Computers and Electronics in Agriculture 141 (2017) 327–339

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Using soil easily measured parameters for estimating soil water capacity: Soft computing approaches

Jalal Shiri [a], Ali Keshavarzi [b,*], Ozgur Kisi [c], Sepideh Karimi [a]

[a] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[b] Laboratory of Remote Sensing and GIS, Department of Soil Science, University of Tehran, P.O. Box: 4111, Karaj 31587-77871, Iran
[c] School of Natural Sciences and Engineering, Ilia State University, Tbilisi, Georgia

ARTICLE INFO

Article history:
Received 3 June 2017
Received in revised form 8 August 2017
Accepted 9 August 2017

Keywords:
Soft computing models
K-fold testing
Soil parameters
Soil moisture

ABSTRACT

The current study examines the applicability of six different soft computing approaches, gene expression programming (GEP), neuro-fuzzy (NF), support vector machine (SVM), multivariate adaptive regression spline (MARS), random forest (RF), and model tree (MT) techniques in modeling two important soil water capacity parameters, field capacity (FC) and permanent wilting point (PWP). Geometric mean particle-size diameter (dg), soil bulk density (BD), clay and silt obtained from 192 soil samples were introduced as input variables to the applied techniques and k-fold testing procedure was used for better comparison of the soft computing models. The best accuracy was provided by the NF models followed by the GEP, while the MT approach gave the worst estimates. The performances accuracies of the soft computing models in estimation of PWP parameter were higher than those in the FC estimation. Further, the soft computing approaches were compared with the traditional multi-variable linear regression (MLR) as well as the previously developed pedotransfer functions (PTFs) and the better FC and PWP estimates which confirms the superiority of the soft computing approaches. The NF model increased the performance of the best PTF (Aina-Periaswamy) by 33% with respect to *GMER* in FC estimation while the *SI* statistics of the best PTF (Ghorbani-Homaee) was decreased by 50% using the soft computing model. The performance of the best PTF (Aina-Periaswamy) with respect to *GMER* was increased by 74% in PWP estimation while the *SI* statistics of the best PTF (Dijkerman) was decreased by 99% using the soft computing model.

1. Introduction

Information on soil hydraulic properties, e.g. soil water content is very necessary in water and solute transport as well as heat and mass transfer in soils (Cornelis et al., 2001). Accurate knowledge about the soil available water capacity (AWC) is very important in various environmental issues including irrigation scheduling, land drainage and reclamation, analyzing soil biologic activity, surface runoff simulation, determining leaching requirement/fraction and crop growth simulation as well as different biophysical models (Rab et al., 2011). Since soil serves as a water circulator, the precise information on its moisture content will be very necessary for better managing the fertilizers application so that no excess runoff of these materials (which would be of high risk for the surface water environments) can be produced.

The AWC is defined as the difference between field capacity (FC) and permanent wilting point (PWP) (Waller and Yitayew, 2016). FC is the amount of soil moisture content held by the soil after the gravitational water was drained from the soil. It is indeed the bulk moisture content retained in the soil at −0.33 bar of hydraulic head (Veihmeyer and Hendrickson, 1931). PWP is defined as a minimum moisture content of a soil which is needed for the crop survival and if the water content decreases lower than PWP, a plant wilts and can no longer recover itself (Veihmeyer and Hendrickson, 1928). In-situe measurement of the FC and PWP moisture contents is very costly and time consuming, so numerous investigations have tried to relate these points to soil easily measured variables (Botula et al., 2012). A survey of the literature shows that soil easily measured variables, e.g. geometric mean particle-size diameter (dg), soil bulk density (BD), geometric standard deviation of soil particles (σ_g) and soil separates (clay, silt, sand) have been used to estimate soil FC and PWP (e.g. Aina and Periaswamy, 1985; Dijkerman 1988; Rab et al., 2011; Mohanty et al., 2015). Meanwhile, the soft computing approaches have been also applied for mapping the

* Corresponding author.
E-mail address: alikeshavarzi@ut.ac.ir (A. Keshavarzi).

http://dx.doi.org/10.1016/j.compag.2017.08.012
0168-1699/

input-output relationships between the FC, PWP and the soil easily measured variables.

Borgesen and Schaap (2005) applied neural networks (NN) to estimate soil water content at different pressures and found that introducing soil organic matter and BD as input vectors improves the modeling accuracy. Merdun et al. (2006) applied NN and multi-variable linear regression (MLR) techniques for estimating soil FC and PWP using 195 soil samples and found that the soil BD and dg are the most influential parameters on FC and PWP. Ahmad et al. (2010) utilized remote sensing data for estimating soil moisture through support vector machine (SVM) technique and found that SVM model performs better than NN and MLR models. Ostovari et al. (2015a) applied Mamdani fuzzy inference system and regression tree techniques for estimating FC using 210 soil samples and introduced the soil clay content, BD and dg as input parameters. The obtained results showed the regression tree's superiority to the fuzzy system. Ostovari et al. (2015b) applied MLR technique to relate the soil FC and PWP to the soil easily measured variables using 255 soil samples and confirmed the superiority of their developed regression-based relations to the other published relations. Based on their results, FC and PWP are mainly affected by the clay and dg. The literature review by the authors showed that there are only limited applications of the soft computing models for modeling soil FC and PWP. Nevertheless, most of the existing literatures have applied a single data set assignment, where the developed models have been trained using a part of the available data and tested using the rest of the available patterns, which might lead to partially valid results (Shiri et al., 2014a, 2014b). The present paper will focus on application of the gene expression programming (GEP), neuro-fuzzy (NF), SVM, multivariate adaptive regression spline (MARS), random forest (RF), and model tree (MT) techniques for estimating these points. Further, a multi-variable linear regression model will be applied and compared with the soft computing techniques. The most robust k-fold testing data assessing scenario will be applied for training and testing the models, where all the available input-target patterns are involved in both the training and testing stages, so there would be no "unseen" part of the data (Roushangar et al.,2014; Shiri et al., 2015, 2017).

2. Materials and methods

2.1. Study area and used data

Data from Mohr plain, Fars province, located in Southwest Iran [between the latitudes of 27°25′N to 27°59′N and longitudes of 52°21′E to 53°05′E with an area about 1900 km^2] were utilized in the current paper for establishing and evaluating the applied models. Fig. 1 shows the geographical position of the studied area. The main land uses are pastures and irrigated farming across the Mehran River.

After preliminary studies of topographic maps (1:25,000), study location was appointed. A simple random sampling scheme was designed using ArcGIS 10.2.2 software for an appropriate determination of soil sampling areas to consider spatial variation of the parameters affecting the field capacity (FC) and permanent wilting point (PWP) in the study region.

A total of 250 soil samples were obtained from two-first vertical depths (0–30 and 30–60 cm depth) of 125 representative soil profiles. In order to investigate the relation between FC and PWP with easily measurable properties and complete the objectives of this study, out of 192 soil samples from two-first vertical depths were selected randomly to design this research.

Depths were assigned to a soil textural class determined by the fractions of each soil separates (sand, silt, and clay) presence in a soil as indicated by the USDA textural triangle (Schoeneberger et al., 2002).

The sampling sites were designed to cover equally the entire area and to incorporate different soil and land use types. The collected disturbed soil samples were air dried, crushed and sieved using 2 mm sieve size. Large plant material and pebbles were separated and discarded.

Rates of clay (<0.002 mm), silt (0.002–0.05 mm), and sand (0.05–2 mm) particles were measured via sieving and sedimentation technique (Gee and Bauder, 1986). The clod method (Blake and Hartge, 1986) was utilized for determining bulk density (BD) with triple replications. The moisture contents at field capacity and wilting point were determined with a pressure plate apparatus at −33 and −1500 kPa, respectively (Cassel and Nielsen, 1986). Water saturation percentage and calcium carbonate equivalent (CCE) were determined using standard methods (Sparks et al., 1996). The dg (mm) and σg were calculated based on three particle size fractions (clay, silt, and sand content) as (Shirazi and Boersma, 1984):

$$d_g = \exp\{0.01[P_{sand} \cdot \mathrm{Ln}(d_{sand}) + P_{silt} \cdot \mathrm{Ln}(d_{silt}) + P_{clay} \cdot \mathrm{Ln}(d_{clay})]\} \quad (1)$$

Table 1 summarizes the statistical parameters of the used data set. From the table, it can be seen that the PWP has negative high skewed distribution (Skewness = −1.279). Differences between the maximum and minimum values are high for the FC and PWP (37.030% for FC and 18.064% for PWP). The geometric mean particle-size diameter (dg) presents the highest variability in terms of the coefficient of variations, skewness and kurtosis (1.752, 5.04, and 30.15, respectively). Among the soil separates, silt and sand present the maximum and minimum variations, respectively.

Table 2 sums up some previously published pedotransfer functions of FC and PWP estimation. As can be seen from these functions, soil separates (clay, sand and silt), BD and dg have been generally used for estimating FC and PWP. Alike to these functions and based on statistical analysis of the available data (not presented here), it was found that the soil clay, silt, BD and dg are the most influential parameters on FC and PWP, so they were utilized as input parameters of the applied soft computing models.

2.2. Data splitting and model assessment

A k-fold testing data assignment procedure was adopted here to feed the applied models with the input-target matrixes, so the complete data was divided into 10 subsets and the models were trained and tested each time using a portion of available patterns. Using this procedure, all the available input-target patterns were seen by the models for constructing the final estimation model. Accordingly, the GEP, NF, SVM, MARS, RF and MT models were trained and tested 60 times (6 techniques ∗ 10 folds). Assessing the models' performance accuracy through k-fold testing would avoid getting partially valid conclusions which might be drawn down using traditional data management scenarios (Marti et al., 2013; Shiri et al., 2014a) as no any unseen input patterns would be remained in models' development.

Assessing the performance accuracy of the employed models was carried out using the geometric mean error (*GMER*), and the scatter index (*SI*) statistical criteria:

$$GMER = \mathrm{Exp}\left[\frac{1}{n}\sum_{i=1}^{n}\mathrm{Ln}\left(\frac{\theta_{im}}{\theta_{io}}\right)\right] \quad (2)$$

$$SI = \frac{RMSE}{\overline{\theta_o}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_{io}-\theta_{im})^2}}{\overline{\theta_o}} \quad (3)$$