Catena 158 (2017) 194–200

Contents lists available at ScienceDirect

Catena

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

Modeling soil cation exchange capacity in multiple countries

Yones Khaledian[a], Eric C. Brevik[b,*], Paulo Pereira[c], Artemi Cerdà[d], Mohammed A. Fattah[e], Hossein Tazikeh[f]

[a] Department of Agronomy, Iowa State University, Ames, IA, USA
[b] Department of Natural Sciences, Dickinson State University, Dickinson, ND, USA
[c] Environmental Management Center, Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania
[d] Soil Erosion and Degradation Research Group, Department of Geography, University of Valencia, Valencia, Spain
[e] Department of Soil and Water Science, Faculty of Agricultural Sciences, University of Sulaimani, Iraq
[f] Department of Soil Sciences, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

ARTICLE INFO

Keywords:
Artificial neural networks
Principal components regressions
Particle swarm optimization
Pedotransfer functions
MONANOVA

ABSTRACT

Cation exchange capacity (CEC), as an important indicator for soil quality, represents soil's ability to hold positively charged ions. We attempted to predict CEC using different statistical methods including monotone analysis of variance (MONANOVA), artificial neural networks (ANNs), principal components regressions (PCR), and particle swarm optimization (PSO) in order to compare the utility of these approaches and identify the best predictor. We analyzed 170 soil samples from four different nations (USA, Spain, Iran and Iraq) under three land uses (agriculture, pasture, and forest). Seventy percent of the samples (120 samples) were selected as the calibration set and the remaining 50 samples (30%) were used as the prediction set. The results indicated that the MONANOVA (R^2 = 0.82 and Root Mean Squared Error (RMSE) = 6.32) and ANNs (R^2 = 0.82 and RMSE = 5.53) were the best models to estimate CEC, PSO (R^2 = 0.80 and RMSE = 5.54) and PCR (R^2 = 0.70 and RMSE = 6.48) also worked well and the results were very similar to each other. While the most influential variables for the various countries and land uses were different and CEC was affected by different variables in different situations, clay (positively correlated) and sand (negatively correlated) were the most influential variables for predicting CEC for the entire data set. Although the MANOVA and ANNs provided good predictions of the entire dataset, PSO gives a formula to estimate soil CEC using commonly tested soil properties. Therefore, PSO shows promise as a technique to estimate soil CEC. Establishing effective pedotransfer functions to predict CEC would be productive where there are limitations of time and money, and other commonly analyzed soil properties are available.

1. Introduction

Soils are an important source of services (filtration of water, exchange of gases with the atmosphere, foundation material for construction) and resources (food, fiber, fuel, clay for construction, medications) for humans (Mol and Keesstra, 2012; Brevik and Sauer, 2015). Soils play a pivotal role in the Earth system to manage the hydrological, erosional, geochemical, and biological cycles (Smith et al., 2015; Willaarts et al., 2016) which makes sustainable soil management and conservation necessary to achieve a sustainable world (Keesstra et al., 2016). Soil conservation includes the maintenance of soil fertility and quality. To determine soil quality, we need a large amount of data that is difficult to obtain due to a lack of measurements, particularly in developing regions, because of their high cost (Masto et al., 2015; Van Leeuwen et al., 2015; Van Hall et al., 2017). The development of easy, accurate and low cost ways to determine soil properties important in the evaluation of soil quality to improve the management of the soil system in areas where data is scarce or non-existent is needed (Costa et al., 2015; Pulido et al., 2015).

Cation exchange capacity (CEC) is a measure of the ability of a soil to hold and exchange cations (Saidi, 2012). Determining soil quality requires identification of soil properties that are important in a soil's ability to carry out its various functions as well as being responsive to changes in land use and land management (Paz-Ferreiro and Fu, 2016; Zolfaghari et al., 2016). CEC plays an important role in soil quality (Brevik, 2009; Khaledian et al., 2016a; Taghizadeh-Mehrjardi, 2016). CEC can be influenced by soil physical (e.g., soil texture), chemical (e.g., pH, mineralogy), and biological (e.g., soil organic matter)

* Corresponding author.
E-mail addresses: ykh@iastate.edu (Y. Khaledian), Eric.Brevik@dickinsonstate.edu (E.C. Brevik), paulo@mruni.eu (P. Pereira), artemio.cerda@uv.es (A. Cerdà), muhammad.fattah@univsul.edu.iq (M.A. Fattah).

http://dx.doi.org/10.1016/j.catena.2017.07.002
Received 7 January 2017; Received in revised form 25 May 2017; Accepted 3 July 2017
0341-8162/

characteristics (Mukherjee and Zimmerman, 2013); therefore, CEC integrates aspects of all three of the indicator groups commonly used to determine soil quality (Brevik, 2009) and there are strong positive relationships between CEC and the soil colloidal fraction (Kweon et al., 2013). However, the conventional analytical methods used for measuring CEC such as the sodium saturation method (Chapman, 1965) are time and money consuming (Busenberg and Clemency, 1973; Seybold et al., 2005; Yukselen and Kaya, 2008).

Pedotransfer functions (PTF) are developed to use data from soil surveys or other sources to predict soil properties that are more difficult to determine in terms of cost and/or time and are therefore determined less frequently than more easily determined properties such as texture, pH, and organic carbon (OC) (McDonald, 1998; Borggaard et al., 2004; Reidy et al., 2016). These functions are useful to help scientists estimate properties like CEC in order to predict the spatial distribution of these important soil properties. Pedotransfer functions have often been developed using multiple linear regression models (Cornelis et al., 2001; Sequeira et al., 2014).

There are a number of studies on statistical methods to estimate soil parameters and assist in environmental management (i.e., McBratney and Odeh, 1997; Chen et al., 2014; Khaledian et al., 2016a; Hosseini et al., 2017) including CEC using general linear models (Seybold et al., 2005), multiple linear regression (Shabani and Norouzi, 2015; Khaledian et al., 2016b), PLS and stepwise regression (Khaledian et al., 2016a), adaptive network-based fuzzy inference system (ANFIS) and artificial neural networks (ANNs) (Ghorbani et al., 2015), and genetic expression programming (GEP) and multivariate adaptive regression splines (MARS) (Emamgolizadeh et al., 2015). However, these studies have not modeled CEC using advanced statistical methods such as algorithms (particle swarm optimization (PSO) and monotone regression tool (MONANOVA)) and compared the performance of these methods, which would represent an advance in the current knowledge.

In spite of the fact that estimation of ecosystem changes using statistical methods is a priority for soil and environmentalist scientists, selecting the correct and most productive methods is still an area that needs additional study. Hence, the goals of this study were to estimate soil CEC with pedo-transfer functions developed using intelligent (PSO, ANNs, and MONANOVA) and regression (PCR) methods, and compare these methods in order to find the best and most productive model(s). In doing so, we could determine the most effective soil properties for modeling CEC in the various countries and land uses evaluated.

2. Materials and methods

2.1. Study areas and soil sampling

Soil samples were collected from three different land uses: agriculture, pasture, and forest lands, in four countries (the USA, Spain, Iran and Iraq) with soil samples taken to a depth of 100 cm. Therefore, large ranges in soil chemical and physical properties were expected.

2.2. Physical and chemical analysis

Soil CEC was determined using the sodium acetate (NaOAc) method (Chapman, 1965). OC was determined using a wet combustion method (Nelson and Sommers, 1982). Soil pH was determined using a pH meter (model WTW 7110) and a 1:1 water/soil suspension (McLean, 1982). Calcium carbonate equivalent (CCE) was measured by the titration method (Page et al., 1982). Soil texture was determined by the Bouyoucos hydrometer method (Gee and Bauder, 1986).

2.3. Statistical analysis

Descriptive analyses such as average, minimum, maximum, standard deviation, and correlation (Pearson) were carried out using SPSS Version 16.0 (IBM Corporation, Armonk, NY). The data (170 soil samples) were divided into a calibration data subset (70%, 120) and prediction data subset (30%, 50). In order to model the data in various countries and land uses, a leave-one-out full cross-validation (LOOV) was used for validation of the models.

Data subsets were used for determining the performance of MONANOVA, PSO and PCR. Matlab (MathWorks, Natick, MA) was used to analyze PSO and XLSTAT (Addinsoft, Paris, France) was used for MONANOVA and PCR.

2.4. ANNS, MONANOVA, PCR, and PSO

Artificial neural networks (ANNs) are a computational model that is part of machine learning. ANNs depend on a large collection of associated artificial neurons similar to axons in a biological brain. In fact, the objective of ANNs is to solve the issues in the same way that human brains do. ANNs focus on not only biological processes in the brain but also neural networks' application to artificial intelligence (McCulloch and Pitts, 1943). In this study, the network was designed with 10 and 1 nodes in the input and output layers, respectively. The minimum possible difference between real and estimated data was aimed in various layers during the training process.

MONANOVA is the combination of a monotonic transformation of responses to a linear regression, like OLS (ordinal linear regression), as a method to improve the linear regression results. MONANOVA works based on the ALS (alternating least squares) algorithm. Its principle consists of alternating between a conventional estimation, such as linear regression or ANOVA, and a monotonic transformation of the dependent variables. The MONANOVA algorithm was introduced by Kruskal (1965).

Particle swarm optimization (PSO) was invented from research into simulating social behavior as an evolutionary technique which makes few or no assumption regarding the issue optimizing and can seek in a very large space. PSO optimizes a problem iteratively, improving a solution by considering a given measure of quality. It solves an issue by a population of particles, or candidate solutions, that these particles move around in the search-space based on simple mathematical formula in different positions and velocities. Each particle's movement is affected by its local best known position and updated as better positions are found by the other particles. Therefore, swarms move toward the best solutions (Kennedy and Eberhart, 1995).

PCR is a regression analysis technique that works based on principal component analysis (PCA). PCR regresses the response (the outcome) on a set of covariates (as predictors) based on a standard linear regression model; however, it uses PCA for predicting the unknown regression coefficients. Furthermore, one of the major uses of PCR is to overcome the multicollinearity problem (Jolliffe, 1982).

To find the best model to estimate soil CEC, the accuracy indicators R^2 and RMSE were used. High coefficient of determination values (R^2) and low RMSE values demonstrate that the models are capable.

3. Results

Descriptive statistics regarding CEC in the study areas are shown in Table 1. The lowest mean CEC values were found in Spain and the highest mean values were found in the agricultural soils of Iran (38.89 cmol/kg) and Iraq (36.7 cmol/kg), followed by CEC values in forest lands (37.55 cmol/kg) in Iran (Table 1). The pasture soils investigated in the USA and Iran had similar CEC values, 26.56 and 27.92 cmol/kg, respectively (Table 1). In terms of the other soil characteristics in different land uses and countries, respectively, the highest levels of clay and silt were measured in forest and agricultural soils, respectively, and Iran, whereas the highest percentage of sand was found in pastures in the USA and Spain. The mean CCE in agricultural and Spanish soils was twice that in the other land uses and countries. Pasture soils had the highest amount of OC. The variation in pH was not remarkable, as in all lands and countries the value was around 7 (Table 1).