2013; Xiangsheng et al., 2016). The ANN technique use an iterative calibration procedure to find a relationship between soil properties, including the capability to detect complex nonlinear relationships between dependent and independent variables. Basically, an ANN consists of many interconnected simple computational elements called nodes (Wösten et al., 2001). Additionally, a hierarchical design with an increasing number of predictors to accommodate different levels of soil data that can be considered as potential predictors

for ρ_b allows the use of limited and more extended sets of predictors that facilitate the practical use of the PTFs (Schaap et al., 2001).

There are not yet equations for predicting ρ_b based on ANN and hierarchical approach with more than OC content, soil particle size distribution and depth as inputs. Moreover, the PTFs for estimating ρ_b have been developed worldwide but no previous studies have developed functions for estimating ρ_b using the soils of Chile. Therefore, the objective of this study was to use a hierarchical approach to develop a series of PTFs using the ANN technique for predicting ρ_b and to compare the estimates with those obtained using 10 already existing PTFs for predicting ρ_b. This study used an independent sample set of 1,007 measured ρ_b values from a wide range of soil conditions from Central Chile.

3.   MATERIALS AND METHODS

3.1.   Soil samples

The soil samples used in this study came from a series of soil surveys performed by the Natural Resources Information Center CIREN (1996a, 1996b, 1997a, 1997b, 1999, 2002) from Chile's Ministry of Agriculture between the years 1971 and 1983. The location of the sampled sites extended from latitudes 32°S to 40°S and longitudes 70° to 73°W and covered the central region of Chile (Fig. 2), which has a wide range of climate conditions (Bonilla and Vidal, 2011). Most of the study area is in the Central Valley, a geological plain between the Western Andes Mountains and the coastal range that extends for approximately 1,000 km from the Valparaíso Region south to the Araucanía Region. This plain is approximately 70 km wide and is composed of a vast thick deposit of heavily mineralized alluvial soils formed by the principal rivers of the region (Bonilla and Johnson, 2012). Most farmland (72%) and national forest areas (54%) of the country are concentrated in this area (INE, 2007).

A total of 1,007 soil samples with ρ_b data (from 243 soil series) were identified in the study area. The samples corresponded to seven soil orders: Alfisols (26 series, 106 horizons), Andisols (41 series, 181 horizons), Inceptisols (74 series, 313 horizons), Mollisols (83 series, 334 horizons), Entisols (10 series, 37 horizons), Ultisols (7 series, 29 horizons), and Vertisols (2 series, 7 horizons). The 1,007 samples covered the 12 USDA soil textures (Fig. 3). A summary of the soil properties is listed in Table 1.
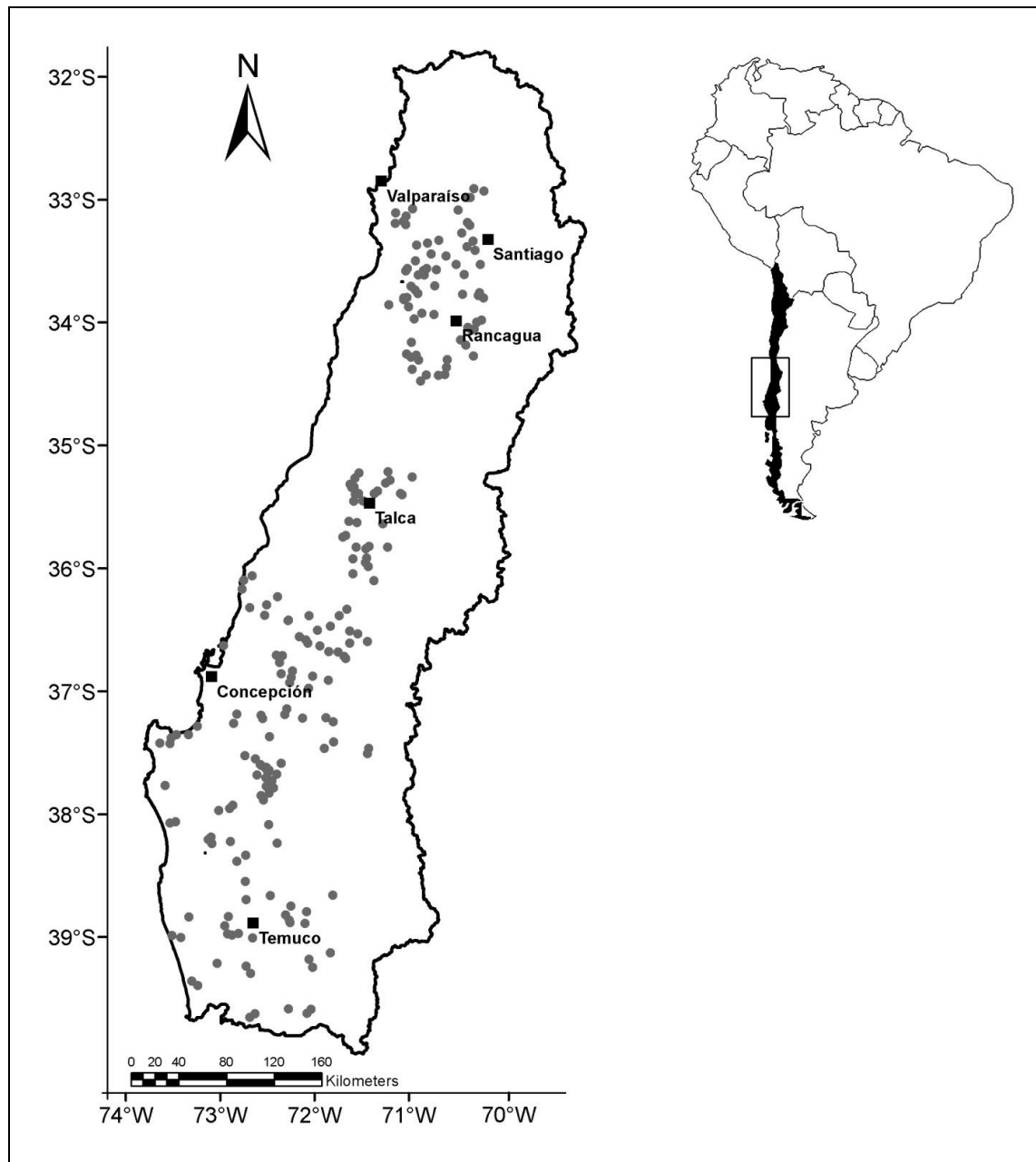
Figure 2: Location of the soil data (soil pits) from soil surveys.

Table 1. Summary of properties of the soil samples used in the study.

Variable	N	Mean	Median	Minimum (x_{min})	Maximum (x_{max})	Standard deviation	Coef. of variation (%)
Sand fraction	1,007	0.39	0.34	0.03	1.00	0.23	59
Silt fraction	1,007	0.36	0.36	0.00	0.85	0.16	43
Clay fraction	1,007	0.25	0.24	0.00	0.78	0.16	62
Bulk density (Mg m^{-3})	1,007	1.33	1.37	0.56	2.10	0.36	27
Depth (m)	1,007	0.48	0.41	0.03	2.25	0.34	71
θ_{33} (% grav.)	754	30.48	27.00	1.80	107.27	16.78	55
θ_{1500} (% grav.)	754	18.70	16.00	1.10	63.60	11.66	62
Organic carbon (%)	1,007	1.74	1.02	0.00	13.03	2.03	117
Basic cations (cmol kg^{-1})	880	11.70	9.67	0.27	43.10	8.79	75
pH	997	6.52	6.40	4.60	9.20	0.81	12
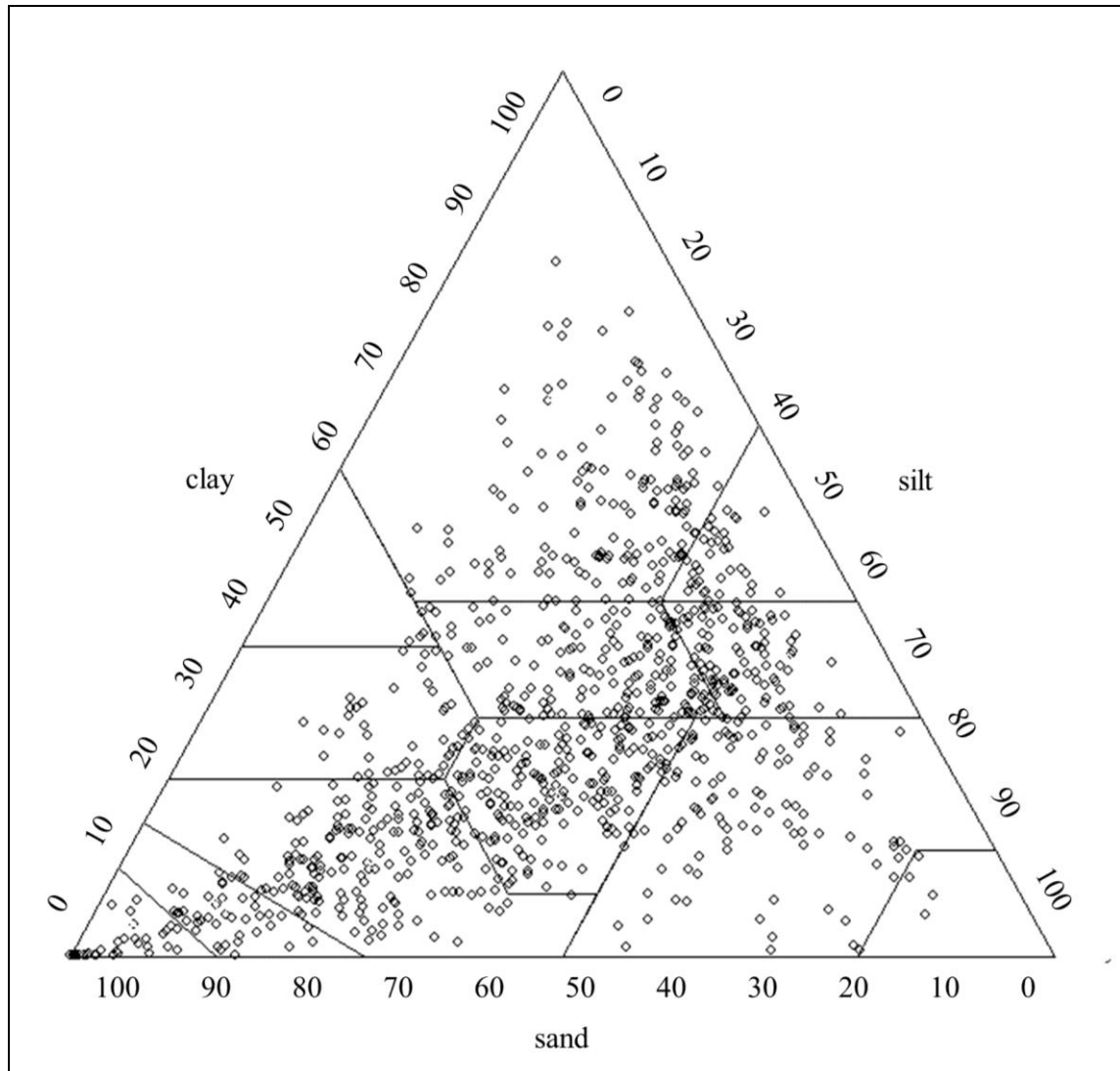
Figure 3: Soil texture distribution among the 1,007 soil samples used in the study.

The soil analysis at the time of the field surveys was conducted according to the standard soil methods used in Chile (INIA, 2006). The ρ_b was determined by the clod method (Blake and Hartge, 1986), and the particle size distribution was determined by the Bouyoucos method (Bouyoucos, 1962) according to the USDA textural classification (sand 2-0.05 mm; silt 0.05-0.002 mm; clay <0.002 mm). The OC content was measured by the dichromate method (Walkley and Black, 1934), and the OM content was obtained by multiplying the OC value by the Van Bemmelen factor of 1.724 (Davis, 1974). The pH was measured by suspension and potentiometric determination (proportion 1:1 soil/water), the water retention parameters (θ_{1500} and θ_{33}) by the pressure plate method (Klute, 1986), and the calcium, magnesium, and potassium contents were measured using extraction with an ammonium acetate solution of 1 mol L^{-1} at pH 7.0 and spectrophotometric determination with atomic absorption and emission (using lanthanum suppression) (INIA, 2006).

3.2.   Artificial neural networks

The hierarchical approach for predicting ρ_b using ANN was developed by building six different networks (named A to F) based on the type and number of input parameters collected from the literature. The first network (A) used sand, silt, and clay content as inputs, whereas the second network (B) used only OM content as input. The network C was a combination of networks A and B: it used sand, silt, and clay content in addition to the OM content. Networks D, E, and F used the same parameters as network C but included pH, basic cations, and soil depth and θ_{1500}, respectively, as inputs.

The ANN were built with a two-layer feed forward network (hidden layer and output layer) using the Levenberg-Marquardt back-propagation algorithm as the training function (Hagan and Menhaj, 1994). The transfer function used was a hyperbolic tangent sigmoid, and the performance function was the mean square error (MSE). The soil samples were randomly divided into three sets: (i) training (70%), (ii) validation (15%), and (iii) testing (15%). The training and validation sets were used to develop the

network, and the testing set was used to evaluate the equation independently of the development of the network. Every time an ANN was executed a different solution for the same problem was generated because of the randomization of these three sets (Tamari et al., 1996). For this reason, a total of 100 iterations were performed and the targets of the ANN were the measured ρ_b values. All simulations were performed with the neural network toolbox of MATLAB^{TM} (The MathWorks Inc., USA). The number of nodes in the hidden layer is commonly estimated based on the number of inputs (Hecht-Nielsen, 1987; Sheela and Deepa, 2013); however, different numbers of nodes were tested (from 1 to 15) in this study in search of the optimal performance of the ANN, which generated 1,500 iterations per network. The criteria for choosing the best iteration per network were: (i) to obtain predicted values in the ρ_b range (from 0 to approximately 2.5 Mg m^{-3}), (ii) the MSE for training and validation must be the lowest, and (iii) the maximum and minimum values of the input variables must be part of the training set (Tamari et al., 1996).

To reduce the "black box" effect usually associated with the ANN structure (Olden and Jackson, 2002) and to isolate the contribution of each input variable within the network, the Relative Importance (RI) of each input parameter in the ANN was computed using the algorithm developed by Garson et al. (1991) and described in detail by Olden and Jackson (2002).

3.3.   **Equations tested for predicting soil bulk density**

Ten already existing PTFs were selected for comparing their performance with those developed with the ANN technique. These PTFs were chosen considering: (i) their being built with soil data from different territories or a wide range of soil types, (ii) their use of different numbers of input variables, and (iii) their generation based on relatively large sample sets. These 10 PTFs were grouped in the same categories as the equations developed with the ANN (from A to F) based on the inputs.

The main characteristics of each are listed in Table 2. As suggested by De Vos et al. (2005), all the available horizons were used for testing the PTFs. The PTFs were tested in their original (without calibration) and calibrated forms. The calibration procedure consisted of computing the equation constants by minimizing the Root Mean Square Error (RMSE) (Eq. 1) between the estimated and measured ρ_b values. Specifically, the Rawls et al. (2004) equation was evaluated for A-horizons (because this equation was developed for this soil layer) and also for the entire set of samples. On the other hand, the Brahim et al. (2012) equation uses the sand content computed for soil particles between 0.20 to 2 mm; because the same range was not available in this soil dataset, the closest range (between 0.25 and 2 mm) was used instead.

Table 2. Pedotransfer functions evaluated in the study. Sample sizes (n), mean ρ_b and r^2 values from the original publications.

PTF	Source	Location	Eq. ρ_b =	ρ_b mean (Mg m^{-3})	n	r^2
1	Saxton et al. (1986)[1]	USA	$(1-(0.332-0.0007251 \bullet \text{Sand}+0.1276 \bullet \text{LOG}(\text{Clay}))) \bullet 2.65 \bullet 0.92$	NR	2,000[2]	NR
2	Adams (1973)[3]	Merionethshire, Wales	$100/((\text{OM}/\rho_{b,o})+((100-\text{OM})/\rho_{b,m}))$	NR	77	NR
3	Tomasella and Hodnett (1998)	Brazil	$1.578-0.054 \bullet \text{OC}-0.006 \bullet \text{Silt}-0.004 \bullet \text{Clay}$	NR	396	0.60
4	Kaur et al. (2002)	Utaranchal, India	$\exp(0.313-0.191 \bullet \text{OC}+0.02102 \bullet \text{Clay}-0.000476 \bullet \text{Clay}^2-0.00432 \bullet \text{Silt})$	1.36	224	0.62
5	Rawls et al. (2004)	USA, and other countries	(see equation below)	NR	2,100	NR

$$1.36411+0.185628 \bullet (0.0845397+0.701658 \bullet w-0.614038 \bullet w^2$$
$$-1.18871 \bullet w^3+0.0991862 \bullet y-0.301816 \bullet w \bullet y-0.153337 \bullet w^2 \bullet y$$
$$-0.0722421 \bullet y^2+0.392736 \bullet w \bullet y^2+0.0886315 \bullet y^3-0.601301 \bullet z$$
$$+0.651673 \bullet w \bullet z-1.37484 \bullet w^2 \bullet z+0.298823 \bullet y \bullet z-0.192686 \bullet w \bullet y \bullet z$$
$$+0.0815752 \bullet y^2 \bullet z-0.0450214 \bullet z^2-0.179529 \bullet w \bullet z^2-0.0797412 \bullet y \bullet z^2$$
$$+0.00942183 \bullet z^3)$$

where :

$$x = -1.2141+4.23123 \bullet \text{sand}$$
$$y = -1.70126+7.55319 \bullet \text{clay}$$
$$z = -1.55601+0.507094 \bullet \text{OM}$$
$$w = -0.0771892+0.256629 \bullet x+0.256704 \bullet x^2-0.140911 \bullet x^3-0.0237361 \bullet y$$
$$-0.098737 \bullet x^2 \bullet y-0.140381 \bullet y^2+0.0140902 \bullet x \bullet y^2+0.0287001 \bullet y^3$$

Table 2. Continued

PTF	Source	Location	Eq. ρ_b =	ρ_b mean (Mg m^{-3})	n	r^2
6	Hollis et al. (2012)	Europe	$0.69794 + \begin{pmatrix} 0.750636 \bullet \exp(-0.230355 \bullet OC) \\ + (0.0008687 \bullet Sand) - (0.0005164 \bullet Clay) \end{pmatrix}$	1.27	925	NR
7	Bernoux et al. (1998)	Brazil	$1.524 - 0.0038 \bullet Clay - 0.05 \bullet OC - 0.045 \bullet pH + 0.001 \bullet Sand$	1.18	323	0.56
8	Brahim et al. (2012)	Tunez	$1.65 - 0.117 \bullet OC - 0.0042 \bullet Clay - 0.0036 \bullet CoreSand + 0.031 \bullet pH$	1.61	348	0.55
9	Benites et al. (2007)	Brazil	$1.56 - 0.0005 \bullet Clay - 0.01 \bullet OC + 0.0075 \bullet BC$	1.36	1,396	0.66
10	Heuscher et al. (2005)	USA, and neighboring countries	$1.685 - 0.198 \bullet OC^{0.5} - 0.0133 \bullet \theta_{1500} + 0.0079 \bullet Clay + 0.00014 \bullet Depth - 0.0007 \bullet Silt$	NR	47,015	0.44

[1] According to del Grosso's modification used in Century Model 4.0 (https://www.nrel.colostate.edu/projects/century/).

[2] Personal communication.

[3] Adams (1973) used $\rho_{b,o}$=0.224 Mg m^{-3} and $\rho_{b,m}$=1.27 Mg m^{-3}.

NR: Not reported.

Sand (%) (fraction in Saxton et al., 1986; and in Rawls et al., 2004). Silt (%), Clay (%) (fraction in Saxton et al. (1986); and in Rawls et al., 2004; and g kg^{-1} in Benites et al., 2007). OC: organic carbon content (%) (g kg^{-1} in Benites et al., 2007). OM: organic matter content (%). CoreSand (%): range [0.2-2] mm. BC: basic cations (cmol kg^{-1}). ρ_b: soil bulk density (Mg m^{-3}). θ_{1500}: soil water content at wilting point (% grav.). Depth (cm).

3.4.   Statistical analysis

The correlation between the variables in the soil dataset was assessed by computing the coefficient of Pearson (r). Additionally, the r^2, RMSE, and the Nash-Sutcliffe Model Efficiency (ME) were used to test the predictive capabilities of the equations developed with the ANN and the existing PTFs. The RMSE and the ME were computed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{1}$$

$$ME = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \tag{2}$$

where y_i and $\hat{y}_i$ are the measured and predicted ρ_b values, respectively; $\bar{y}$ is the mean of measured values; and n is the number of observations. The r^2 is a measure of the strength of the linear relationship between the measurements and predictions and indicates the fraction of the variation that is shared between them (De Vos et al., 2005). The RMSE is a measure of the overall error of the prediction, and the ME is a measure of the accuracy of the prediction and indicates how much the fit is related to the 1:1 line. Ideally, the r^2 and ME should be close to 1 and the RMSE close to 0.

4.   RESULTS AND DISCUSSION

The main results of this study were (i) a description of the soil properties in the sample dataset, (ii) a description of the ANN networks developed and the characterization of the RI of their input parameters, (iii) the development of a new and hierarchical set of equations for estimating ρ_b using the ANN, (iv) a comparison of the performance of the 10 PTFs collected from literature both without calibration and calibrated, and (v) a comparison between the equations developed with the ANN and the calibrated PTFs.

4.1.   Soil properties in the soil data set

The mean ρ_b value in the soil dataset was 1.33 Mg m^{-3} (Table 1). As shown in Fig. 4, the histogram of ρ_b values did not show a normal distribution (p<0.005 by the Anderson-Darling test) and most of the ρ_b were between 1.2 and 1.8 Mg m^{-3} (65% of the samples). The OC content in the samples ranged from 0.00 to 0.13, with a coefficient of variation (CV) of 117% (Table 1); as has been reported in previous studies (Bernoux et al., 1998; Brahim et al., 2012), this high CV value is because OC content is highly variable in soils.
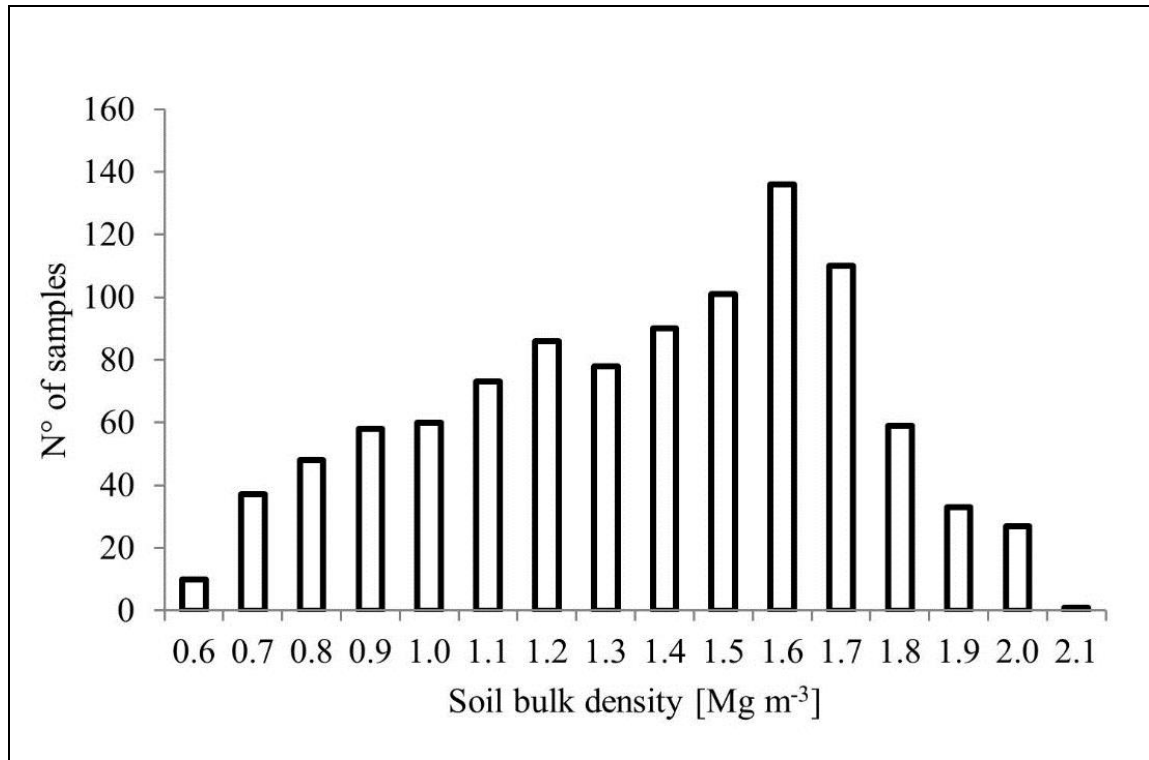
Figure 4: Soil bulk density histogram for the soil samples (n=1,007).

Table 3 shows the matrix of correlation among the soil properties based on the data from Table 1. When analyzing the entire soil sample set (n=1,007), the ρ_b value decreased as the OC content increased (r=−0.58). Despite diverse studies that have reported OC content as the main input parameter explaining the variation in ρ_b, OC content in this dataset explained only 34% of the variation in ρ_b. This low OC predictor score was caused by the low values of OC content (Table 1), which was also shown by Bernoux et al. (1998). In fact, the functions that used OC content as a main (or sole) input were developed in soils with high OC contents. Those studies included Curtis and Post (1964), Adams (1973), Alexander (1980), and Huntington et al. (1989). The study of Leonaviciute (2000) found that ρ_b changes with soil depth, whereas other studies have reported that these two soil properties are independent (Heuscher et al., 2005; Martin et al., 2009). The low Pearson coefficient with soil depth (r=-0.01) endorses the results of Heuscher et al. (2005) and Martin et al. (2009). Additionally, the differences in the ρ_b mean for each soil horizon in the soil dataset were not statistically significant. A low correlation was found between ρ_b and clay, silt, and sand content (r=−0.20 r=−0.37

r=0.38, respectively). In contrast, the correlation was higher with θ_{33} and θ_{1500} than with other soil properties (r=−0.82 and r=−0.79, respectively), the same result as was reported by Patil and Chaturvedi (2012) when they predicted ρ_b in waterlogged soils. This high negative correlation was caused by the fact that more water retention at specific suction requires more available pore space and thus more porosity and, consequently, a small ρ_b.

Table 3. Correlation matrix (with the Pearson coefficient) between soil properties of Table 1 for the entire soil sample set.

Variable	Sand	Silt	Clay	ρ_b	Depth	θ_{33}	θ_{1500}	OC	BC	pH
Sand	1									
Silt	-0.74	1								
Clay	-0.74	0.09	1							
ρ_b	0.38	-0.37	-0.20	1						
Depth	0.01	-0.07	0.06	0.02	1					
θ_{33}	-0.74	0.67	0.45	-0.82	0.00	1				
θ_{1500}	-0.74	0.59	0.54	-0.79	0.05	0.92	1			
OC	-0.31	0.35	0.11	-0.58	-0.39	0.68	0.61	1		
BC	-0.24	0.22	0.15	0.31	0.03	-0.06	-0.13	-0.28	1	
pH	0.14	-0.01	-0.20	0.38	0.20	-0.25	-0.30	-0.35	0.60	1

ρ_b: soil bulk density. θ_{33}: soil water content at field capacity. θ_{1500}: soil water content at wilting point. OC: organic carbon. BC: basic cations.

4.2.   Artificial neural networks

The ANN developed for predicting ρ_b had three steps (pre-processing, neural network, and post-processing). After testing a wide range of numbers of nodes, no statistically significant improvement in the r^2 values was found when using four or more nodes in each type of equation. Therefore, the equations were built with three nodes and the input for equations A to F.

The pre-processing (input variables standardization) was accomplished by using the following equation:

$$xp = (xp_{max} - xp_{min})/(x_{max} - x_{min}) \bullet (x - x_{min}) + xp_{min} \tag{3}$$

where xp is the standardized input variable, xp_{min} and xp_{max} are the minimum and maximum values for xp (equal to -1 and 1, respectively), x is the input value of the variable, and x_{min} and x_{max} are the minimum and maximum values for the input variable. The values for x_{min} and x_{max} are listed in Table 1 for each variable.

The general form of the neural network is presented in Eq. 4. The normalized inputs xp_k (k=1 to K) for the node j (j=1 to 3) are multiplied by the weights IW_{jk} and summed with the bias term b_{1j}. Then, the sigmoid function is applied and multiplied by the layer weights (LW_j) and summed together with the constant b_2 to generate the output a^2 as follows:

$$a^2 = \sum_{j=1}^{3} LW_j \bullet \text{tansig}(\sum_{k=1}^{K} IW_{jk} \bullet xp_k + b_{1j}) + b_2 \tag{4}$$

where the values for LW_j, IW_{jk}, b_{1j}, and b_2 are listed in Table 4 to write and use the equations. Because the output a^2 is standardized, it must be reverse-processed to obtain the predicted values. This is accomplished by using Eq. 5:

$$\hat{y} = (\hat{y}_{max} - \hat{y}_{min})/(a^2_{max} - a^2_{min}) \bullet (a^2 - a^2_{min}) + \hat{y}_{min} \tag{5}$$

where $\hat{y}$ is the predicted ρ_b value, $\hat{y}_{max}$ and $\hat{y}_{min}$ are the maximum and minimum value of $\hat{y}$ (2.1 and 0.56 Mg m^{-3} for this soil set, respectively), a^2_{max} and a^2_{min} are 1 and -1, respectively, and a^2 is the network output of Eq. 4.

Table 4. Values of parameters LW_j, IW_{jk}, b_{1j} and b_2 for the Artificial Neural Network equations.

Type A	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
Sand	-3.916	1.349	-1.472	6.103	0.495	-0.217	-0.164	0.302	0.450	-0.159
Silt	-16.624	-1.141	-2.965							
Clay	18.548	-2.076	0.937							
Type B	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
OC	-3.006	13.775	21.172	-0.314	8.150	20.146	0.225	-0.210	-0.214	-0.179
Type C	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
Sand	-21.503	0.981	2.074	90.575	-0.249	1.605	-0.120	0.465	0.253	-0.318
Silt	-27.942	-0.416	0.021							
Clay	-19.216	1.324	-2.272							
OC	115.213	-2.292	0.138							
Type D	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
Sand	-1,710.493	1.543	-0.704	209.298	-4.553	0.055	-0.143	0.317	0.457	0.109
Silt	673.488	-0.172	-0.361							
Clay	1214.670	0.308	0.347							
OC	1567.943	-5.798	0.342							
pH	525.105	-5.098	4.255							
Type E	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
Sand	6.710	0.706	0.754	-25.336	0.848	3.818	0.193	0.565	0.262	-0.347
Silt	6.402	-0.513	-0.428							
Clay	5.717	0.076	-0.158							
OC	-33.069	-1.489	4.270							
BC	3.246	2.024	-1.945							

Table 4. Continued

Type F	IW			B_1			LW			B_2
k/j	1	2	3	1	2	3	1	2	3	
Sand	-0.541	212.413	0.799	-0.998	-866.738	0.547	1.391	0.144	0.862	-0.176
Silt	-0.362	279.561	0.265							
Clay	0.668	165.583	-1.525							
OC	-0.850	0.137	0.398							
Depth	0.070	109.134	-0.854							
θ_{1500}	-1.828	82.895	2.256							

The RI of the variables showed that in equation A, the three inputs (sand, silt, and clay) had almost the same contribution (RI=34%, 34%, and 32%, respectively). For equation C and D, the most important input was the clay content (RI=35% and 32%, respectively). However, clay had the lowest correlation with ρ_b among the inputs of equation C and D (Table 3), which proved that the effect of clay was relevant but non-linear in these equations. However, for equation E and F, OC content was the most significant input (RI=31% and 22%, respectively) and is consistently with the fact that OC content showed the highest correlation with ρ_b within the inputs of equation E. However, in the case of equation F, OC content has not the highest correlation with ρ_b among their inputs (Table 3), which proved that the effect of OC content was relevant but non-linear in this equation. The results of the RI analysis on input variables indicate that the process of estimating ρ_b cannot be explained for a single variable and cannot be simplified to a correlation analysis.

In terms of network performance, Table 5 shows the ME and RMSE obtained with the ANN for the entire sample dataset and the training, validating and testing sample sets. The relationships between the measured and predicted ρ_b are shown in Fig. 5. These equations were developed with measured ρ_b values in the range of 0.56 to 2.1 Mg m^{-3} (Table 1) and the estimates can be different when using for soils out of this range.

Table 5. Evaluation for equations A through F developed with the Artificial Neural Network techniques.

| | Samples dataset | | | | | | | | | | | |
| | Entire | | | Testing | | | Training | | | Validating | | |
Equation	N	ME	RMSE	N	ME	RMSE	N	ME	RMSE	N	ME	RMSE
A	1,007	0.22	0.32	705	0.20	0.32	151	0.22	0.32	151	0.28	0.31
B	1,007	0.38	0.28	705	0.40	0.28	151	0.39	0.28	151	0.33	0.29
C	1,007	0.44	0.27	705	0.39	0.29	151	0.47	0.26	151	0.36	0.28
D	997	0.49	0.26	697	0.40	0.29	150	0.50	0.25	150	0.49	0.25
E	880	0.55	0.24	616	0.47	0.26	132	0.57	0.23	132	0.56	0.26
F	754	0.72	0.17	528	0.69	0.19	113	0.72	0.17	113	0.76	0.16

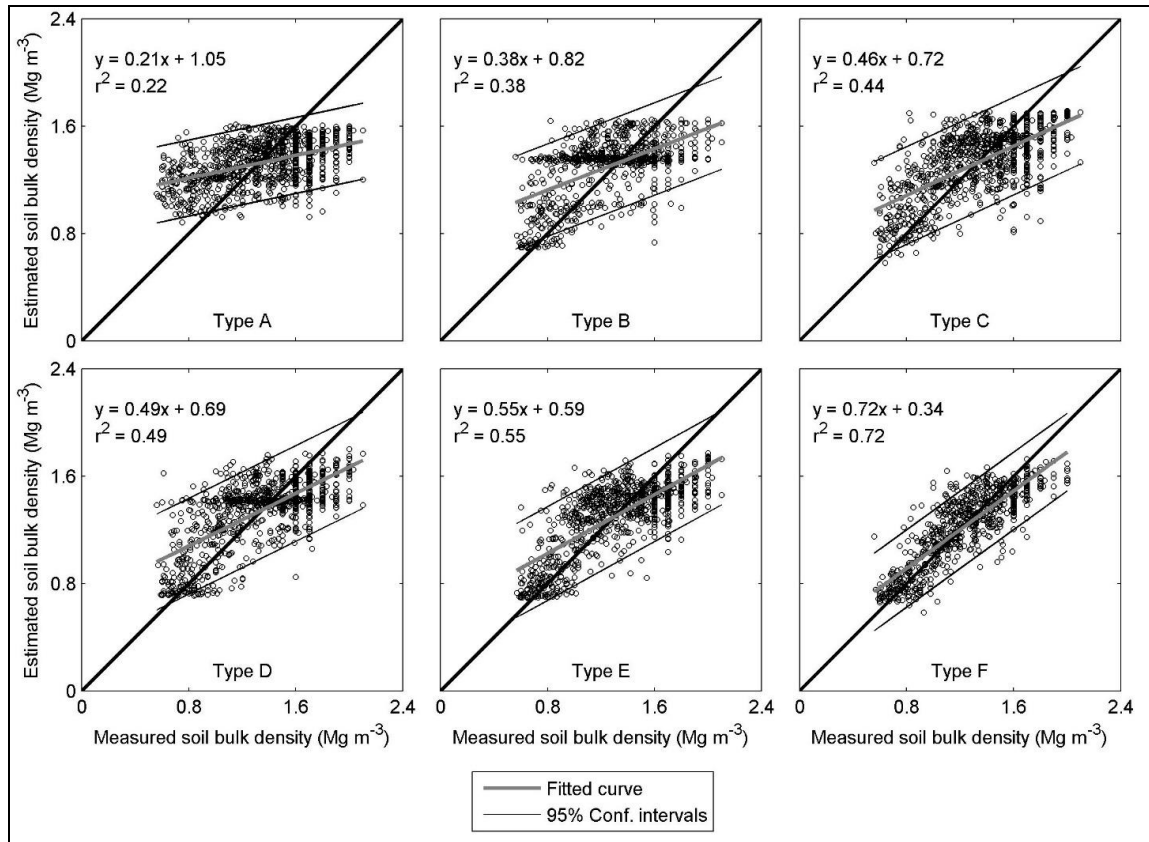All the networks have three nodes. The r^2 is not shown because it is equal to the ME value.

Figure 5: Comparison between measured soil bulk density (Mg m^{-3}) and predicted values using the equations developed with the ANN technique for the entire soil sample set. The lines around the fitted curve are the 95% confidence intervals.

The performance of the equations improved (accuracy and overall error) as the number of inputs increased (from A to F). The lowest performance was found with equation A, which used only sand, silt, and clay content (ME=0.22, RMSE=0.32 Mg m^{-3}). This is explained because sand, silt, and clay were not highly correlated with ρ_b (Table 3). In contrast, equation B, with only OC content, performed better than equation A (ME=0.38, RMSE=0.28 Mg m^{-3}). This result indicates that an equation based on OC (or OM) provides better ρ_b estimates than a soil particle size distribution-based equation, even when applied to soils with low OC (or OM) content such as those used in this study.

Equation C, which used OC, sand, silt, and clay content, had better performance for accuracy but was just slightly better for overall error compared with equation B (ME=0.44, RMSE=0.27 Mg m^{-3}). Moreover, the performance improvement between

equations B and C was smaller compared with the increment between equations A and B. However, a significant improvement was found when pH (equation D, ME=0.49, RMSE=0.26 Mg m^{-3}) and basic cations (equation E, ME=0.55, RMSE=0.24 Mg m^{-3}) were included.

The best estimates were found when using equation F, with the highest ME (0.72) and the lower RMSE (0.17 Mg m^{-3}). Differently from the other equations, equation F uses θ_{1500} and a high correlation was found between ρ_b and θ_{1500} (Table 3).

Based on 332 soil samples, Al-Qinna and Jaber (2013) used ANN to develop an equation for predicting ρ_b based on OC, sand, silt, and clay content (the same as equation C). Using six nodes, their results showed an r^2 and RMSE of 0.63 and 0.11 Mg m^{-3}, and 0.27 and 0.14 Mg m^{-3} in calibration and validation, respectively. Their results had a higher r^2 and a lower RMSE compared with this study. Their soils were more compacted (mean ρ_b = 1.68 Mg m^{-3}) and had less variability (CV of 10%). On the other hand, Xiangsheng et al. (2016) used 495 soil samples as well as ANN to develop three equations for predicting ρ_b. Depending on the inputs, they obtained r^2=0.69 and RMSE=0.15 when using OC content, r^2=0.71 and RMSE=0.14 when clay and silt were added, and r^2=0.71 and RMSE=0.14 when the soil depth was added to the previous variables. Their results showed better r^2 and RMSE values compared with the results presented in this study although the soils used in their study were almost negligible in clay (mean of 0.01). However, their results showed the same pattern as this study; the performance of the equations improved with the number of inputs, even though the increments were very small.

An analysis based on textural classes was also performed with the ANN. All the equations overestimated the ρ_b values when measured ρ_b was low (silt loam, silty clay, and silty clay loam) and underestimated the values when measured ρ_b was high (sand, loamy sand, and sandy clay loam).

4.3.　　　　Evaluation of the previous PTFs

The performance of the 10 previous PTFs without calibration and calibrated is shown in Table 6 for the entire soil dataset. The comparison between the measured and predicted ρ_b values for the calibrated PTFs is shown in Fig. 6. Without calibration, all the equations overestimated ρ_b when the measured value was low. Only the equations developed by Bernoux et al. (1998), Tomasella and Hodnett (1998), and Heuscher et al. (2005) could predict ρ_b values as low as 0.60 Mg m^{-3}. On the other hand, the equations consistently underestimated ρ_b when the measured value was high, and only the Saxton et al. (1986) and Heuscher et al. (2005) equations could predict values as high as 2 Mg m^{-3}.

Table 6. Performance of the tested pedotransfer functions using the entire soils dataset.

PTF	Eq.	Reference	n	Without calibration			Calibrated	
				r^2	ME	RMSE	ME	RMSE
						Mg m^{-3}		Mg m^{-3}
1	A	Saxton et al. (1986)	1,007	0.05	-0.03	0.36	0.17	0.33
2	B	Adams (1973)	1,007	0.36	-0.02	0.36	0.36	0.29
3	C	Tomasella and Hodnett (1998)	1,007	0.36	0.13	0.33	0.39	0.28
4	C	Kaur et al. (2002)	1,007	0.38	-0.51	0.44	0.41	0.28
5a	C	Rawls et al. (2004)-A horizons	167	0.39	0.28	0.24	0.48	0.21
5b	C	Rawls et al. (2004)-All horizons	632	0.26	0.23	0.30	0.32	0.28
6	C	Hollis et al. (2012)	1,007	0.38	0.33	0.29	0.39	0.28
7	D	Bernoux et al. (1998)	997	0.25	-0.25	0.40	0.42	0.27
8	D	Brahim et al. (2012)	901	0.33	0.16	0.34	0.44	0.27
9	E	Benites et al. (2007)	880	0.22	-0.61	0.46	0.41	0.28
10	F	Heuscher et al. (2005)	754	0.43	0.24	0.29	0.67	0.19

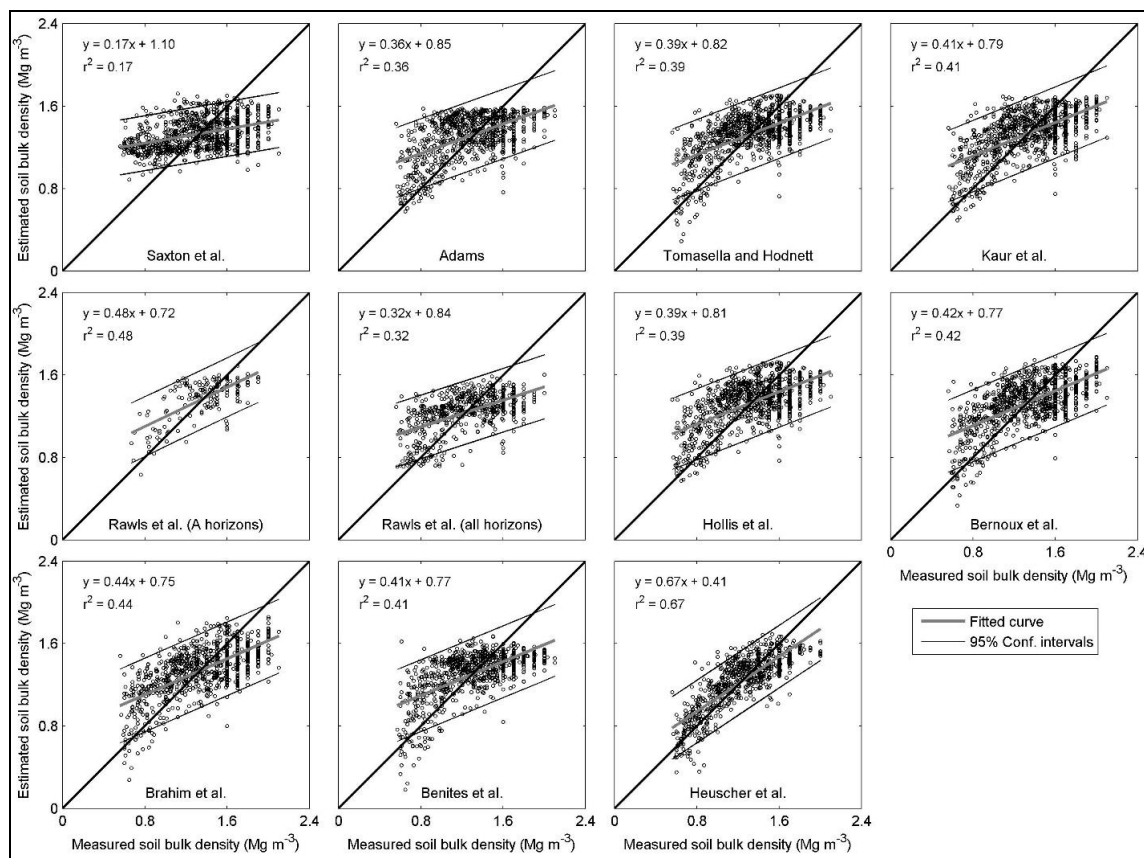The r^2 in the calibrated section is not shown because is equal to the ME value.

Figure 6: Comparison between measured soil bulk density (Mg m^{-3}) and predicted values with the existing pedotransfer functions calibrated for the entire soil sample set. The lines around the fitted curve are the 95% confidence intervals.

The estimates showed an ME between -0.61 and 0.33. The equations developed by Adams (1973), Saxton et al. (1986), Bernoux et al. (1998), Kaur et al. (2002), and Benites et al. (2007) showed negative ME values, which means that it is better to use the average ρ_b value of the soil dataset of this study instead these equations. However, the calibration reduced the difference between the predicted and measured ρ_b values, with an ME between 0.17 and 0.67. The RMSE for these equations without calibration was between 0.29 Mg m^{-3} and 0.46 Mg m^{-3}. With a ρ_b mean of 1.33 Mg m^{-3}, the RMSE was between 19% and 35% of the mean ρ_b. The calibration reduced the overall prediction error to between 0.19 Mg m^{-3} and 0.33 Mg m^{-3}, or between 14% and 25% of the mean ρ_b.

When the equations were calibrated with this new soil dataset, the performance followed the same trend as the equations developed with ANN in section 3.2: the estimates improved with the number of inputs in the equations. The estimates that differed more from the measured ρ_b values were obtained with equation A (Saxton et al., 1986), which used only sand and clay. Equations C (Tomassella and Hodnett, 1998; Kaur et al., 2002; Rawls et al., 2004 and Hollis et al., 2012), which used OC (or OM) and sand, silt and/or clay, had similar accuracy and overall error (Table 6) and provided better estimates than equation A. In terms of ME, equation B (Adams, 1973) was similar to equations C but with a larger overall error as a result of using only OM content as input (RMSE=0.34 Mg m^{-3}).

Compared with equation C, and differently from the relationships developed with the ANN, no significant improvement in the quality of the estimates was observed when pH was added (equation D: Bernoux et al., 1998 and Brahim et al., 2012) or basic cations were added (equation E: Benites et al., 2007). However, the best estimates were obtained with equation F (Heuscher et al., 2005). The ME and RMSE demonstrated that this PTF was the most reliable for this soil dataset when predicting ρ_b. The other PTFs showed some limitations when tested on Chilean soils as reported by Casanova et al. (2016) after testing 10 PTFs in two soils of Central Chile.

In the case of equations C, the PTF developed for the A-horizons (Rawls et al., 2004) worked better for those horizons than when used to predict ρ_b in the entire soil profile. However, when compared with the other equations C, the estimates were similar in RMSE and better in ME. This improvement in accuracy was highly related to complex combinations of variables in the original equation (Table 2).

Moreover, to remove any possible bias related to the number of samples used in the comparison, an additional test was performed for all the PTFs with the same number of soil samples (n=611). The results were not significantly different from those presented in Table 6 when using different number of soil samples. Furthermore, the calibrated form of the PTFs evaluated are shown in Table 7 for practical use.

Table 7. Mathematical form of the calibrated pedotransfer functions developed in this study.

PTF	Source	Eq. $\rho_b =$
1	Saxton et al. (1986)	$(1-(0.708-0.004\bullet Sand+0.076\bullet LOG(Clay)))\bullet 2.65\bullet 0.959$
2	Adams (1973)	$100/((OM/0.181)+((100-OM)/1.586))$
3	Tomasella and Hodnett (1998)	$1.710-0.089\bullet OC-0.004\bullet Silt-0.003\bullet Clay$
4	Kaur et al. (2002)	$\exp(0.332-0.087\bullet OC+0.006\bullet Clay-0.0001\bullet Clay\char94 2-0.003\bullet Silt)$
5	Rawls et al. (2004)	

$1.303+0.223\bullet(0.088+3.085\bullet w+1.075\bullet w^2$

$-4.223\bullet w^3+0.491\bullet y+1.473\bullet w\bullet y+0.290\bullet w^2\bullet y$

$0.333\bullet y^2+0.606\bullet w\bullet y^2+0.084\bullet y^3-0.335\bullet z$

$+0.844\bullet w\bullet z-1.308\bullet w^2\bullet z+0.113\bullet y\bullet z-0.090\bullet w\bullet y\bullet z$

$+0.139\bullet y^2\bullet z-0.219\bullet z^2-0.254\bullet w\bullet z^2-0.058\bullet y\bullet z^2$

$+0.040\bullet z^3)$

where :

$x=-1.2141+4.23123\bullet sand$

$y=-1.70126+7.55319\bullet clay$

$z=-1.55601+0.507094\bullet OM$

$w=-0.0771892+0.256629\bullet x+0.256704\bullet x^2-0.140911\bullet x^3-0.0237361\bullet y$

$-0.098737\bullet x^2\bullet y-0.140381\bullet y^2+0.0140902\bullet x\bullet y^2+0.0287001\bullet y^3$

6	Hollis et al. (2012)	$0.238+\left(\begin{array}{l}1.142\bullet\exp(-0.127\bullet OC)\\+(0.003\bullet Sand)-(0.0007\bullet Clay)\end{array}\right)$
7	Bernoux et al. (1998)	$0.574+0.003\bullet Clay-0.075\bullet OC+0.097\bullet pH+0.005\bullet Sand$
8	Brahim et al. (2012)	$0.639-0.075\bullet OC-0.0006\bullet Clay-0.004\bullet CoreSand+0.119\bullet pH$
9	Benites et al. (2007)	$1.477-0.004\bullet Clay-0.09\bullet OC+0.008\bullet BC$
10	Heuscher et al. (2005)	$1.820-0.160\bullet OC\char94 0.5-0.015\bullet\theta_{1500}-0.0006\bullet Clay-$
		$0.002\bullet Depth-0.001\bullet Silt$

In addition, as shown in section 3.2, an analysis based on textural classes was also performed for the calibrated PTFs. As the equations developed with ANN, the same condition was found. However, equations C and D overestimated the ρ_b values in sandy soils, which indicates that these types of PTFs are less accurate in soils either with extremely high sand or low clay contents (0.94 and 0.01 on average, respectively).

4.4.   Comparison between the ANN and the calibrated PTFs

The performance of the ANN was compared with those of the calibrated PTFs evaluated in section 3.3. This test provided a fair comparison because the constants in both groups of equations were computed with the same soil sample set. The results showed a better ME and RMSE in all the ANN compared with the already existing PTFs. The improvement in ME was between 0.02 and 0.14. However, the increment in the ME in equation B was not statistically significant after the Fisher test (p=0.24) because the ANN technique cannot significantly enhance the quality of estimates with one input (Hecht-Nielsen, 1990).

Equations A, C, D, E, and F provided statistically significant improvements in the ME (p=0.04, p=0.04, p=0.04, p=0.00 and p=0.03, respectively) compared with the previous PTFs, which demonstrates the convenience of using the ANN technique compared with classical multivariable regression analysis. The highest increment in ME was with equation E, which proved the suitability of the use of basic cations as input compared with the use of pH in equation D. However, the improvement in the RMSE was not relevant (between 0.01 Mg m^{-3} and 0.04 Mg m^{-3}) and the error was on the order of the error of the core and clod analytical methods (Blake and Hartge, 1986).

5.   CONCLUSIONS

The use of artificial neural network techniques proved to be a suitable method for building six equations for predicting ρ_b across a wide range of soil types and inputs. The ANN developed in this study were two-layer feed forward networks with 3 nodes, and their capability for predicting ρ_b was related to the number of parameters used in the network. The networks were established with the inputs of sand, silt and clay content (r^2=0.22); organic carbon content (r^2=0.38); sand, silt, clay and organic carbon content (r^2=0.44); sand, silt, clay, organic carbon content and pH (r^2=0.49); sand, silt, clay, organic carbon content and basic cations (r^2=0.55); and sand, silt, clay, organic carbon content, depth and soil water content at wilting point (r^2=0.72). Depending on the field and soil survey data, these ANN are meant to be selected and used hierarchically to predict ρ_b values. The inputs required for the networks are usually available in soil surveys, and the networks can be easily coupled with other soil and hydrology models to provide an alternative for estimating ρ_b when its measured value is not available.

Compared with other existing PTFs developed for predicting ρ_b, the equations developed with ANN enhanced the quality of estimates when they were evaluated with the same type and number of input parameters. This condition demonstrates that this technique is a promising option for relationships for predicting variable soil properties such as bulk density. Although the classical regression relationships are still useful for predicting the soil bulk density because of their simplicity and intuitive formulation, the use of this new set of equations is highly recommended to produce a more robust estimate and reduce the overall error. The set of ρ_b values used in this study came from many soil horizons and pits and covered a wide range of soil types, including cultivated and forest soils, in addition to a variety of climate conditions. Finally, if the data are available, the network based on sand, silt, clay and organic carbon content, soil depth, and soil water at the wilting point should be used with the set of the equations presented in this study to obtain the best estimates.

REFERENCES

Adams, W.A. 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. Journal of Soil Science 24:10-17.

Alexander, E.B. 1980. Bulk densities of California soils in relation to other soil properties. Soil Science Society of America Journal 44:689-692.

Al-Qinna, M.I., and S.M. Jaber. 2013. Predicting soil bulk density using advanced pedotransfer functions in an arid environment. Transactions of the American Society of Agricultural and Biological Engineers. 56(3): 963-976.

Baker, L., and D. Ellison. 2008. Optimisation of pedotransfer functions using an artificial neural network ensemble method. Geoderma 144(1-2): 212–224.

Blake, G.R., and K.H. Hartge. 1986. Bulk density. In: A. Klute, editor, Methods of soil analysis: Part 1. Physical and mineralogical methods. ASA and SSSA, Madison, WI. p. 371-373.

Benites, V.M., P.L.O.A. Machado, E.C.C. Fidalgo, M.R. Coelho, and B.E. Madari. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139:90–97.

Bernoux, M., D. Arrouays, C. Cerri, B. Volkoff, and C. Jolivet. 1998. Bulk Densities of Brazilian Amazon Soils Related to Other Soil Properties. Soil Sci. Soc. Am. J. 62: 743-749.

Bonilla, C. and J. Cancino. 2001. Soil water content estimation using pedotransfer functions. Chilean Journal of Agricultural Research 61(3):326-338. (In Spanish)

Bonilla, C.A., and K.L. Vidal. 2011. Rainfall erosivity in central Chile. Journal of Hydrology 401:126-133.

Bonilla, C. A., and O. I. Johnson. 2012. Soil erodibility mapping and its correlation with soil properties in Central Chile. Geoderma. 189-190:116-123.

Boucneau, G., M. Van Meirvenne, and G. Hofman. 1998. Comparing pedotransfer functions to estimate soil bulk density in northern Belgium. Pedologie Themata. 5:67-70.

Bouyoucos, G.J. 1962. Hydrometer method improved for making particle size analysis of soils. Agronomy Journal. 54:464-465.

Brahim, N., M. Bernoux, and T. Gallali. 2012. Pedotransfer functions to estimate soil bulk density for Northern Africa: Tunisia case. J. Arid Environ. 81:77–83.

Brown, J. D. and Heuvelink, G. B. M. 2006. Assessing Uncertainty Propagation through Physically Based Models of Soil Water Flow and Solute Transport. Encyclopedia of Hydrological Sciences. 6:79.

Calhoun, F.G., N.E. Smeck, B.L. Slater, J.M. Bigham, and G.F. Hall. 2001. Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. Soil Sci. Soc. Am. J. 65:811-819.

Casanova, M., E. Tapia, O. Seguel, and O. Salazar. 2016. Direct measurement and prediction of bulk density on alluvial soils of central Chile. Chilean Journal of Agricultural Research. 76(1):105-113.

CIREN, 1996a. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región Metropolitana. Centro de Información de Recursos Naturales. (In Spanish).

CIREN, 1996b. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región de Libertador B. O'Higgins. Centro de Información de Recursos Naturales. (In Spanish).

CIREN, 1997a. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región de Valparaíso. Centro de Información de Recursos Naturales. (In Spanish).

CIREN, 1997b. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región del Maule. Centro de Información de Recursos Naturales. (In Spanish).

CIREN, 1999. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región del Bío-Bío. Centro de Información de Recursos Naturales. (In Spanish).

CIREN, 2002. Estudio Agrológico. Descripción de suelos. Materiales y símbolos. Región de la Araucanía. Centro de Información de Recursos Naturales. (In Spanish).

Curtis, R.O., and B.W. Post. 1964. Estimating bulk density from organic matter in some Vermont forest soils. Soil Science Society of America Journal. 28:285-286.

Dam, R.F., B.B. Mehdi, M.S.E. Burgess, C.A. Madramootoo, G.R. Mehuys, I.R. Callum. 2005. Soil bulk density and crop yield under eleven consecutive years of corn with different tillage and residue practices in a sandy loam soil in central Canada. Soil and Tillage Research. 84:41-53.

Davis, B.E. 1974. Loss-on-ignition as an estimate of soil organic matter. Soil Science Society of America. Proceedings. 38:150-151.

De Vos, B., M. Van Meirvenne, P. Quataert, J. Deckers, and B. Muys. 2005. Predictive Quality of Pedotransfer Functions for Estimating Bulk Density of Forest Soils. Soil Sci. Soc. Am. J. 69:500-510.

Garson, G.D. 1991. Interpreting Neural-network Connection Weights. Artificial Intelligence Expert 6(4): 46-51.

Hagan, M.T., and M.B. Menhaj. 1994. Training Feedforward Networks with the Marquardt Algorithm. IEEE Trans. Neural Networks 5(6): 989–993.

Han, G-Z., G-L. Zhang, Z-T. Gong, and G-F. Wang. 2012. Pedotransfer Functions for Estimating Soil Bulk Density in China. Soil Sci. 177:158–164.

Hecht-Nielsen, R. (1987). Kolmogorov's Mapping Neural Network Existence Theorem.

In Proceedings of the IEEE First International Conference on Neural Networks, San Diego, CA, USA (11–13).

Hecht-Nielsen, R. (1990). Neurocomputing. Addison-Wesley, Reading, MA.

Heuscher, S. A, C.C. Brandt, and P.M. Jardine. 2005. Using Soil Physical and Chemical Properties to Estimate Bulk Density. Soil Sci. Soc. Am. J. 69:1–7.

Hollis, J.M., J. Hannam, and P.H. Bellamy. 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. Eur. J. Soil Sci. 63:96–109.

Huntington, T.G., C.E. Johnson, A.H. Johnson, T.G. Siccama, and D.F. Ryan. 1989. Carbon, organic matter, and bulk density relationships in a forested spodosol. Soil Science. 148:380-386.

INE, 2007. VII Censo Nacional Agropecuario y Forestal. Instituto Nacional de Estadística, Santiago, Chile. (In Spanish).

INIA, 2006. Métodos de análisis recomendados para los suelos de Chile. Serie Actas INIA N°34, Santiago, Chile. (In Spanish).

Jalabert, S.S.M., M.P. Martin, J.-P. Renaud, L. Boulonne, C. Jolivet, L. Montanarella, and D. Arrouays. 2010. Estimating forest soil bulk density using boosted regression modelling. Soil Use Manag. 26:516–528.

Kaur, R., S. Kumar, and H.P. Gurung. 2002. A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Aust. J. Soil Res. 40:847–857.

Klute, A. 1986. Bulk density. In: A. Klute, editor, Methods of soil analysis: Part 1. Physical and mineralogical methods. ASA and SSSA, Madison, WI. p. 635-656.

Koekkoek, E.J.W., and H. Booltink. 1999. Neural network models to predict soil water retention. Eur. J. Soil Sci. 50: 489–495.

Lagos-Avid M.P., and C.A. Bonilla. 2017. Predicting the particle size distribution of eroded sediment using artificial neural networks. Science of the Total Environment 581–582:833–839.

Leonaviciute, N. 2000. Predicting soil bulk and particles densities by pedotransfer functions from existing soil data in Lithuania. Geografijos Metrastis. 33:317-330.

Martin, M.P., D. Lo Seen, L. Boulonne, C. Jolivet, K.M. Nair, G. Bourgeon, and D. Arrouays. 2009. Optimizing Pedotransfer Functions for Estimating Soil Bulk Density Using Boosted Regression Trees. Soil Sci. Soc. Am. J. 73:485-493.

Merdun, H., Ö. Çınar, R. Meral, and M. Apan. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil Tillage Res. 90(1-2):108–116.

Nanko, K., S. Ugawa, S. Hashimoto, A. Imaya, M. Kobayashi, H. Sakai, S. Ishizuka, S. Miura, N. Tanaka, M. Takahashi, and S. Kaneko. 2014. A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. Geoderma. 213:36–45.

Olden, J.D., and D.A. Jackson. 2002. Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling. 154(1-2): 135–150.

Pachepsky, Y. A., Rawls, W.J. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. Soc. Am. J. 63:1748-1757.

Patil, N.G., and A. Chaturvedi. 2012. Estimation of bulk density of waterlogged soils from basic properties. Arch. Agron. Soil Sci. 58:499–509.

Prévost, M. 2004. Predicting Soil Properties from Organic Matter Content following Mechanical Site Preparation of Forest Soils. Soil Sci. Soc. Am. J. 68:943-949.

Rawls, W.J., 1983. Estimating soil bulk density from particle size analysis and organic matter content. Soil Science. 135:123-125.

Rawls, W.J., A. Nemes, and Y. Pachepsky. 2004. Effect of soil organic carbon on soil hydraulic properties. Dev. Soil Sci. 30(C):95–114.

Saxton, K.E., W.J. Rawls, J.S. Romberger, and R.I. Papendick. 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50:1031-1036.

Schaap, M.G., F.J. Leij, and M.T. Van Genuchten. 2001. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology. 251:163 – 176.

Sheela, K. G. and Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. Mathematical Problems in Engineering. 2013, 1-11.

Suuster, E., C. Ritz, H. Roostalu, E. Reintam, R. Kõlli, and A. Astover. 2011. Soil bulk density pedotransfer functions of the humus horizon in arable soils. Geoderma 163:74–82.

Tamari, S., J.H.M. Wösten, and J.C. Ruiz-Suárez. 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. Soil Science Society of America Journal. 60(6): 1732–1741.

Tomasella, J., and M.G. Hodnett. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci. 163:190–202.

Tremblay, S., R. Ouimet, and D. Houle. 2002. Prediction of organic carbon content in upland forest soils of Quebec, Canada. Can. J. For. Res. 32(5): 903–914.

Vasiliniuc, I., and C.V. Patriche. 2015. Validating soil bulk density pedotransfer functions using a Romanian dataset. Carpathian Journal of Earth and Environmental Sciences. 10:225–236.

Walkley, A.J., Black, I.A. 1934. Estimation of soil organic matter carbon by the chromic acid titration method. Soil Science. 37:38-39.

Wösten, J.H.M., Y.A. Pachepsky, and W.J. Rawls. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251:123 – 150.

Xiangsheng, YI, LI Guosheng, and YIN Yanyu. 2016. Pedotransfer functions for estimating soil bulk density: a case study in the Three-River headwater Region of Qinghai Province, China. Pedosphere 26(3):362-373.