


Modeling of available soil phosphorus (ASP) using multi-objective group method of data handling

Ali Keshavarzi¹  · El-Sayed Ewis Omran² · Sayed M. Bateni³ · Biswajeet Pradhan⁴ · Duraisamy Vasu⁵ · Ali Bagherzadeh⁶

Received: 31 July 2016 / Accepted: 13 August 2016 / Published online: 20 August 2016
© Springer International Publishing Switzerland 2016

Abstract The accurate estimation of available soil phosphorus (ASP) is required for the effective management of environmental resources and improving agricultural productivity. Direct measurement of ASP at large scales is time-consuming and expensive. Therefore, many empirical models or pedotransfer functions (PTFs) have been developed to obtain ASP. This study introduces a new model that estimates ASP from easily measurable soil properties by using the multi-objective group method of data handling (mGMDH) approach. Organic carbon, clay content, calcium carbonate equivalent, pH and ASP were obtained from 100 soil samples, which were collected by the stratified random sampling strategy in Mashhad Plain, Khorasan-e-Razavi Province, Northeast Iran. The

suggested model for the prediction of ASP generated satisfactory results in terms of the statistical performance criteria. The coefficient of determination (R^2) and mean square error (MSE) values of the mGMDH model in training step were 0.606 and 4.9671 mg kg⁻¹, respectively, whereas those were 0.504 and 5.6685 mg kg⁻¹, in the testing step. It was concluded that only 50 % of variations of ASP could be described by selecting easy-to-measure soil variables and owing to the key role of phosphorus in soil fertility and precision agriculture, the other related factors such as terrain attributes and fertilizer management should also be considered as they influence the distribution of the soil phosphorus.

Keywords Group method of data handling · Prediction · Available soil phosphorus · Mashhad plain · Iran

✉ Ali Keshavarzi
alikeshtarzi@ut.ac.ir

¹ Laboratory of Remote Sensing and GIS, Department of Soil Science, University of Tehran, P.O.Box: 4111, Karaj 31587-77871, Iran

² Soil and Water Department, Faculty of Agriculture, Suez Canal University, 41522 Ismailia, Egypt

³ Department of Civil and Environmental Engineering and Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI 96822, USA

⁴ Department of Civil Engineering, Geospatial Information Science Research Center (GISRC), Faculty of Engineering, University Putra Malaysia, Serdang, Selangor Darul Ehsan 43400, Malaysia

⁵ Division of Soil Resource Studies, ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra 440 033, India

⁶ Department of Agriculture, Islamic Azad University, Mashhad Branch, Emamyeh Boulevard, P.O.Box: 91735-413, Mashhad, Iran

Introduction

Phosphorus (P) is an essential nutrient for plants, which is found in various forms in the soil. After nitrogen (N), P is one of the most important elements required by the plants and it is involved in processes such as photosynthesis, cell respiration and cell division (Whitney 1988). Total phosphorus (TP) concentration in soils ranges from 10 to 1000 g kg⁻¹, depending on the soil horizon, soil substrate, pedogenic processes, and land use types and its intensity (Kruse et al. 2015). Mineral forms of phosphorus in soils usually exist as a primary mineral such as apatite, hydroxyapatite and oxy-apatite, and hydrated oxides of iron, aluminum and manganese where P is insoluble (Grant et al. 2005). The available forms present in the soil solution significantly influence crop production. Moreover, soil P dynamics is influenced by soil physical and chemical

properties and farming practices. A better understanding of the factors controlling its distribution and availability is required to manage P nutrition for cropping systems (Roger et al. 2014). Several factors can affect the amount of available P to plants. Organic matter, soil pH, lime, and the presence of iron and aluminum oxides are the main factors affecting available soil P (ASP) for plants (Freeman and Rowell 1981; Mohebbi Sadegh 2014). It is noteworthy to mention that even sandy soils fix high amounts of P because of quick reaction of P with Fe, Al and Ca, which precipitate as insoluble forms (Sarhadi-Sardoui et al. 2003). Havlin et al. (1999) demonstrated that by adding lime to the soil the amount of available P could be significantly reduced, indicating the availability of phosphorus in calcareous soils, which is a challenge for plant nutrition. As large amount of phosphorus fertilizer is added in calcareous soils of Iran consistently, investigation of the effective factors on P availability is a prerequisite to decide on soil P management.

The development of models, which simulate soil processes expanded rapidly in recent years. They are meant for the improvement and understanding of important processes and acts as a tool for clarifying or predicting problems related to agricultural activities and the environment (McBratney et al. 2002). Efforts to predict and assess the spatial distribution of soil P have been well established. However, traditional ways to explore soil P variation have no longer been considered appropriate since it is labor intensive, time consuming and lacks spatial exhaustiveness. Therefore, the potential of other methods for soil P prediction needs to be explored. Determination of ASP at large scales is time-consuming and expensive (Keshavarzi et al. 2015). Hence, many empirical models or pedotransfer functions (PTFs) have been developed to assess soil P (Keshavarzi et al. 2015).

An artificial neural network (ANN) is a computational structure inspired by the study of biological neural processing. ANN generally consists of many interconnected processing elements widely known as neurons. The two important elements of neural networks are types of neural interconnection arrangements available and different algorithms used to set the strength of the connections. These algorithms are used to model the complex interaction of the environmental systems' interactions. It allows modeling of complex systems without requiring the explicit formulation of the relationships that may exist between variables (Omran 2012). ANNs are successfully applied for the estimation of several difficult-to-measure soil characteristics (Merdun et al. 2006; Landeras et al. 2008; Keshavarzi et al. 2015). One of the advantages of using ANN compared with conventional models is that it does not require determining a specific function to express the relationship between input and output, which is

achieved by the training process (Schaap and Bouten 1996). These methods to estimate difficult-to-measure soil characteristics such as P could save time and are inexpensive (Keshavarzi et al. 2015). The most common models in the estimation of difficult-to-measure soil characteristics are multivariate regression equations that are used and developed by various researchers (McBratney et al. 2002). However, the models developed for one region may not be applicable to another (Wagner et al. 2001).

The inherent complexity in designing feed-forward ANNs in terms of understanding the most appropriate topology and coefficients has a great impact on their performance (Neyshaburi et al. 2015). In the case of weight or coefficient training procedure, the gradient descent algorithm (i.e. back propagation) is the most popular one (Nariman-Zadeh et al. 2005). It is believed that such learning algorithms are often trapped in a local minimum and are incapable of finding a global minimum due to multimodality and/or non-differentiability of many error functions (Porto 1997). There have been many efforts to deploy population-based stochastic search algorithms such as evolutionary methods to design ANNs since it would be particularly useful for dealing with complex problems having large search spaces with many local optima (Ivakhnenko 1971). Evolutionary methods are effective tools for system identification (Kristinsson and Dumont 1992; Liu and Kadirkamanathan 1999) that generate polynomial neural networks to model either simulated or experimental data of any kind (Ahmadi et al. 2007).

In recent years, the use of ANNs leads to successful application of different algorithms in which group method of data handling (GMDH) algorithm is one. The GMDH algorithm is a self-organizing approach by which gradually complicated models are generated based on the evaluation of their performances on a set of multi-input single-output data pairs. Along with artificial neural networks (ANNs) and genetic programming (GP), GMDH is a data driven technique suitable for automatic generation of models linking the input and output variables. Pachepsky et al. (1998), who were the pioneers of using GMDH in soil science, developed point PTFs from data on soil texture, bulk density, penetration resistance, and water content at different suction values to predict the soil water retention curve (SWRC). Pachepsky and Rawls (1999) applied the GMDH technique to investigate the effect of data grouping on the improvement of PTF accuracy and reliability. Ungaro et al. (2005) compared the point and parametric PTFs developed by ANNs and GMDH and found that GMDH-driven point PTFs performed better with the exception of water content at 1500 kPa. For the parametric PTFs, ANNs provided satisfactory results, but were less accurate at the training step. The GMDH procedure has also been successfully employed by Wösten et al. (2001),

Tomasella et al. (2003), and Nemes and Rawls (2006) to build PTFs for soil hydraulic properties.

As corroborated in the aforementioned literature review, multi-objective group method of data handling (mGMDH) has not yet been commonly used to estimate available soil P. Therefore, the objective of this research is to introduce mGMDH to demonstrate its potential in the development of PTFs and estimation of available soil P. In the GMDH procedure, only one objective function should be minimized while in the mGMDH two or more objective functions should be simultaneously minimized that may lead to different results (Neyslaburi et al. 2015). Thus, the main objective of this study is to employ mGMDH to derive PTFs for modeling of ASP using easy-to-measure soil properties.

Materials and methods

Study area and data used

The present study was conducted in Mashhad Plain with an area of 6131 km², Khorasan-e-Razavi Province, Northeast Iran (Fig. 1). The study area is located between latitude 35°59'N to 37°04'N and longitude 58°22'E to 60°07'E. The general physiographic trend of the plain extends in a NW–SE direction with an average of 160 km in length surrounded by the two mountainous zones of Kopet-dagh northward and Binaloud southward based on a visual interpretation of the satellite imagery and field observations. The topographic elevation values of the study area

vary between 900 m a.s.l and 1500 m a.s.l, while the main topographic elevation ranges over 1200 m a.s.l.

Geologically, the alluvial plain has developed into a thick sediment-dominated environment belonging to the quaternary period. The soil textures are loam, sandy loam and sandy clay loam. The dominant soil types include calcareous cambisols, gypsic regosols, calcareous regosols and calcareous fluvisols, which cover pediment plains, plateau and upper terraces and gravelly colluvial fans, respectively. The study area consists of six cities with a population of about 2,481,290 and 519 villages with a population of about 422,610, scattered over the entire plain. The main land use practiced in the study area is irrigated farming around the Kashfrod River, characterized by semi-arid climate with mean annual precipitation of 222.1 mm and mean annual temperature of 15.8 °C. The maximum rainfall is received in the month of March (44.8 mm) and the minimum in September (1.2 mm). Utilizing the stratified random sampling, 100 soil samples were collected from 0 to 30 and 30 to 60 cm depth. The collected soil samples were air dried, crushed and sieved using a 2 mm sieve size and subsequently subjected to analysis. The laboratory tests on the soil samples, including particle size distribution (i.e., clay, silt and sand fractions) and determination of soil organic carbon (OC), calcium carbonate equivalent (CCE), pH and ASP were carried out as described by Sparks et al. (1996). Data were tested for normality through Kolmogorov–Smirnov test. Logarithmic transformation was performed for OC and ASP data, which were processed further using the mGMDH models (Fig. 2).

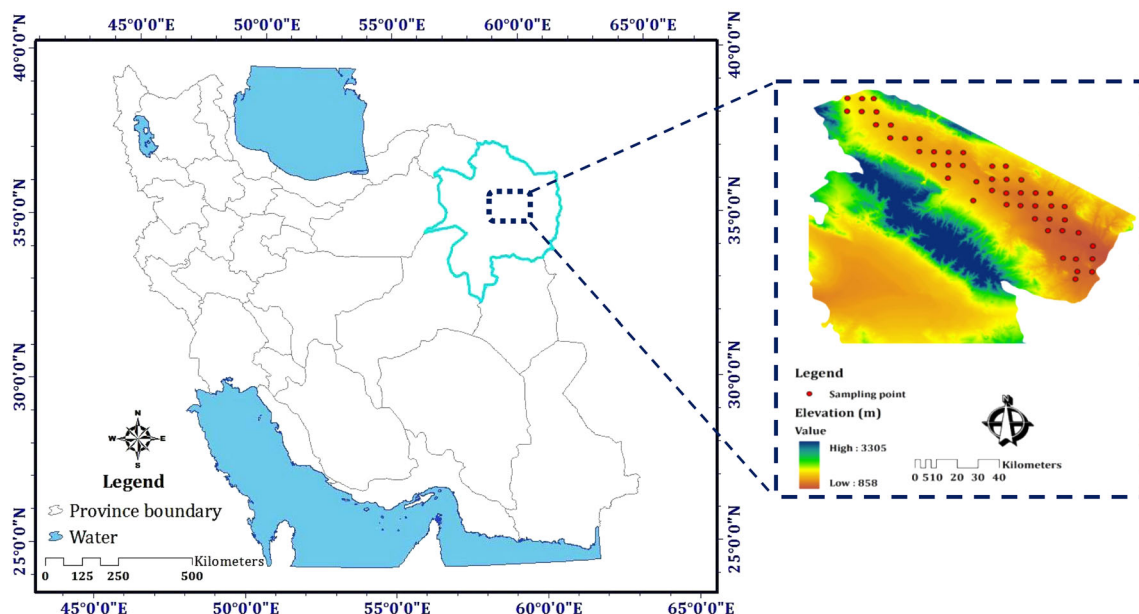


Fig. 1 Location and geographical position of the study area

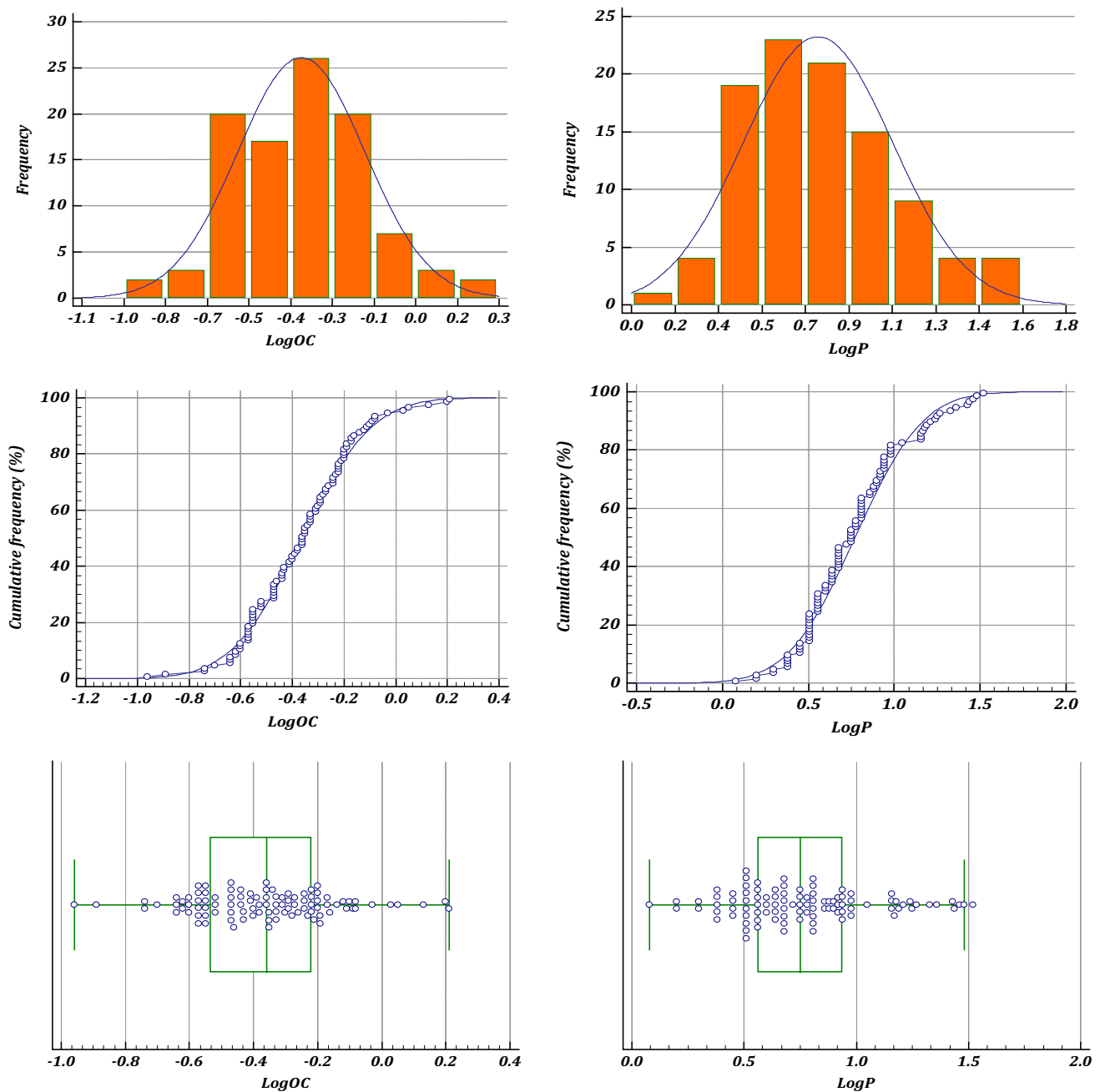


Fig. 2 Distribution plots (histogram), cumulative frequency plots and box-and-whisker plots of transforming variables

After data processing and data normality confirmation, data points were randomized and 80 % of the data was applied as training data, while remaining 20 % was utilized as a test data, respectively, using cross validation technique (Fijani et al. 2013; Barzegar et al. 2016a, b), which was utilized to divide the data set. Data points were standardized for equalization before the models’ training, which prevents excessive shrinking weights. The data points were converted between 0 and 1 as most of the output threshold functions, which were found within this

range. In order to cover all possible ranges of the data pattern, data division was carried out randomly, as mentioned earlier.

The applied data were then normalized using the following equation to fix them between (0,1).

$$y_{\text{normal}} = (y_0 - y_{\text{min}}) / (y_{\text{max}} - y_{\text{min}}) \tag{1}$$

where, y_{max} and y_{min} represent the maximum and minimum values of each record, and y_0 shows the observed (recorded) values.

The group method of data handling (GMDH)

In this approach, GMDH was used to circumvent the difficulty of knowing a prior knowledge of the mathematical model of the process being considered. Therefore, it can be used to model complex systems without having specific knowledge of the systems (Müller and Lemke 2000; Neyshaburi et al. 2015). The main idea of GMDH is to build an analytical function in a feed forward network based on a quadratic node transfer function whose coefficients are to be obtained using a regression technique. Different from the other machine learning approach, GMDH is a self-organizing inductive modeling technique based on explicit polynomial regression algorithms. The method selects automatically the essential input variables and builds a hierarchical, polynomial regression given a degree of complexity specified by the user. Different from ANNs techniques, which are essentially deductive in their nature, the GMDH technique, as GP (Parasuraman and Elshorbagy 2007), does not require an arbitrary, a priori structure of the network connecting inputs and outputs. The structure of the model and the dependence of outputs on the most significant inputs are found generating a network structure based on the characteristics of the data set during the estimation process itself; the GMDH networks can then be considered as universal structure identifiers (Ungaro et al. 2014).

Accuracy and reliability criteria

The following statistical indicators were selected in the performance evaluation of constructed models (Barzegar and Asghari Moghaddam 2016):

1. Coefficient of determination (R^2), and
2. Mean square error (MSE).

$$R^2 = \left[\frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\left[\sum_{i=1}^N (P_i - \bar{P})^2 \sum_{i=1}^N (O_i - \bar{O})^2 \right]^{1/2}} \right]^2 \tag{2}$$

$$MSE = [N^{-1} \sum_{i=1}^N (P_i - O_i)^2] \tag{3}$$

where, N is the number of observations, P_i is the predicted values, O_i is the observed data, and \bar{P} and \bar{O} are the mean values for P_i and O_i , respectively. The coefficient of determination (R^2) measures the degree of correlation among the observed and predicted values. R^2 values range from 0 to 1, with 1 indicating a perfect relationship between the data and the line drawn through them, while 0 represents no statistical correlation between the data and the line. The MSE evaluates the variance of errors independent of the sample size. The MSE indicates the

discrepancy between the observed and predicted values. A perfect fit between observed and predicted values would have an MSE of 0 (Barzegar et al. 2016c).

Model development

Multi-objective group method of data handling

Four characteristics, including OC (%), clay (%), CCE (%), and pH were used as the input variables and the ASP (mg kg^{-1}) as output was utilized in the mGMDH models. By means of a GMDH algorithm, a model can be represented as a set of neurons at which different pairs in every layer are connected through a quadratic polynomial and, thus, produce new neurons in the next layer. Such description can be used in modeling to connect inputs to outputs. The formal definition of the identification problem is to find an approximate relationship between a set of input variables ($x_1, x_2, x_3, \dots, x_n$) and an output variable y (Neyshaburi et al. 2015). Therefore, the mathematical description can be fully represented by a system of partial quadratic polynomials (Eq. (4)):

$$\hat{y} = a_0 + a_1x_i + a_2x_{ij} + a_3x_ix_j + a_4x_i^2 + a_5x_j^2 \tag{4}$$

Consisting of only two variables (neurons) that predict output \hat{y} for a given set of input variables ($x_1, x_2, x_3, \dots, x_n$) as close as possible to its actual value y (Atashkari et al. 2005; Neyshaburi et al. 2015). The GMDH algorithm used multi-objective optimization. A complete overview of mGMDH algorithms was given by Atashkari et al. (2005). In multi-objective optimization problems, there are several objective or cost functions (a vector of objectives) to be optimized (minimized or maximized) simultaneously. These objectives often conflict with each other so that as one objective function improves, another deteriorates. Therefore, there is no single optimal solution that is best with respect to all the objective functions. Instead, there is a well-known set of optimal solutions called Pareto optimal solutions (Coello and Christiansen 2000) for multi-objective optimization problems (Bayat et al. 2011).

In general, the vector $X^* = [X_1^*, X_2^*, X_3^*, \dots, X_n^*]^T$ can be optimized for:

$$F(X) = [f_1(X), f_2(X), \dots, f_k(X)]^T \tag{5}$$

Subject to m inequality constraints $g_i(X)$:

$$g_i(X) \leq 0 \quad (i = 1, \dots, m) \tag{6}$$

and p equality constraints $h_j(X)$:

$$h_j(X) = 0 \quad (j = 1, \dots, p) \tag{7}$$

where, T is the transpose sign, $X(\in R^n)$ is the vector of decision or design variables, and $F(X)(\in R^k)$ is the vector of objective functions, which must be minimized or

maximized. It is assumed, however, that, all of the objective functions are to be minimized without loss of generality (Atashkari et al. 2005; Bayat et al. 2011; Neyshaburi et al. 2015). Two of the objectives were trained and testing mean squared errors, which required minimization. Evolutionary algorithms were used for multi-objective optimization because of their natural properties which is suitable for these types of problems. Genetic algorithms were used in a mGMDH neural network for each neuron searching its optimal set of connections to the preceding layer (Atashkari et al. 2005; Neyshaburi et al. 2015).

Results and discussion

Table 1 shows the descriptive statistics including minimum, maximum, mean, standard deviation (SD) and the range of the data utilized as well as the correlation coefficient between the ASP. Also, input variables for both training and testing data sets are listed in Table 1. After training the proposed model, the model was tested using 20 samples. The results of the developed models in the training and testing steps are presented in Figs. 3 and 4. Figure 5 demonstrates the scatter plots of the target versus output in training and testing steps for mGMDH models.

Table 2 shows the statistical performance criteria used to evaluate the model. The suggested model for the prediction of ASP generated satisfactory results in terms of the statistical performance criteria. Therefore, this model was acceptable for prediction of ASP in the Mashhad plain, Northeast Iran. The proposed model obtained relatively lower prediction errors in training step as compared to the testing step indicating that this model exhibited relatively better generalization as compared to the prediction. The R^2

and MSE values of the mGMDH model in training step were 0.606 and 4.9671 mg kg⁻¹, respectively, whereas those were 0.504 and 5.6685 mg kg⁻¹, in the testing step. This result concurs with the studies of Atashkari et al. (2005), Bayat et al. (2011) and Neyshaburi et al. (2015). It was concluded that this may be related to the embedded genetic algorithm and automated selection of essential input variables, and builds polynomial equations for the modeling.

When using mGMDH models to predict the ASP, the relations between characteristics need to be described by well-defined equations. The choice of training and test data set as a part of the data mining process can lead to more accurate of the mGMDH model in predicting ASP. However, the generality offered by the GMDH in this study is still limited by some issues. The first issue is the choice of soil for estimations in which soil properties influence the prediction results. In most calcareous soils of Iran that are evolved in arid and semi-arid regions, a large amount of ASP is retained by the reactions of absorption and illuviation of carbonate minerals (Musavi and Sepehr 2013). Second, chemistry of P in soils is very complicated because inorganic P can react with elements such as calcium, iron and aluminum and convert them to phosphates. Additionally, organic P can be found with a variety of shapes and resistant to microbial degradation in soil (Soltani et al. 2011). Therefore, the variation in the amount of OC and the lime in the study area can make a difference to the amount of P in soils (Keshavarzi et al. 2015). Third, the clay content, Fe and Al oxides enhance P sorption (Singh and Gilkes 1991; Freese et al. 1992; Frossard et al. 1995), whereas soil OC has the reverse effect (Dubus and Becquer 2001). Demaria et al. (2013) emphasized that soil pH and metal ions have a significant effect on soil P contents and

Table 1 Descriptive statistics of data sets utilized for training and testing

Variable	Minimum	Maximum	Mean	SD	Range	Correlation with ASP
Training data						
OC (%)	0.11	1.61	0.49	0.26	1.50	0.368**
Clay (%)	4.00	41.00	19.24	7.39	37.00	0.251*
CCE (%)	4.90	39.70	21.74	8.65	34.80	-0.112**
pH	7.70	8.40	8.10	0.12	0.70	-0.283**
ASP (mg kg ⁻¹)	1.20	33.50	7.75	6.71	32.30	1
Testing data						
OC (%)	0.14	1.58	0.46	0.29	1.44	0.350**
Clay (%)	6.00	38.00	19.10	7.31	32.00	0.273*
CCE (%)	7.50	36.00	20.68	8.44	28.50	-0.129**
pH	7.72	8.34	8.02	0.10	0.62	-0.280**
ASP (mg kg ⁻¹)	2.80	17.90	6.94	5.82	15.10	1

SD standard deviation

* Significant at 0.05 level

** Significant at 0.01 level

Fig. 3 Measured and predicted ASP (mg kg^{-1}) using the mGMDH model in the training step

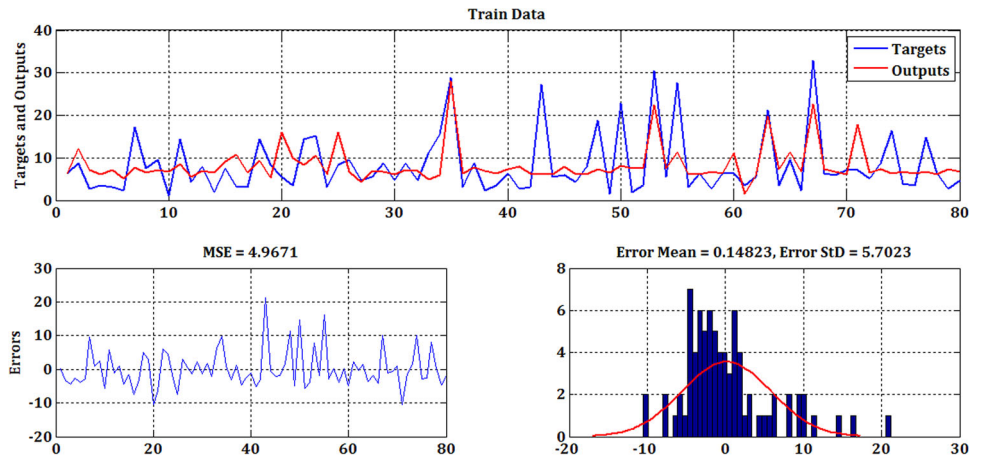


Fig. 4 Measured and predicted ASP (mg kg^{-1}) using the mGMDH model in the testing step

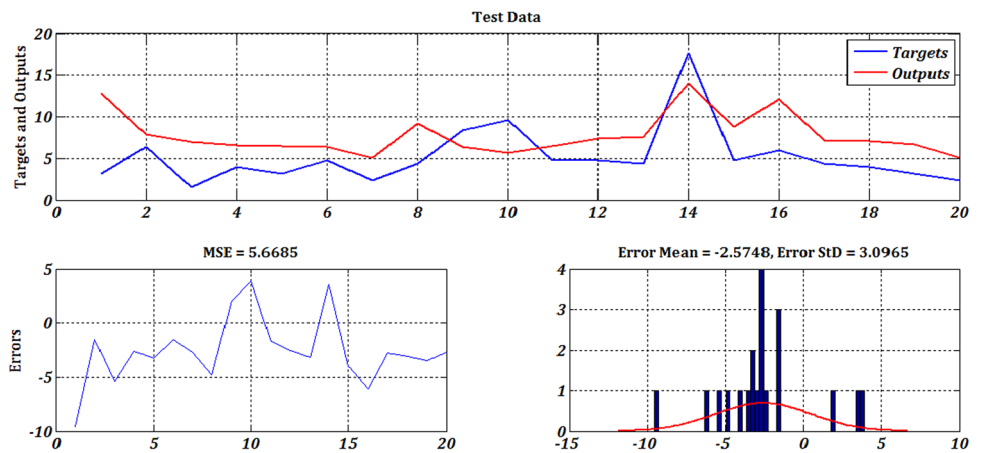
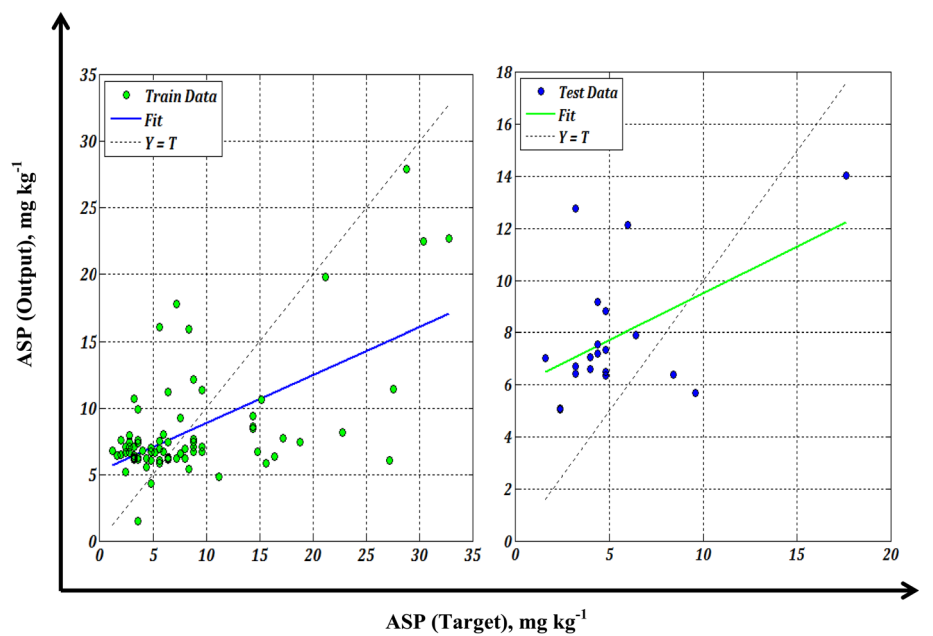


Fig. 5 Scatter plots of the target versus output in training (left) and testing (right) steps for mGMDH model



the variation in phosphorus contents might be due the variation in soil properties since soil OC and clay distribution were considerably different in the studied area.

Fourth, land use may affect the relationship among variables, which was also demonstrated by similar results (Jia et al. 2011; Lemerrier et al. 2008; Reijneveld et al. 2010).

Table 2 The results of the statistical criteria used to evaluate the mGMDH model in the training and testing steps for prediction of ASP (mg kg^{-1})

Model	Training step		Testing step	
	R ²	MSE (mg kg^{-1})	R ²	MSE (mg kg^{-1})
mGMDH	0.606	4.9671	0.504	5.6685

Finally, topography influences soil properties due to local re-distribution of water, solar radiation and soil material (Gessler et al. 2000; Kozar et al. 2002).

Conclusions

The mGMDH was successfully used and assessed for the prediction and modeling of available soil phosphorus (ASP) in the Mashhad plain, Northeast Iran based on easy-to-measure soil properties. The soil properties used are OC (%), Clay (%), CCE (%), and pH. The coefficient of determination (R^2) and mean square error (MSE) were utilized to evaluate the model's performance. The R^2 and MSE values of the mGMDH model in testing step were 0.504 and 5.6685 mg kg^{-1} , respectively. From the results, it was concluded that mGMDH can be used for the estimation of ASP and is recommended for analysis due to its capability and better performance.

References

- Ahmadi H, Mottaghtalab M, Nariman-Zadeh N (2007) Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. *J Appl Poult Res* 16:494–501. doi:10.3382/japr.2006-00074
- Atashkari K, Nariman-Zadeh N, Pilechi A, Jamali A, Yao X (2005) Thermodynamic Pareto optimization of turbojet engines using multi-objective genetic algorithms. *Int J Therm Sci* 44:1061–1071. doi:10.1016/j.ijthermalsci.2005.03.016
- Barzegar R, Asghari Moghaddam A (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Model Earth Syst Environ*. doi:10.1007/s40808-015-0072-8
- Barzegar R, Adamowski J, Moghaddam AA (2016a) Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. *Stoch Environ Res Risk Assess* 1–23. doi:10.1007/s00477-016-1213-y
- Barzegar R, Asghari Moghaddam A, Baghban H (2016b) A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. *Stoch Environ Res Risk Assess* 30(3):883–899
- Barzegar R, Sattarpour M, Nikudel MR, Asghari Moghaddam A (2016c) Comparative evaluation of artificial intelligence models for prediction of uniaxial compressive strength of travertine rocks, case study: Azarshahr area, NW Iran. *Model Earth Syst Environ* 2:76. doi:10.1007/s40808-016-0132-8
- Bayat H, Neyshabouri MR, Mohammadi K, Nariman-Zadeh N (2011) Estimating water retention with pedotransfer functions using multi-objective group method of data handling and ANNs. *Pedosphere* 21:107–114. doi:10.1016/S1002-0160(10)60085-9
- Coello CA, Christiansen AD (2000) Multiobjective optimization of trusses using genetic algorithms. *Comput Struct* 75:647–660. doi:10.1016/S0045-7949(99)00110-8
- Demaria P, Sinaj S, Flisch R, Frossard E (2013) Soil properties and phosphorus isotopic exchangeability in cropped temperate soils. *Commun Soil Sci Plant Anal* 44:287–300
- Dubus IG, Becquer T (2001) Phosphorus sorption and desorption in oxide-rich Ferralsols of New Caledonia. *Aust J Soil Res* 39:403–414
- Fijani E, Nadiri AA, Asghari Moghaddam A, Tsai F, Dixon B (2013) Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh-Bonab plain aquifer Iran. *J Hydrol* 530:89–100
- Freeman JS, Rowell DL (1981) The adsorption and precipitation of phosphate on to calcite. *J Soil Sci* 32:75–84
- Freese D, Van der Zee SEATM, Van Riemsdijk WH (1992) Comparison of different models for phosphate sorption as a function of the iron and aluminum oxides in soils. *Soil Sci* 43:729–738
- Frossard E, Brossard M, Hedley MJ, Metherell A (1995) Reactions controlling the cycling of P in soils. In: Tiessen H (ed) *Phosphorus cycling in terrestrial and aquatic ecosystems: a global perspective*. SCOPE/John Wiley, New York, pp 107–137
- Gessler PE, Chadwick OA, Chamran R, Althouse L, Holmes K (2000) Modeling soil-landscape and ecosystem properties using terrain attributes. *Soil Sci Soc Am J* 64:2046–2056
- Grant C, Bittman S, Montreal M, Plenchette C, Morel C (2005) Soil and fertilizer phosphorus: effects on plant P supply and mycorrhizal development. *Can J Plant Sci* 85:3–14
- Havlin JL, Beaton JD, Tisdale SL, Nelson WL (1999) *Soil fertility and fertilizer an introduction to nutrient management*, 6th eds. Macmillan pub Co. New York, pp 154–196
- Ivakhnenko AG (1971) Polynomial theory of complex systems. *IEEE Trans Syst Man Cybern* 1:364–378. doi:10.1109/TSMC.1971.4308320
- Jia S, Zhou D, Xu D (2011) The temporal and spatial variability of soil properties in an agricultural system as affected by farming practices in the past 25 years. *J Food Agric Environ* 9:669–676
- Keshavarzi A, Sarmadian F, Omran EE, Iqbal M (2015) A neural network model for estimating soil phosphorus using terrain analysis. *Egypt J Remote Sens Space Sci* 18:127–135
- Kozar B, Lawrence R, Long DS (2002) Soil phosphorus and potassium mapping using a spatial correlation model incorporating terrain slope gradient. *Precis Agric* 3:407–417
- Kristinsson K, Dumont G (1992) System identification and control using genetic algorithms. *IEEE Trans Syst Man Cybernet* 22:1033–1046. doi:10.1109/21.179842
- Kruse J, Abraham A, Amelung W, Baum C, Bol R, Kuhn O, Lewandowski H, Niederberger J, Oelmann Y, Ruger C, Santner J, Siebers M, Siebers N, Spohn M, Vestergren J, Vogts A, Leinweber P (2015) Innovative methods in soil phosphorus research: A review. *J Plant Nutr Soil Sci* 1–46. doi:10.1002/jpln.201400327
- Landeras G, Ortiz-Barredo A, Lopez JJ (2008) Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agric Water Manage* 95:553–565
- Lemercier B, Gaudin L, Walter C, Arousseau P, Arrouays D, Schwartz C, Saby NPA, Follain S, Abrassart J (2008) Soil phosphorus monitoring at the regional level by means of a soil test database. *Soil Use Manag* 24:131–138

- Liu GP, Kadirkamanathan V (1999) Multi-objective criteria for neural network structure selection and identification of nonlinear systems using genetic algorithms. *IEEE Control Theory Appl* 146:373–382. doi:[10.1049/ip-cta:19990501](https://doi.org/10.1049/ip-cta:19990501)
- McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer functions to soil inference systems. *Geoderma* 109:41–73
- Merdun H, Ozer C, Meral R, Apan M (2006) Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil Till Res* 90:108–116
- Mohebbi Sadeqh MJ (2014) Investigation of relationships between available phosphorus, potassium and some soil properties in agricultural lands of Varamin—Iran. *Inter J Agric Biosci* 3(1):7–12
- Müller JA, Lemke F (2000) Self-organizing data mining: an intelligent approach to extract knowledge from data. Hamburg, Libri
- Musavi R, Sepehr E (2013) Phosphorus efficiency of some barley genotypes in the presence of phosphate-solubilizing microorganisms. *EJGCTS* 4(16):27–40 (**In Persian**)
- Nariman-Zadeh N, Darvizeh A, Jamali A, Moeini A (2005) Evolutionary design of generalized polynomial neural networks for modelling and prediction of explosive forming process. *J Mater Process Technol* 164–165:1561–1571. doi:[10.1016/j.jmatprotec.2005.02.020](https://doi.org/10.1016/j.jmatprotec.2005.02.020)
- Nemes A, Rawls WJ (2006) Evaluation of different representations of the particle-size distribution to predict soil water retention. *Geoderma* 132:47–58. doi:[10.1016/j.geoderma.2005.04.018](https://doi.org/10.1016/j.geoderma.2005.04.018)
- Neyshaburi MR, Bayat H, Mohammadi K, Nariman-Zadeh N, Irannejad M (2015) Improvement in estimation of soil water retention using fractal parameters and multiobjective group method of data handling. *Arch Agron Soil Sci* 61(2):257–273. doi:[10.1080/03650340.2014.924107](https://doi.org/10.1080/03650340.2014.924107)
- Omran EE (2012) A neural network model for mapping and predicting unconventional soils at a regional level. *Appl Remote Sens* 2(2):35–44
- Pachepsky YA, Rawls WJ (1999) Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Sci Soc Am J* 63:1748–1757. doi:[10.2136/sssaj1999.6361748x](https://doi.org/10.2136/sssaj1999.6361748x)
- Pachepsky YA, Rawls WJ, Giménez D, Watt JPC (1998) Use of soil penetration resistance and group method of data handling to improve soil water retention estimates. *Soil Till Res* 49:117–126. doi:[10.1016/S0167-1987\(98\)00168-8](https://doi.org/10.1016/S0167-1987(98)00168-8)
- Parasuraman K, Elshorbagy A (2007) Estimating saturated hydraulic conductivity using genetic programming. *Soil Sci Soc Am J* 71:1676–1684
- Porto VW (1997) Evolutionary computation approaches to solving problems in neural computation. In: Back T, Fogel DB, Michalewicz Z (eds) *Handbook of evolutionary computation*. IOP Publishing, Bristol, pp D1.2:1–6
- Reijneveld JA, Ehlert PAI, Termorshuizen AJ, Oenema O (2010) Changes in the soil phosphorus status of agricultural land in the Netherlands during the 20th century. *Soil Use Manag* 26:399–411
- Roger A, Libohova Z, Rossier N, Joost S, Maltas A, Frossard E, Sinaj S (2014) Spatial variability of soil phosphorus in the fribourg canton, switzerland. *Geoderma* 217–218:26–36
- Sarhadi-Sardoui J, Ronagashi A, Maftoun M, Karimian N (2003) Growth and chemical composition of corn in three calcareous sandy soil of Iran as affected by applied phosphorus and manure. *J Agric Sci Tech* 5:77–84
- Schaap MG, Bouten W (1996) Modeling water retention curves of sandy soils using neural networks. *Water Resour Res* 32(10):3033–3040
- Singh B, Gilkes RJ (1991) Phosphorus sorption in relation to soil properties for the major soil types of south-western Australia. *Aust J Soil Res* 29:603–618
- Soltani SM, Davatgar N, Kavooosi M, Darighoftar F (2011) Phosphorus fractionation of paddy fields and their relations with physical and chemical properties of soils (Case study: some-e-Sara city, Guilan province). *Water Soil Conserv* 18(2):159–176 (**In Persian**)
- Sparks DL, Page AL, Helmke PA, Leoppert RH, Soltanpour PN, Tabatabai MA, Johnston GT, Summer ME (1996) *Methods of soil analysis*. Soil Science Society of America, Madison
- Tomasella J, Pachepsky YA, Crestana S, Rawls WJ (2003) Comparison of two techniques to develop pedotransfer functions for water retention. *Soil Sci Soc Am J* 67:1085–1092. doi:[10.2136/sssaj2003.1085](https://doi.org/10.2136/sssaj2003.1085)
- Ungaro F, Calzolari C, Pistocchi A, Malucelli F (2014) Modelling the impact of increasing soil sealing on runoff coefficients at regional scale: a hydrogeological approach. *J Hydrol Hydromech* 62(1):33–42. doi:[10.2478/johh-2014-0005](https://doi.org/10.2478/johh-2014-0005)
- Ungaro F, Calzolari C, Busoni E (2005) Development of pedotransfer functions using a group method of data handling for the soil of the Pianura Padano-Veneta region of north Italy: water retention properties. *Geoderma* 124:293–317. doi:[10.1016/j.geoderma.2004.05.007](https://doi.org/10.1016/j.geoderma.2004.05.007)
- Wagner B, Tarnawski VR, Hennings V, Muller U, Wessolek G, Plagge R (2001) Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set. *Geoderma* 102(3–4):275–297
- Whitney DA (1988) *Phosphorus facts: soil, plant, and fertilizer*. Cooperative Extension Service, Manhattan
- Wösten JHM, Pachepsky YA, Rawls WJ (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J Hydrol* 251:123–150. doi:[10.1016/S0022-1694\(01\)00464-4](https://doi.org/10.1016/S0022-1694(01)00464-4)