Applied Soil Ecology 114 (2017) 123–131

Contents lists available at ScienceDirect

Applied Soil Ecology

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

Comparison of multiple statistical techniques to predict soil phosphorus

Mehdi Hosseini[a,*], Sareh Rajabi Agereh[b], Yones Khaledian[c], Hossein Jafarzadeh Zoghalchali[d], Eric C. Brevik[e], Seyed Ali Reza Movahedi Naeini[b]

[a] Department of Soil Science, Tarbiat Modares University, Tehran, Iran
[b] Department Soil and Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Iran
[c] Department of Agronomy, Iowa State University, IA, USA
[d] Department of Soil Science, Mazandaran Agricultural Sciences and Natural Resources Research and Education Center, Iran
[e] Department of Natural Sciences, Dickinson State University, Dickinson, ND, USA

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form 1 February 2017
Accepted 8 February 2017
Available online 23 March 2017

Keywords:
Soil phosphorus
Soil organic matter (SOM)
Genetic algorithm (GA)
Partial least squares (PLS)
Artificial neural network (ANN)

ABSTRACT

Soil scientists have tested various models to predict soil physical, chemical and biological properties over the last few decades. Determination of soil phosphorus (P) in soils is difficult due to the sensitivity of its measurement, especially in developing nations because of lack of sufficient facilities and limitation of financial resources. In this study, advanced statistical methods (intelligent and regression models) were evaluated, such as genetic algorithm (GA), artificial neural network (ANN), fuzzy inference system (FIS), adaptive neuro-fuzzy inference system (ANFIS), partial least squares (PLS), principal components regression (PCR), ordinary least squares (OLS) and multiple regression (MR), to determine the best model to predict P. This research was carried out at Mazandaran Research Center of Agricultural and Natural Resources, Sari, Iran. Four properties of soils, clay, sand, soil organic matter (SOM) and pH were presented to the models as independent parameters to predict soil P. Such advanced quantitative models have never been compared with each other in order to find the best model for prediction. The results revealed that PLS (among regression models) and GA and ANN (among intelligent models) are promising approaches for the estimation of soil P with higher R^2 and value account for (VAF) and lower mean absolute percentage error (MAPE) and root mean square error (RMSE) compared to the other models. The ANN model predicted soil P more accurately than the other models with $R^2 = 0.912$ and RMSE = 4.019. The GA and PLS models both provided formulas to predict soil P with good fits. Results of the sensitivity analysis showed that SOM was a more effective factor predicting soil P relative to the other variables and SOM provides an important source of P in soil.

1. Introduction

Phosphorus (P) is one of the nutrients that determines the function and primary productivity of terrestrial ecosystems (Elser et al., 2007; Smith et al., 2015) and can be one of the most limiting for crop production. One of the main roles of P in living organisms is in the transfer of energy. Organic compounds that contain P are used to transfer energy from one reaction to drive another reaction within cells (Turner et al., 2013). Adequate P availability for plants stimulates early plant growth and hastens maturity. Soil P changes and dynamics are also influenced by complex interactions between physical, chemical and biological processes that potentially occur within the rhizosphere (Hinsinger, 2013; Messiga et al., 2015).

Pedo-transfer functions are useful for helping environmental agencies estimate soil P in order to establish strategies to control this important element, especially in Third World nations (McDonald, 1998; Borggaard et al., 2004). Pedo-transfer functions consist of equations or sets of equations that allow estimation of the value of a soil property from others that are simpler and faster to determine. Usually, pedo-transfer functions are developed from multiple linear regression models; however, more recently, these functions have also been developed by using the artificial neural networks (ANN) approach, which are capable of identifying non explicit relationships among the variables of a dataset (Cagliari et al., 2011).

* Corresponding author.
E-mail address: mehdi.h.2009@gmail.com (M. Hosseini).

http://dx.doi.org/10.1016/j.apsoil.2017.02.011

A genetic algorithm (GA) models the process of natural selection and is commonly used for optimization of objective function by means of the parameters space coding. This technique has been applied to address complex problems in various scientific disciplines following an iterative process to maximize or minimize objective function (Momenbeik et al., 2010). The GA can obtain acceptable results using operators, namely reproduction, crossover and mutation. GA has been used for soil research such as optimizing simulated fertilizer additions (Cropper and Comerford, 2005), modeling the angle of shearing resistance of soils using soft computing systems (Kayadelen et al., 2009; Levasseur et al., 2008), and predicting soil–water characteristic curves using genetic programming (Johari et al., 2006, 2010).

Intelligent methods have also been used in soil research. This has included the prediction of swell potential of clayey soils (Yilmaz and Kaynar, 2011) using multiple regression, ANN, and adaptive neuro-fuzzy inference system (ANFIS) models, ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neuro-fuzzy models (Padmini et al., 2008), a fuzzy logic-based approach to assess imprecision of soil water contamination modeling (Freissinet, 1998), prediction of soil properties using fuzzy membership values (Zhu, 2010), fuzzy knowledge-based model for prediction of soil loosening and draft efficiency in tillage (Marakoglu and Carman, 2010), fuzzy sets in soil classification (McBratney and Gruijter, 1992) and mapping (Odeh et al., 1992; Taghizadeh-Mehrjardi et al., 2016), neural network models for predicting organic matter content in soils (Ingleby and Crowe, 2001), comparison of artificial neural network and regression pedo-transfer functions for prediction of water retention and saturated hydraulic conductivity (Merdun et al., 2006).

Even though there are a myriad of studies on multivariate statistical methods to predict soil parameters and assist in environmental management (i.e., McBratney and Odeh, 1997; McBratney et al., 2000; Ruggieri et al., 2011; Chen et al., 2014; Pereira et al., 2017), only a few studies have assessed regression methods such as partial least squares regression (PLSR), principal components regression (PCR), ordinary least-squares regression (OLS) and multiple regression (MR) to predict soil properties (e.g., Khaledian et al., 2017a,b). PLSR, for instance, which is accomplished via a small set of latent variables (the components), is a powerful technique for multivariate calibration when the data are highly correlated (Li et al., 2002). In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regressors (Jolliffe, 1982). Moreover, in statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the differences between the observed responses in some arbitrary dataset and the responses predicted by the linear approximation of the data (Hayashi., 2000). This research seeks to enrich the soil science literature with an analysis of transfer functions using new approaches that have the potential to reduce the cost and efforts required for predicting important soil parameters.

The objectives of this study were to estimate soil P using transfer functions employing intelligent (GA, ANFIS, ANN, and FIS) and regression (PCR, PLSR, OLS and MR) methods as new approaches and compare them to find the best model for predicting P and to determine the most effective factors related to P using sensitivity analyses.

2. Materials and methods

2.1. Study area

Soil samples were taken from agricultural lands with wheat crops around Sari City, in Mazandaran province northeast Iran, located between 53°23′55″- 53°31′10″E and 36°59′58″–36°57′11″N. The soils were classified as Haplic Luvisols and Gleyic Cambisols according to WRB (IUSS working group WRB, 2014). The soils in the study area developed in alluvium and are deep (around 1 m) with brown to very dark gray colors. The soil texture is loam to clay loam with illite clay and granular or blocky aggregates. The average annual temperature and precipitation are 23.5 °C and 2661 mm respectively. The elevation ranges from 400 to 600 m above sea level (Mazandaran Regional Water Company, 2017).

2.2. Soil sampling

Soil samples were collected in September 2015. Soils were sampled by stratified random sampling, stratified by physiographic units, slope, and aspect. Soil samples were taken to a depth of 30 cm using a soil probe (diameter 2 cm). In total, 613 soil samples were taken from sites with a northern aspect, to eliminate aspect affects as a variable (Lozano-García et al., 2016). Samples were also obtained from similar altitudes during the study.

2.3. Physical and chemical analysis

Clay, sand, organic matter and pH were independent variables and soil P was the dependent variable. Soil P was determined using the method of Olsen et al. (1954). Particle size distribution was determined by the Bouyoucos hydrometer method (Gee and Bauder, 1986) and soil texture was classified according to Schoeneberger et al. (2012). Organic carbon (OC) was determined using a wet combustion method (Nelson and Sommers, 1982). P

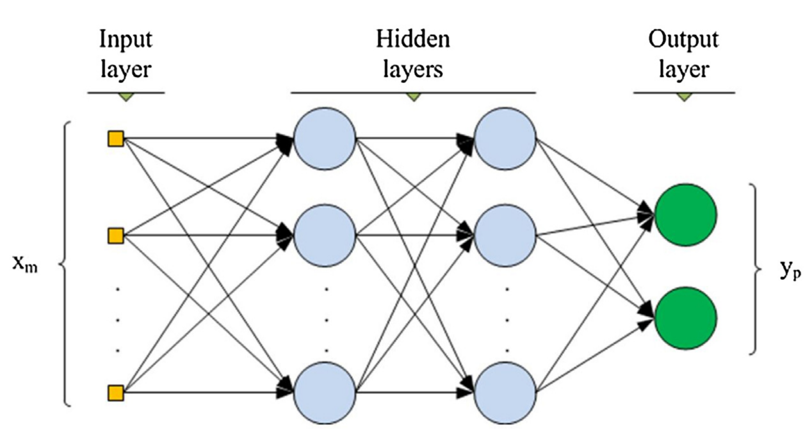

Fig. 1. A simple multilayer of artificial neural network (ANN) configuration.