Modeling Soil Cation Exchange Capacity in Arid Region of Iran: Application of Novel Hybrid Intelligence Paradigm.

Mahmood Shahabi
University of Tabriz

Mohammad Ali Ghorbani
University of Tabriz

Sujay Raghavendra Naganna (sujay.gopan@gmail.com)
Siddaganga Institute of Technology https://orcid.org/0000-0002-0482-1936

Sungwon Kim
Dongyang University

Sinan Jasim Hadi
Ankara University

Zaher Mundher Yaseen
Al-Ayen University

Samed Inyurt
Zonguldak Bulent Ecevit University

Research Article

Keywords: Cation exchange capacity, extreme learning machine, multiple model strategy, artificial neural network, genetic algorithm.

Posted Date: May 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-471256/v1

Modeling soil cation exchange capacity in arid region of Iran: Application of novel hybrid intelligence paradigm.

Mahmood Shahabi[1], Mohammad Ali Ghorbani[1], Sujay Raghavendra Naganna[2,*], Sungwon Kim[3], Sinan Jasim Hadi[4], Zaher Mundher Yaseen[5], Samed Inyurt[6]

[1]Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.

[2]Department of Civil Engineering, Siddaganga Institute of Technology, Tumakuru 572103, Karnataka, India.

*Correspondence E-mail: sujay.gopan@gmail.com

[3]Department of Railroad and Civil Engineering, Dongyang University, Yeongju, 36040, Republic of Korea

[4]Department of Real Estate Development and Management, Faculty of Applied Sciences, Ankara University, Ankara, Turkey

[5]New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah, 64001, Iraq

[6]Department of Geomatics Engineering, Faculty of Engineering, Zonguldak Bulent Ecevit University, Zonguldak, Turkey.

1

ABSTRACT

The potential of the soil to hold plant nutrients is governed by cation exchange capacity (CEC) of any soil. Estimating soil CEC aids in conventional soil management practices to replenish the soil solution that supports plant growth. In the present study, a multiple model integration scheme driven by hybrid GANN (MM-GANN) was developed and employed to predict the accuracy of soil CEC in Tabriz plain, an arid region of Iran. The standalone models (i.e., artificial neural network (ANN) and extreme learning machine (ELM)) were implemented for incorporating in the MM-GANN. In addition, it was tested to enhance the prediction accuracy of the standalone models. The soil parameters such as clay, silt, pH, carbonate, calcium equivalent (CCE), and soil organic matter (OM) were used as model inputs to predict soil CEC. By the use of several evaluation criteria, the results showed that the MM-GANN model involving the predictions of ELM and ANN models calibrated by considering all the soil parameters (e.g., Clay, OM, pH, Silt, and CCE) as inputs provided superior soil CEC estimates with an NSE = 0.87. The proposed MM-GANN model is a reliable intelligence based approach for the assessment of soil quality parameters intended for sustainability and management prospects.

Keywords: Cation exchange capacity; extreme learning machine; multiple model strategy; artificial neural network; genetic algorithm.

1. INTRODUCTION

Cation exchange capacity (CEC) refers to the extent of the soil's capacity to preserve exchangeable cations, the like of which has a direct bearing on soil fertility triangle (Wolf, 1999). Soil CEC is a sensitive indicator of natural and human-induced perturbations over soil profile and groundwater. Monitoring changes in soil CEC can assist in predicting whether soil quality has degraded, improved, or sustained under diverse agricultural or forestry schemes. In the course of conventional soil management practices to replenish the soil solution that supports plant growth, the negatively charged clay particles and organic substances adsorb and hold on positively charged soil nutrients (e.g. NH_4^+, K^+, Mg^{2+} and Ca^{2+} etc.) via electrostatic forces (Ketterings et al., 2007). Depending on the soil structure, CEC clearly demonstrates the shrink-swell potential of any soil; a high CEC value (>40 meq/100g) denotes that a soil structure will recuperate gradually, and at sometimes can show expansive behavior. In contrast, a soil with low CEC value (<10 meq/100g) will have reduced capacity to hold water and end up being acidic rapidly (Thomas et al., 2000). Soil CEC can fluctuate according to clay percentage, soil pH, ionic strength, soil-to-solution ratio, clay type and changing organic matter composition. For agriculture, the preferred value of CEC is >10 meq/100g for exchange between plant root hairs and soils (Mengel, 1993). The leaching of contaminants into the underlying aquifer system is usually affected by CEC and percent base saturation which are eloquent indices of soil fertility and nutrient retention capacity. In areas of intensive irrigation, the continuous use of inorganic fertilizers (in excess) inundates the soil profile with more nutrients and thereby flush a plume of contaminants to the groundwater (Böhlke, 2002). Therefore, in the early stages of agriculture, it is necessary to estimate CEC for

determining the supplemental nutrient needs or to remove excess salts which influence over soil structure and agricultural productivity. Soil CEC is a sensitive indicator of natural and human-induced perturbations over soil profile and groundwater. Monitoring changes in soil CEC can assist in predicting whether soil quality has degraded, improved, or sustained under diverse agricultural or forestry schemes.

Various methods for direct measurement of soil CEC have been reported extensively over the literature (Delavernhe et al., 2018; Dohrmann, 2006a, 2006b). Multiple comparison of CEC estimation techniques is presented by Conradie and Kotze, (1989). In addition, there were several ancillary approaches such as pedotransfer functions (PTF) for estimating CEC based on easily measured soil's physical & chemical properties as reported by (Khorshidi and Lu, 2017; Liao et al., 2015; Obalum et al., 2013). Several others researchers conducted studies on the functional relationships between CEC, water retention and particle size distribution. Lambooy (1984) studied the influence of CEC on the water retention characteristics of soils. Parfitt et al. (1995) estimated the CEC using multiple regression models taking into account soil organic carbon and clay content. Krogh et al. (2000) modeled the CEC rates of Danish soils using clay and organic matter content as input variables through multiple linear regression analysis. Van Erp et al. (2001) evidenced that the actual CEC of agricultural soils found to be directly related with estimated charge of organic carbon and clay in the soil at the actual pH of the soil. Using soil organic and non-carbonate clay contents as predictors, Seybold et al. (2004) explained the variation in CEC for several soil horizons based on soil pH, mineralogy class, taxonomic family and CEC-activity class. Fooladmand (2008) derived PTF's using multiple linear regression between CEC and soil textural data including sand content, clay content,

geometric mean particle-size diameter, the soil particle-size distribution, and soil organic matter content. Several PTF's relating soil CEC with soil's sand, silt or clay fractions, and soil organic carbon content evaluated by (Khodaverdiloo et al., 2018).Scholars took into account of calibration dataset size on the prediction accuracy of soil CEC. These classical pedotransfer function-based approaches often suffer from a high degree of inaccuracy due to spatial scale dependence, non-linear relationships among variables and incompetence to handle mixed data (Van Looy et al., 2017). Hence, the motivation of the current state-of-art directed on new research era where more intelligent models should be explored for this field.

The recent researches have focused on improving the estimation accuracy of soil CEC by means of artificial intelligence (AI) techniques. Artificial neural network (ANN) model based PTF's have become popular to predict/estimate soil CEC of different soil types under diverse climatic zones (Amini et al., 2005; Bayat et al., 2014; Seyedmohammadi et al., 2016; Tang et al., 2009; Zolfaghari et al., 2016). Kalkhajeh et al., (2012) conducted the accurate prediction of soil CEC using different soft computing models. They compared the performance of multiple linear regressions (MLR), adaptive neuro-fuzzy inference system (ANFIS), multi-layer perceptron (MLP), and radial basis function (RBF) based ANN models for predicting the soil CEC using the bulk density, calcium carbonate, organic carbon, clay, and silt content (%) of the soil as input variables. The MLP model gave the most reliable prediction of soil CEC. A set of AI techniques along with empirical PTF's were developed and evaluated by (Ghorbani et al., 2015), authors determined the most influential soil properties that influence soil CEC through sensitivity analysis. The ANFIS model provided the superior performance to RBF, MLP, MLR, and empirical PTF's while

estimating soil CEC. Arthur (2017) presented an ANN based methodology for estimating CEC from soil water content at different relative humidity ranges. Relatively few studies were accomplished using support vector machine (SVM), random forests (RF), genetic expression programming (GEP), multivariate adaptive regression splines (MARS), and subtractive clustering algorithm based ANFIS for estimating soil CEC using readily measured soil properties as inputs (Akpa et al., 2016; Emamgolizadeh et al., 2015; Jafarzadeh et al., 2016; Liao et al., 2014). A hybrid model integrating ant colony optimization (ACO) algorithm with ANFIS improved the prediction accuracy of soil CEC accompanied by optimal choice of input subset which comprised of soil properties (e.g. soil organic matter, clay, silt, pH and bulk density) (Shekofteh et al., 2017). Although there has been a noticeable progress on the AI implementation with the field of geoscience, the enthusiasm of developing and exploring more reliable intelligent predictive models is still ongoing research era. As a result, the inspiration of developing a multiple learning intelligent model is investigated here for the soil CEC.

Hybrid soft computing approaches involving evolutionary algorithms coupled with AI techniques facilitate the development of more sophisticated models with higher prediction accuracy. Hence, in the present study, a hybrid approach involving the multi-layer perceptron neural network optimized with genetic algorithm (GANN) was developed and employed to enhance the prediction efficiency of soil CEC in Tabriz plain, an arid region of Iran. In addition, a multiple model integration scheme driven by hybrid GANN (MM-GANN) was also developed and tested to improve the prediction efficiency. To the best of author's knowledge, this multiple model integration scheme driven by GANN approach is a unique one in the literature with

reference to soil CEC prediction. Standalone MLP artificial neural network (ANN) and extreme learning machine (ELM) models were also implemented for incorporating in the multiple model integration scheme and for comparative evaluation with MM-GANN model predictions.

2. THEORETICAL OVERVIEW

2.1 Artificial Neural Network (ANN)

Artificial neural network (ANN) is one of the most versatile algorithms that has proven capable to simulate highly complex and nonlinear relationships between a set of input variables (predictors) and the output data (predictand) (McClelland and Rumelhart, 1988). A three-layered perceptron network with one hidden layer is as shown in **Figure 1**. The network is trained on a set of reference data by adjusting the parameters of ANN model with the assistance of a Levenberg-Marquardt back propagation (BP) algorithm. The network architecture involving a set of processing units (neurons), a specific topology of weighted links connecting the neurons and the learning paradigm that updates the connection weights determine the efficiency of ANN model. Every single input (X_n), weighted by an element (w_{ij}) of the weight matrix (W) are summated and provided to the transfer function or activation function (φ) along with a bias (B) term. The activation function constructs a non-linear decision boundary via linear combinations of the weighted inputs and then applies a threshold to transform the net inputs from all the neuronal unit into an output signal (Haykin, 2009; Kim and Singh, 2015). The Levenberg-Marquardt BP learning rule incrementally adjusts the weight and bias terms to minimize the mean square error (MSE) of the network (Nourani et al., 2013). The quantum of progressions made in adjusting the synaptic weights and biases at every epoch is determined by the learning

rate parameter. Smaller learning rates end up in longer training time however, warrant stability that steers to minimum errors.

2.2 Extreme Learning Machine (ELM)

Extreme learning machine (ELM) model proposed by Huang et al., (2004) for a single layer feedforward network (SLFN) has been widely used for the prediction, forecasting, and estimation in many engineering fields (Acharya et al., 2014; Şahin et al., 2014; Abdullah et al., 2015; Deo and Şahin, 2015; Niu et al., 2018; Yaseen et al., 2018). The previous researches have proved the outstanding advantages of ELM model over the traditional AI techniques. In addition, the ELM model can be used easily and has improved the parameters such as learning speed, use of non-differentiable activation functions while training SLFN, achieving the least training error for superior generalization performance (Huang et al., 2004, 2006). The ELM model based on the principle of empirical risk minimization stands apart from most of the other popular gradient-based learning algorithms for training feedforward neural networks by its solitary learning process which needs solitary iteration only (Huang et al., 2006). While implementing an ELM model, one has to set the number of hidden layer nodes to obtain the optimal model architecture. Figure 2 provides the basic schematic structure of an ELM model. For *N* arbitrary distinct input samples $(X_k, Y_k) \in R^n \times R^n$, the standard SLFN with *L* hidden layer nodes can be described as following equation (1).

$$\sum_{i=1}^{L} \beta_i g(X_k, ; c_i, w_i) = Y_k, \qquad k = 1, 2, 3, \cdots N \tag{1}$$

where $c_i \in R$ = assigned bias of the ith hidden node, $w_i \in R$ = assigned input weight connecting the ith hidden and input layer nodes, β_i = the weight connecting the ith hidden and output layer nodes, $g(X_k,;c_i,w_i)$ = the output of the ith hidden layer node with respect to the input X_k. Each input is assigned to the hidden nodes in the ELM model. The output weights can be derived by finding the least square solutions to the linear system. The main differences between the ELM model and traditional AI techniques is that the parameters of feedforward network including input weights and hidden layer biases are not required to be adjusted previously in the ELM model. For additional information on the ELM model, its several architectures and mathematical formulations refer to Ding et al. (2015), Huang et al. (2011), Wang et al. (2011), and Martínez-Martínez et al. (2011).

2.3 Hybrid Genetic Algorithm - Neural Network (GANN)

Genetic algorithm (GA) belongs to a class of search iterative approaches based on the 'Darwinian' theory of natural selection and genetics that provide optimum solutions for the combinatorial optimization, heuristic search or process planning problems (Goldberg and Holland, 1988; Holland, 1992). The GA implements genetic operators like reproduction, crossover, and mutation for upgradation and search of the best population by imitating the natural evolution process artificially. The GA is initiated with individuals - an initial population of possible solutions, with a specified objective (fitness function) wherein every single individual is symbolized using a chromosome – a distinct form of encoding (Goldberg and Deb, 1991). The chromosomes of a population are nominated for reproduction based on the fitness value and the fittest individuals so selected are manipulated using crossover and

mutation. The rudimentary idea here is the hope that superior parents can probabilistically produce superior offspring's. The offspring's of the next generation are generated by applying the GA operators - crossover and mutation, upon the selected parents. The iteration process continues until the search converges to the termination criterion (Goldberg and Holland, 1988; Jain and Srinivasulu, 2004; Kim and Kim, 2012). The schematic illustration of GA cycle is represented in **Figure 3**. The advantages of GA include: (1) rapid convergence to the global optima, (2) superior multi-directional global search even in complex search surfaces, (3) use of probabilistic transition rules, and the not deterministic ones in the search spaces where the gradient information is missing. The training of an ANN model is somewhat a cyclic process. However, in case of GA, the intelligent search technique allows the user to configure weight initialization range, the number of hidden layer neurons and update the weights, and bias terms of an ANN model. Even though the weights of ANN model are initialized randomly, the GA does not adhere to a simple random walk. Based on the parameter settings, it effectively exploits the information to gamble on new search points for expected improved performance (Goldberg et al., 1991).

2.4 Multiple Model integration scheme driven by hybrid GANN strategy

The proposed multiple models integration scheme involves the development of ANN and ELM models individually using input combinations as defined in their model structures. The discrete outputs (predicted series) of individual ANN and ELM models are then unified as inputs for the GANN model to obtain superior soil CEC predictions. The implementation of this multiple models scheme involves two phases. At the first phase, the best performing ANN and ELM models are identified by simulating with all possible combinations of inputs. Later in the second phase, the discrete outputs (predicted series) of best ANN and ELM models are unified as inputs to simulate the GANN model. The GA optimizes the number of hidden layer neurons and updates the weights and bias terms of an ANN. The final output derived out of this proposed scheme is referred to as multiple model integration scheme driven by hybrid GANN (MM-GANN) strategy (**Figure 4**).

3. CASE STUDY AND DATA DESCRIPTION

The study area (Tabriz plain) considered encompasses an area of 150000 hectares (45°25'– 46°12'E, 37°50'–38°20'N) located in the East Azerbaijan province of Iran. The topography consists of rugged, mountainous rims and the Urmia Lake is positioned near the southwestern part (**Figure 5**). The altitude is around 1360 m above mean sea level. The climate of Tabriz plain is characterized by cold winters and hot summers with a desert steppe area. The average minimum temperature ranges from -1.9°C to -2.2°C in winter, and the average maximum temperature ranges from 25.1°C to 27.5°C in the summer with a mean annual precipitation of 220 mm. The descriptive statistics of CEC and other soil parameters are tabulated in **Table 1**. The spatial

distribution of observed soil CEC is presented in **Figure 6**. For visualizing the spatial variations, the IDW interpolation method has used for all models. The clay and soil organic matter were found to have relatively significant positive correlation, whilst, the sand was found to have negative correlation with soil CEC. The silt, pH and carbonate calcium equivalent (CCE) parameters were not so significantly correlated to soil CEC.

4. MODEL STRUCTURES INVESTIGATED & PERFORMANCE EVALUATION METRICS

The input–output combinations formulated for the development of ANN and ELM models were based on the permutation and combination of different soil parameters as inputs with soil CEC as persistent output. The input-output structures put on trial are as listed in **Table 2**. The performance of the models developed are assessed based on the statistical indices: Root Mean Square Error (RMSE); Mean Absolute Error (MAE); and Nash Sutcliffe Efficiency (NSE).

$$\text{Root Mean Square Error, } RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}$$

$$\text{Mean Absolute Error, } MAE = \frac{\sum_{i=1}^{n}\left|y_i - x_i\right|}{n}$$

$$\text{Nash Sutcliffe Efficiency, } NSE = 1 - \frac{\sum_{i=1}^{n}\left(y_i - x_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

where, x_i - true value; y_i – model estimated value; $\bar{x}$ - mean of true values; $\bar{y}$ - mean of model estimated values and n- number of data points.

5. RESULTS AND DISCUSSION

5.1 Performance of ANN and ELM models

ANN and ELM models were simulated for predicting the soil CEC based on the input-output combinations as mentioned in **Table 2**. The model structure (input nodes-hidden layer nodes-output nodes) and performance statistics of ANN model for each input combination are presented in **Table 3**. In this study, the proposed ANN, ELM, GANN, and MM-GANN models were developed based on MATLAB interface coding. The combination involving all the soil parameters (i.e., Clay + OM + pH + Silt + CCE) provided the better estimates of soil CEC with an NSE = 0.842. The input structure involving four soil parameters (i.e., Clay, OM, pH, and Silt) also provided reasonably good soil CEC estimates with an NSE = 0.826. Despite having significant correlation between clay and soil CEC, the single input – output ANN model (i.e., Clay CEC) failed to provide good soil CEC predictions. The spatial distribution map of ANN predicted soil CEC is presented in **Figure 7**. The ability of ANN model to formulate priori explicit hypotheses about a possible non-linear relationship among several input variables makes it illustrious from other AI techniques.

The performance statistics of ELM models for each input combination are presented in **Table 4**. The combination involving all the soil parameters (i.e., Clay + OM + pH + Silt + CCE) provided the better estimates of soil CEC with an NSE = 0.835. The ELM model efficiency was marginally lesser than that of ANN model. The ELM model simulated with four inputs (i.e., Clay, OM, pH, and Silt) had reasonably substandard performance when compared to that of ANN model with similar input structure. The spatial distribution map of ELM predicted soil CEC is presented in **Figure 8**. The scatter plot of the three highest performed models presented in **Figure**

9 displays the strength, direction, and form of the relationship between the observed and estimated soil CEC by ANN and ELM models. According to the **Figure 9**, the ELM model outperformed the ANN model although they have very close performance in terms of the statistical indices (**Tables 3** and **4**). The ELM model is known for its superior learning speed and virtuous generalization performance than the ANN model.

5.2 Performance of MM-GANN models

ANN and ELM models predictions were employed as inputs to the GANN model to predict soil CEC. To select the optimal input combinations for applying next step problems, the previous literatures demonstrated that few combinations, if possible, were recommended for enhancing the accuracy of improved models based on the different fields (Kim and Kim, 2008; Kim et al., 2015, 2019; Kim and Singh, 2014). Within this category, it is worth mentioning that only three highest performed combinations were considered for processing of hybrid data-driven modeling. The parameters of GA algorithm for adjusting the weights and bias terms of ANN model are presented in **Table 5**. Also, the performance statistics of MM-GANN models are presented in **Table 6.** The MM-GANN models involving the predictions of ELM and ANN models calibrated by considering all the soil parameters (i.e., Clay, OM, pH, Silt, and CCE) as inputs provided superior performance with an NSE = 0.87 in the test phase. The advantage of giving standalone model outputs as inputs attributes to the hybrid model is to comprehend and establish inherent complex relationships between the predictors and the predictand due to improved correlation among them. This is apparently enhancing the learning process of the hybrid model where the predicted output using the standalone models are relative informative predictors. The spatial distribution map of MM-GANN predicted soil CEC is presented in **Figure 10**

which is very much similar to that of observed soil CEC map. The MM-GANN models developed with the predictions of ELM and ANN models calibrated by considering three and four soil parameters as inputs also provided reasonably good soil CEC predictions with NSE = 0.80 and 0.854, respectively. The scatter plots of MM-GANN models presented in **Figure 9** also portrayed the goodness of fit of the model predictions against the observed soil CEC. In the **Figure 9**, it is evident that the third combination of MM-GANN model indicated a very close linearly fitted line to the 1:1 line especially for the combination that had all the parameters. The Taylor diagrams plotted for the best ANN, ELM, and MM-GANN models are shown in **Figure 11**. According to Taylor diagram, it was very much evident that the MM-GANN provided superior estimates of soil CEC compared to ELM and ANN models based on statistical indices wherein the MM-GANN model was the closest to the observed data/point. The point density plots presented in **Figure 12** also supported the above statement by exposing the tradeoff between observed soil CEC against the modelled.

5.3 Validation with previous works and further discussion

It is worth to validate the current research results with the reliable published researches in the literature with reference to the same kind of study area (i.e., semi-arid region). It was selected the correlation coefficient (R^2) indices as an indicator of the prediction capability. The best R^2 obtained for MM-GANN, ELM, and ANN models are $R^2 \approx 0.88$, 0.85, and 0.84. One of the earliest research performed on the soil CEC simulation along Zayandehroud River in Isfahan, Iran, by Amini et al. (2005) established two classical ANN algorithms (i.e., feed forward neural network and generalized regression neural network). The applied models performed with poor

prediction results with $R^2 \approx 0.69$ and 0.66. Another study was conducted by Emamgolizadeh et al. (2015) for prediction soil CEC on collected soil information from Semnan, Mashahad, and Taybad provinces in Iran. The authors developed two new data intelligence models namely genetic expression programming (GEP) and multivariate adaptive regression spline (MARS). The GEP and MARS models attained an $R^2 \approx 0.80$ and 0.86. Overall, the current study showed a convincing correlation performance over the state-of-the-art researches.

Although the current research was the first approach to develop and assess the multiple model integration scheme driven by hybrid GANN (MM-GANN) for improving the accuracy of standalone models (i.e., ANN and ELM), the certified limitation should be addressed for future research. As can be seen from tables and figures, the MM-GANN model can improve the prediction accuracy of soil CEC when the inputs involving the predictions of ELM and ANN models calibrated by considering all the soil parameters (e.g., Clay, OM, pH, Silt, and CCE) are provided. However, one of the disadvantages of MM-GANN model can be classified to select the best standalone model for enhancing the prediction accuracy of soil CEC. Therefore, it is recommended to incorporate the prediction results of other data-driven models as the inputs of MM-GANN model which can enhance the model's performance. In addition, this concept can be expanded and applied to other engineering fields such as structural, hydrologic, water resources, climatic, and different time series prediction/forecasting

6. CONCLUSIONS

Over the past two decades, there is a noticeable demand for soil data assessment with regards to pollution and land degradation. The new era of soil process modeling using data intelligence models has been rapidly boosted. The current study was to develop new hybrid intelligence model based on multi model genetic algorithm neural network for soil cation exchange capacity. Two classical artificial intelligence models namely ANN and ELM were employed to evaluate their performance in predicting soil CEC along with the proposed MM-GANN. Several correlated soil parameters including clay, silt, pH, carbonate calcium equivalent (CCE), and soil organic matter (OM) were used in the form of input attributes to the proposed and the comparable predictive intelligence models. Overall, the proposed MM-GANN model which receives the predicted values of ANN and ELM models as inputs performs well in the prediction of soil CEC. In general, the proposed multiple model integration scheme driven by hybrid GANN (MM-GANN) serves as an effective pedotransfer function to predict soil CEC using readily available soil parameters (i.e., Clay, OM, pH, Silt, and CCE) as input variables. In specific, the current investigation concludes with the following remarks:

- ELM model performed with superior predictability over ANN model based on the examined statistical indices.
- The prediction function of ELM model has faster learning process compared to the traditional ANN model as it does not require any internal parameter tuning.
- The proposed hybrid MM-GANN model outperforms both stand-alone ANN and ELM models in terms of all the statistical indices.
- Using multiple model schema with inputs taken from the predicts of stand-alone models improves the accuracy of the predictions.

- Overall, the proposed hybrid intelligence model exhibited a robust and reliable modeling strategy for modeling soil cation exchange capacity at this particular studied region.

Before this end, it is worth to state the possibility for future research. As a fact, soil CEC is influenced by several morphological parameters (Sharma et al., 2015; Tan and Dowling, 1984); thus, integrating a feature selection as prior modeling phase for the prediction process is highly recommended to be established (Shekofteh et al., 2017). In addition, owing to the associated variability with each soil CEC type, it is an ideal proposition to estimate each type individually.

Conflict of interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no financial gains for this work that could have influenced its outcome.

Author contribution statement

M.S. and M.A.G. conceived of the presented idea. M.S. developed the models and performed the computations. M.A.G. wrote the manuscript with support from S.R.N., S.K., S.J.H and Z.M.Y. S.I and S.K. verified the methods. M.A.G. supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding

This research received no external funding.

Ethical Approval

Not Applicable

Consent to Participate

Not Applicable

388 **Consent to Publish**

389 Not Applicable

390 **Availability of data and materials**

391 The datasets generated during and/or analyzed during the current study are available
392 from the authors on reasonable request.

REFERENCES

Acharya, N., Shrivastava, N.A., Panigrahi, B.K., Mohanty, U.C., 2014. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Climate dynamics 43(5-6), 1303-1310. doi:10.1007/s00382-013-1942-2

Abdullah, S.S., Malek, M.A., Abdullah, N.S., Kisi, O., Yap, K.S., 2015. Extreme learning machines: a new approach for prediction of reference evapotranspiration. Journal of Hydrology, 527, 184-195. doi:10.1016/j.jhydrol. 2015.04.073

Akpa, S.I.C., Ugbaje, S.U., Bishop, T.F.A., Odeh, I.O.A., 2016. Enhancing pedotransfer functions with environmental data for estimating bulk density and effective cation exchange capacity in a data-sparse situation. Soil Use and Management 32, 644–658. doi:10.1111/sum.12310

Amini, M., Abbaspour, K.C., Khademi, H., Fathianpour, N., Afyuni, M., Schulin, R., 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. European Journal of Soil Science 56, 551–559. doi:10.1111/j.1365-2389.2005.0698.x

Arthur, E., 2017. Rapid estimation of cation exchange capacity from soil water content. European Journal of Soil Science 68, 365–373. doi:10.1111/ejss.12418

Bayat, H., Davatgar, N., Jalali, M., 2014. Prediction of CEC Using Fractal Parameters by Artificial Neural Networks. International Agrophysics 28, 143–152. doi:10.2478/intag-2014-0002

Böhlke, J.-K., 2002. Groundwater recharge and agricultural contamination. Hydrogeology Journal 10, 153–179. doi:10.1007/s10040-001-0183-3

Conradie, M., Kotze, W.A.G., 1989. Comparison of methods for estimating cation exchange capacity of orchard soils. South African Journal of Plant and Soil 6, 136–137. doi:10.1080/02571862.1989.10634497

Delavernhe, L., Pilavtepe, M., Emmerich, K., 2018. Cation exchange capacity of natural and synthetic hectorite. Applied Clay Science 151, 175–180. doi:10.1016/j.clay.2017.10.007

Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research 153, 512-525. doi: 10.1016/j.atmosres.2014.10.016

Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R., 2015. Extreme learning machine: algorithm, theory and applications. Artificial Intelligence Review 44, 103–115. doi:10.1007/s10462-013-9405-z

Dohrmann, R., 2006a. Cation exchange capacity methodology I: An efficient model for the detection of incorrect cation exchange capacity and exchangeable cation results. Applied Clay Science 34, 31–37. doi:10.1016/j.clay.2005.12.006

Dohrmann, R., 2006b. Cation exchange capacity methodology II: A modified silver–thiourea method. Applied Clay Science 34, 38–46. doi:10.1016/j.clay.2006.02.009

Emamgolizadeh, S., Bateni, S.M., Shahsavani, D., Ashrafi, T., Ghorbani, H., 2015. Estimation of soil cation exchange capacity using Genetic Expression Programming (GEP) and Multivariate Adaptive Regression Splines (MARS). Journal of Hydrology 529, 1590–1600. doi:10.1016/j.jhydrol.2015.08.025

Fooladmand, H.R., 2008. Estimating cation exchange capacity using soil textural data and soil organic matter content: A case study for the south of Iran. Archives of Agronomy and Soil Science 54, 381–386. doi:10.1080/03650340802158193

Ghorbani, H., Kashi, H., Hafezi Moghadas, N., Emamgholizadeh, S., 2015. Estimation of Soil Cation Exchange Capacity using Multiple Regression, Artificial Neural Networks, and Adaptive Neuro-fuzzy Inference System Models in Golestan Province, Iran. Communications in Soil Science and Plant Analysis 46, 763–780. doi:10.1080/00103624.2015.1006367

Goldberg, D.E., Deb, K., 1991. A comparative analysis of selection schemes used in

genetic algorithms, in: Foundations of Genetic Algorithms. Elsevier, pp. 69–93.

Goldberg, D.E., Holland, J.H., 1988. Genetic Algorithms and Machine Learning. Mach. Learn. 3, 95–99. https://doi.org/10.1023/A:1022602019183

Haykin, S., 1999. Neural networks: a comprehensive foundation by Simon Haykin. Knowl. Eng. Rev. https://doi.org/10.1017/S0269888998214044

Holland, J.H., 1992. Genetic Algorithms. Sci. Am. 267, 66–72. https://doi.org/10.1038/scientificamerican0792-66

Huang, G.-B., Wang, D.H., Lan, Y., 2011. Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics 2, 107–122. doi:10.1007/s13042-011-0019-y

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501. doi:10.1016/j.neucom.2005.12.126

Huang, G., Zhu, Q., Siew, C., 2004. Extreme Learning Machine : A New Learning Scheme of Feedforward Neural Networks, in: IEEE International Joint Conference on Neural Networks (IJCNN). pp. 985–990. doi:10.1109/IJCNN.2004.1380068

Jafarzadeh, A.A., Pal, M., Servati, M., FazeliFard, M.H., Ghorbani, M.A., 2016. Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction. International Journal of Environmental Science and Technology 13, 87–96. doi:10.1007/s13762-015-0856-4

Jain, A., Srinivasulu, S., 2004. Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. Water Resour. Res. https://doi.org/10.1029/2003WR002355

Kalkhajeh, Y.K., Arshad, R.R., Amerikhah, H., Sami, M., 2012. Comparison of

multiple linear regressions and artificial intelligence-based modeling techniques for prediction the soil cation exchange capacity of Aridisols and Entisols in a semi-arid region. Aust. J. Agric. Eng. 3, 39.

Ketterings, Q., Reid, S., Rao, R., 2007. Cation Exchange Capacity. Cornell University Cooperative Extension, Agronomy Fact Sheet #22, Ithaca, NY.

Khodaverdiloo, H., Momtaz, H., Liao, K., 2018. Performance of Soil Cation Exchange Capacity Pedotransfer Function as Affected by the Inputs and Database Size. Clean Soil Air Water 46, 1700670. doi:10.1002/clen.201700670

Khorshidi, M., Lu, N., 2017. Determination of Cation Exchange Capacity from Soil Water Retention Curve. Journal of Engineering Mechanics 143, 04017023. doi:10.1061/(ASCE)EM.1943-7889.0001220

Kim, S., Kim, H.S., 2008. Neural Networks and Genetic Algorithm Approach for Nonlinear Evaporation and Evapotranspiration Modeling. Journal of Hydrology 351, 299-317. doi:10.1016/j.jhydrol.2007.12.014

Kim, S., Seo, Y., Rezaie-Balf, M., Kisi, O., Ghorbani, M.A., Singh, V.P., 2019. Evaluation of Daily Solar Radiation Flux using Soft Computing Approaches based on Different Meteorological Information: Peninsula vs Continent. Theoretical and Applied Climatology 137, 693-712. doi:10.1007/s00704-018-2627-x

Kim, S., Singh, V.P., 2014. Modeling Daily Soil Temperature using Data-Driven models and Spatial Distribution. Theoretical and Applied Climatology 118, 465-479. doi:10.1007/s00704-013-1065-z

Kim, S., Singh, V.P., Lee, C.J., Seo, Y., 2015. Modeling the Physical Dynamics of Daily Dew Point Temperature using Soft Computing Techniques. KSCE Journal of Civil Engineering 19, 1930-1940. doi:10.1007/s12205-014-1197-4

Krogh, L., Breuning-madsen, H., Greve, M.H., 2000. Cation-Exchange Capacity Pedotransfer Functions for Danish Soils. Acta Agriculturae Scandinavica, Section B - Soil & Plant Science 50, 1–12. doi:10.1080/090647100750014358

Lambooy, A.M., 1984. Relationship between cation exchange capacity, clay content and water retention of Highveld soils. South African Journal of Plant and Soil 1, 33–38. doi:10.1080/02571862.1984.10634106

Liao, K., Xu, S., Wu, J., Zhu, Q., An, L., 2014. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. Journal of Plant Nutrition and Soil Science 177, 775–782. doi:10.1002/jpln.201300176

Liao, K., Xu, S., Zhu, Q., 2015. Development of ensemble pedotransfer functions for cation exchange capacity of soils of Qingdao in China. Soil use and management 31, 483–490. doi:10.1111/sum.12207

Martínez-Martínez, J.M., Escandell-Montero, P., Soria-Olivas, E., Martín-Guerrero, J.D., Magdalena-Benedito, R., Gómez-Sanchis, J., 2011. Regularized extreme learning machine for regression problems. Neurocomputing 74, 3716–3721. doi:10.1016/j.neucom.2011.06.013

McClelland, J.L., Rumelhart, D.E., 1988. Explorations in parallel distributed processing: A handbook of models, programs, and exercises. Explor. Parallel Distrib. Process. a Handb. Model. Programs Exerc. 344, ix, 344 p. https://doi.org/10.2307/1423065

Mengel, D.B., 1993. Fundamentals of Soil Cation Exchange Capacity (CEC). Purdue University Cooperative Extension Service, Agronomy Guide AY-238, West Lafayette.

Nourani, V., Khanghah, T.R., Sayyadi, M., Prof, A., Student, M.S., Student, B.S., 2013. Application of the Artificial Neural Network to monitor the quality of treated water 3, 38–45.

Niu, W.J., Feng, Z.K., Cheng, C.T., Zhou, J.Z., 2018. Forecasting daily runoff by extreme learning machine based on quantum-behaved particle swarm optimization. Journal of Hydrologic Engineering 23(3), 04018002. doi:10.1061/(ASCE)HE.1943-5584.0001625

Obalum, S.E., Watanabe, Y., Igwe, C.A., Obi, M.E., Wakatsuki, T., 2013. Improving on the Prediction of Cation Exchange Capacity for Highly Weathered and Structurally Contrasting Tropical Soils from their Fine-Earth Fractions. Communications in Soil Science and Plant Analysis 44, 1831–1848. doi:10.1080/00103624.2013.790401

Parfitt, R.L., Giltrap, D.J., Whitton, J.S., 1995. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. Communications in Soil Science and Plant Analysis 26, 1343–1355. doi:10.1080/00103629509369376

Şahin, M., Kaya, Y., Uyar, M., Yıldırım, S., 2014. Application of extreme learning machine for estimating solar radiation from satellite data. International Journal of Energy Research 38(2), 205-212. doi:10.1002/er.3030

Seybold, C.A., Grossman, R.B., Reinsch, T.G., 2004. Predicting Cation Exchange Capacity for Soil Survey Using Linear Models. Soil Science Society of America Journal 69, 856–863. doi:10.2136/sssaj2004.0026

Seyedmohammadi, J., Esmaeelnejad, L., Ramezanpour, H., 2016. Determination of a suitable model for prediction of soil cation exchange capacity. Modeling Earth Systems and Environment 2, 156. doi:10.1007/s40808-016-0217-4

Sharma, A., Weindorf, D.C., Wang, D.D., Chakraborty, S., 2015. Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). Geoderma 239, 130–134. doi:10.1016/j.geoderma.2014.10.001

Shekofteh, H., Ramazani, F., Shirani, H., 2017. Optimal feature selection for predicting soil CEC: Comparing the hybrid of ant colony organization algorithm and adaptive network-based fuzzy system with multiple linear regression. Geoderma 298, 27–34. doi:10.1016/j.geoderma.2017.03.010

Tan, K.H., Dowling, P.S., 1984. Effect of organic matter on CEC due to permanent and variable charges in selected temperate region soils. Geoderma 32, 89–101. doi:10.1016/0016-7061(84)90065-X

Tang, L., Zeng, G., Nourbakhsh, F., Shen, G.L., 2009. Artificial Neural Network Approach for Predicting Cation Exchange Capacity in Soil Based on Physico-

Chemical Properties. Environmental Engineering Science 26, 137–146. doi:10.1089/ees.2007.0238

Thomas, P.J., Baker, J.C., Zelazny, L.W., 2000. An Expansive Soil Index for Predicting Shrink–Swell Potential. Soil Science Society of America Journal 64, 268–274. doi:10.2136/sssaj2000.641268x

van Erp, P.J., Houba, V.J.G., van Beusichem, M.L., 2001. Actual Cation Exchange Capacity of Agricultural Soils and its relationship with pH and content of organic carbon and clay. Communications in Soil Science and Plant Analysis 32, 19–31. doi:10.1081/CSS-100102990

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Reviews of Geophysics 55, 1199–1256. doi:10.1002/2017RG000581

Wang, Y., Cao, F., Yuan, Y., 2011. A study on effectiveness of extreme learning machine. Neurocomputing 74, 2483–2490. doi:10.1016/j.neucom.2010.11.030

Wolf, B., 1999. Cation exchange capacity and anion exchange, in: The Fertile Triangle : The Interrelationship of Air, Water, and Nutrients in Maximizing Soil Productivity. Food Products Press, New York, pp. 177–188.

Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.W., 2018. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology 569, 387-408. doi:10.1016/j.jhydrol.2018.11.069

Zolfaghari, A.A., Taghizadeh-Mehrjardi, R., Moshki, A.R., Malone, B.P., Weldeyohannes, A.O., Sarmadian, F., Yazdani, M.R., 2016. Using the nonparametric k-nearest neighbor approach for predicting cation exchange capacity. Geoderma 265, 111–119. doi:10.1016/j.geoderma.2015.11.012

Figures

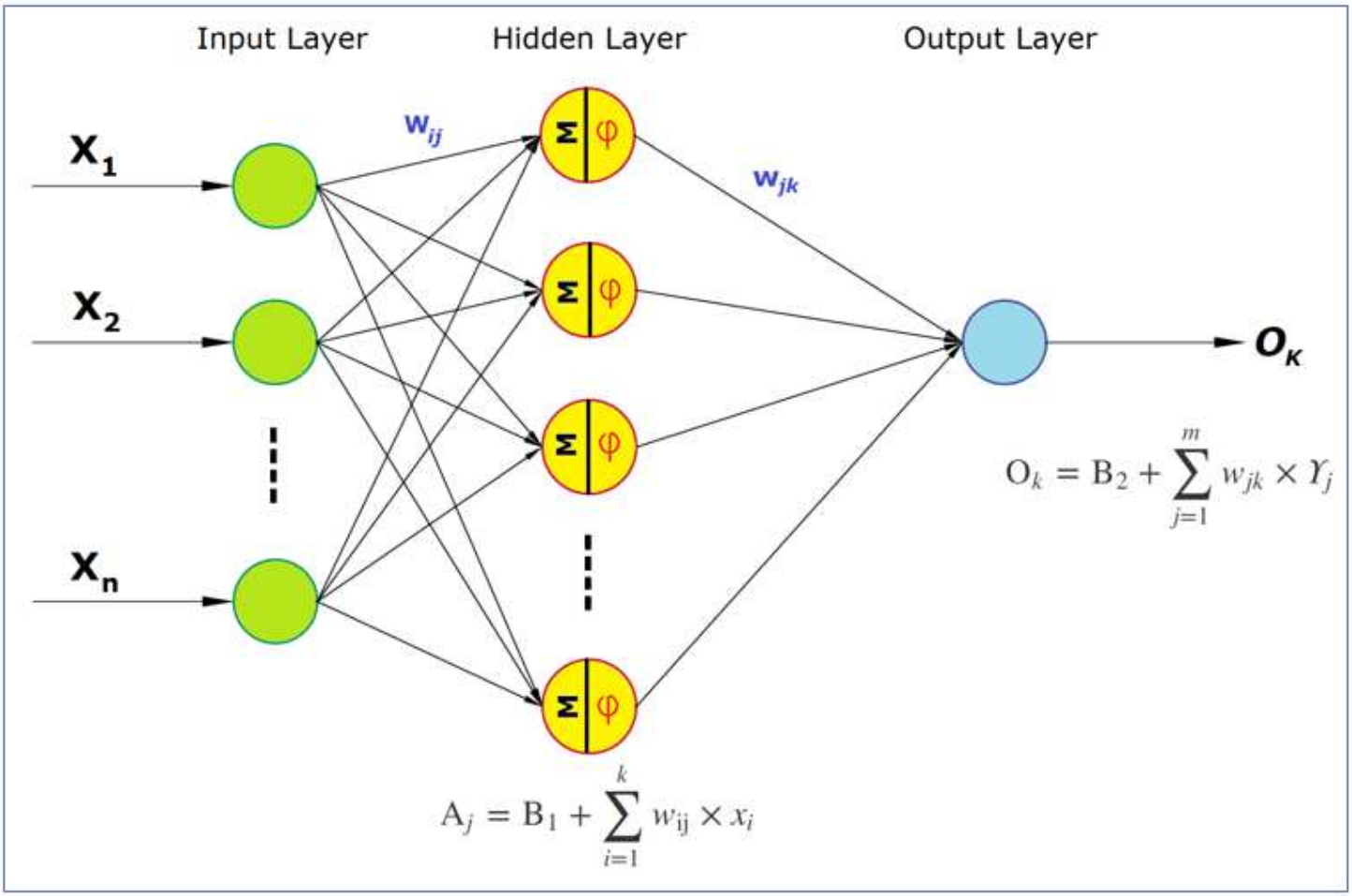


Figure 1

The architecture of the MLP network model.

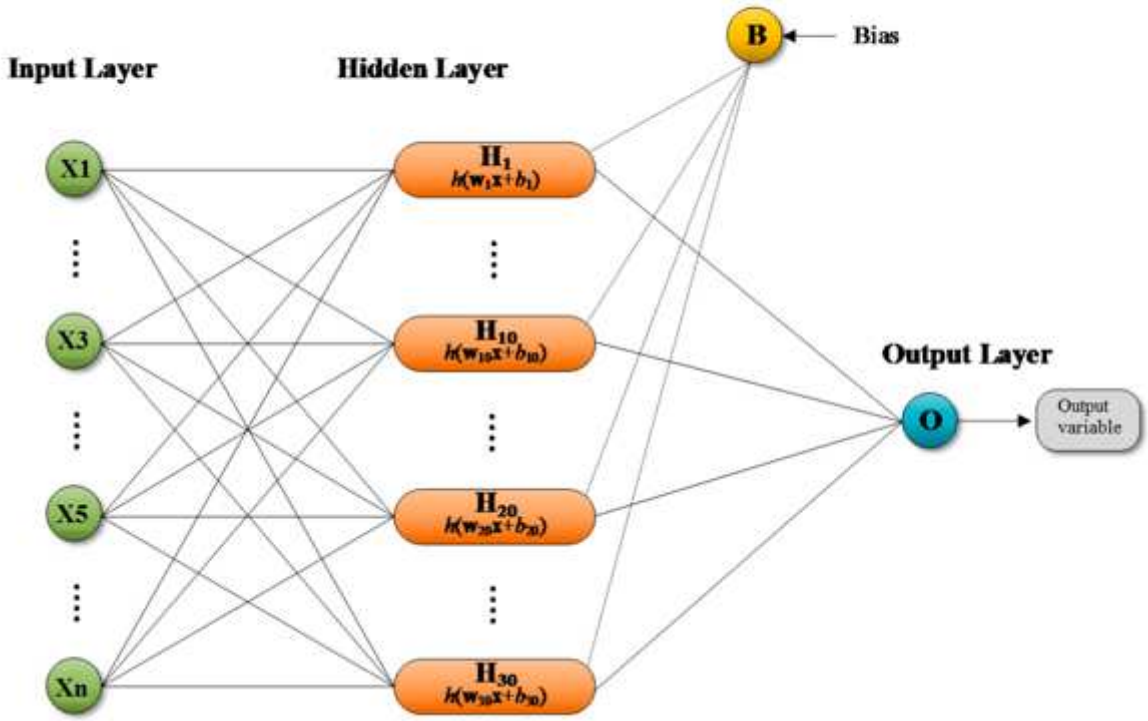


Figure 2

Architecture of ELM Model

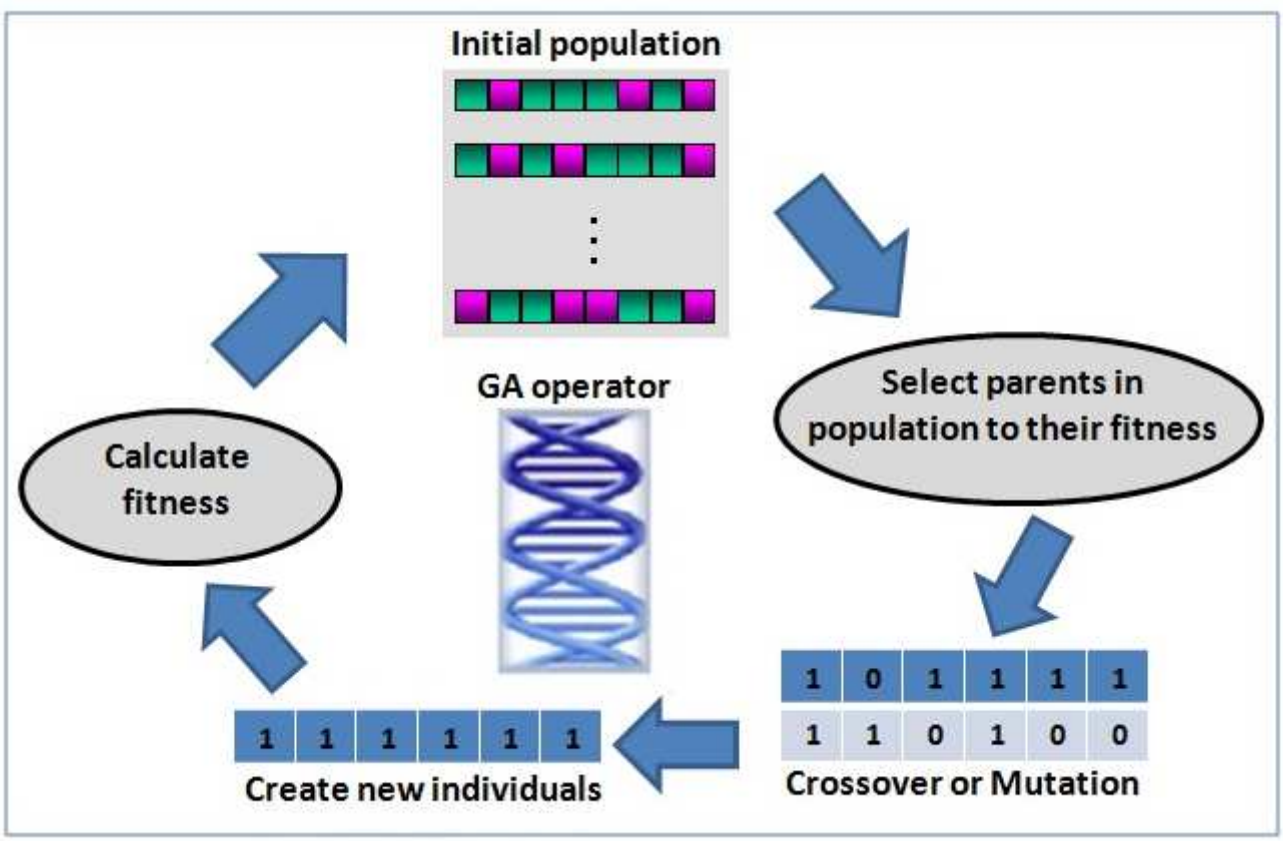


Figure 3

The schematic diagram of genetic algorithm.

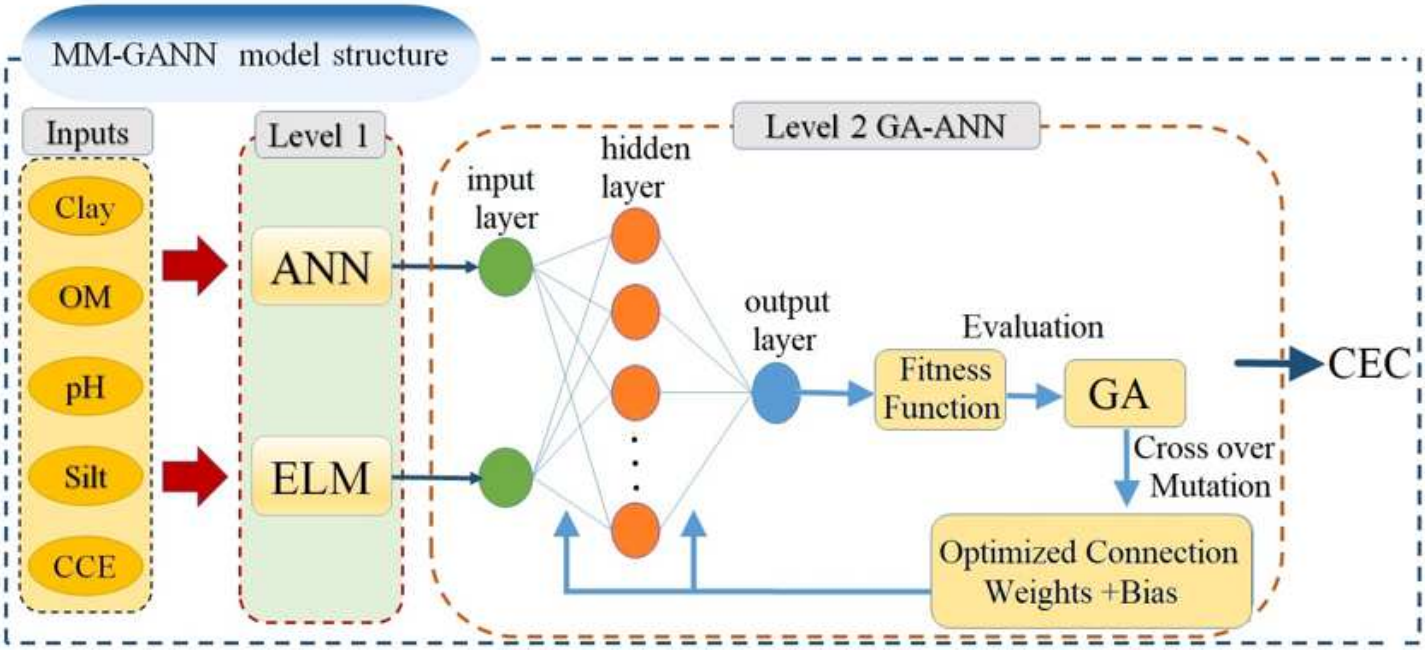


Figure 4

The structure of the proposed MM-GANN model for predicting soil cation exchange capacity.

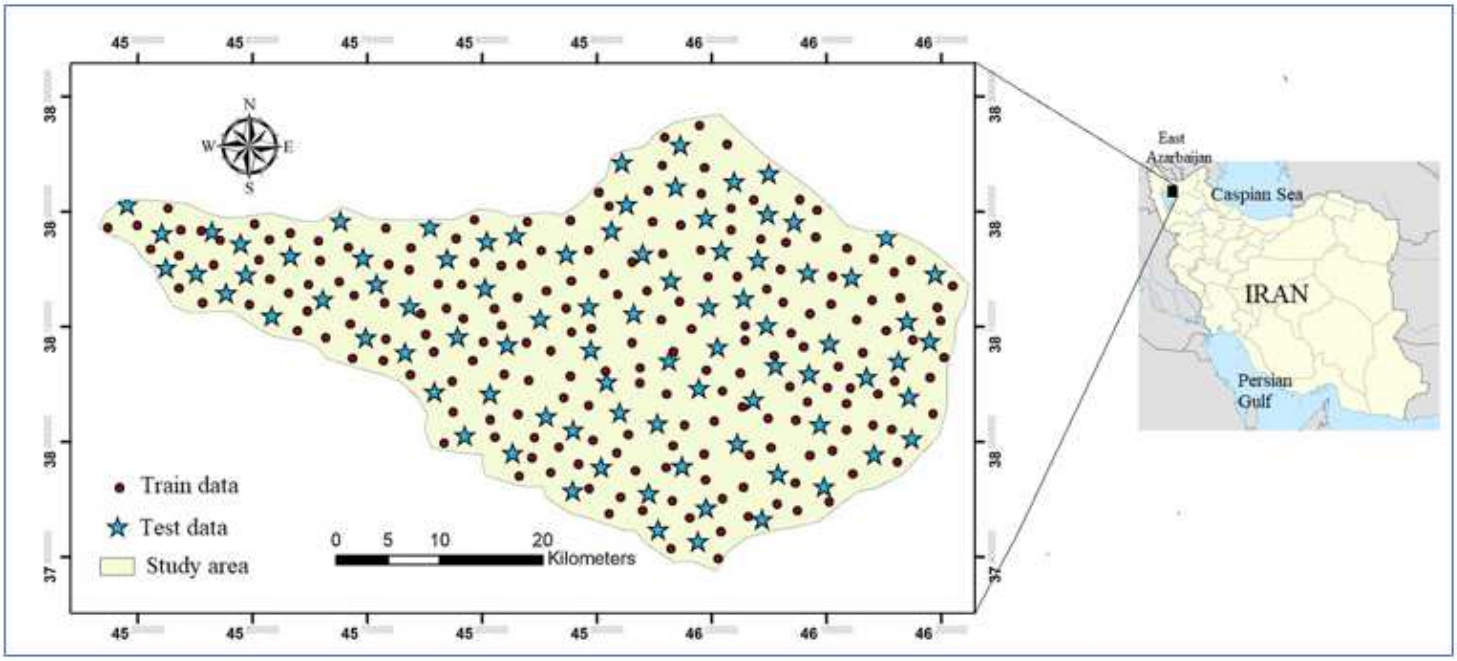


Figure 5

The location of the investigated study area along with sampling points. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever

on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

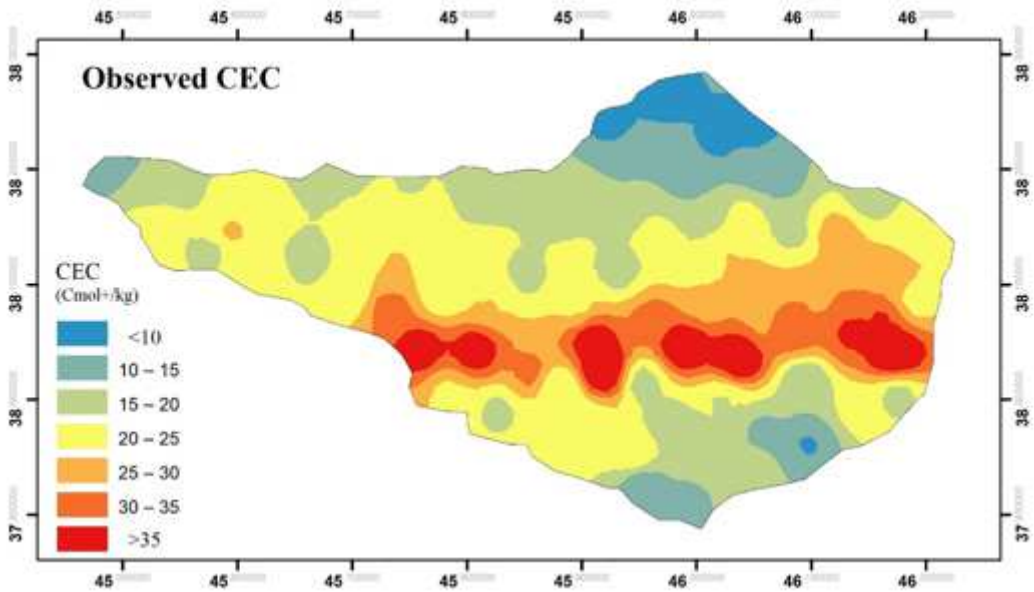


Figure 6

The spatial distribution map of observed soil cation exchange capacity. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

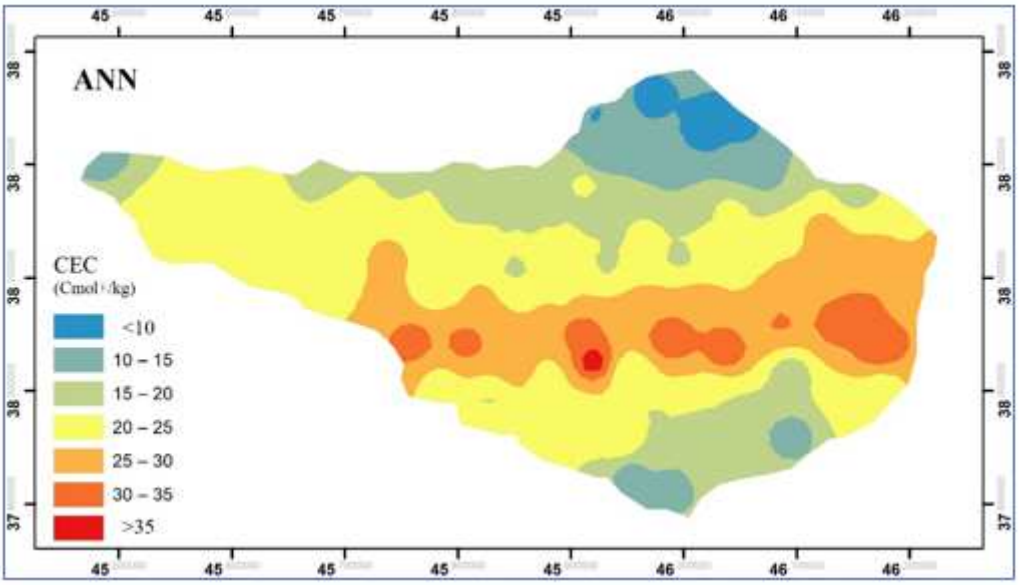


Figure 7

The spatial distribution map of applied ANN predicted soil CEC (ANN model with 5 inputs). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

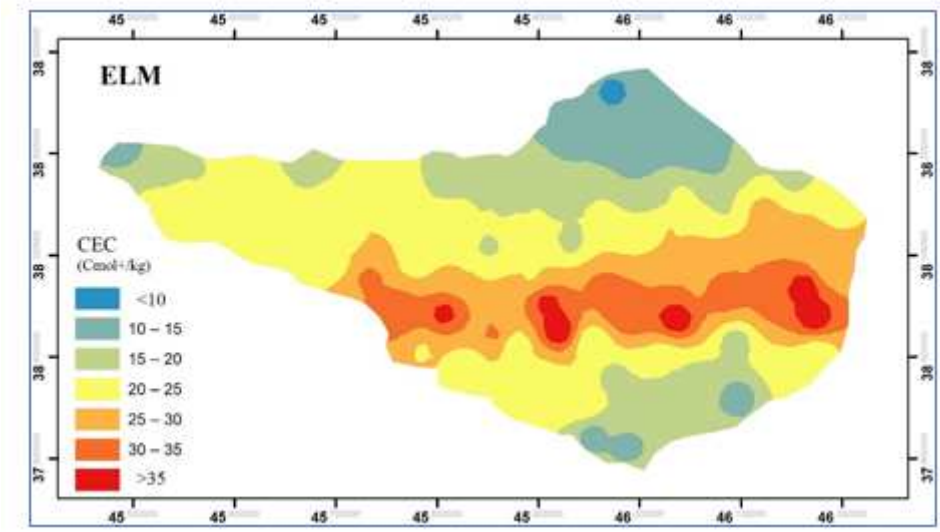


Figure 8

The spatial distribution map of the applied ELM predicted soil CEC (ELM model with 5 inputs). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

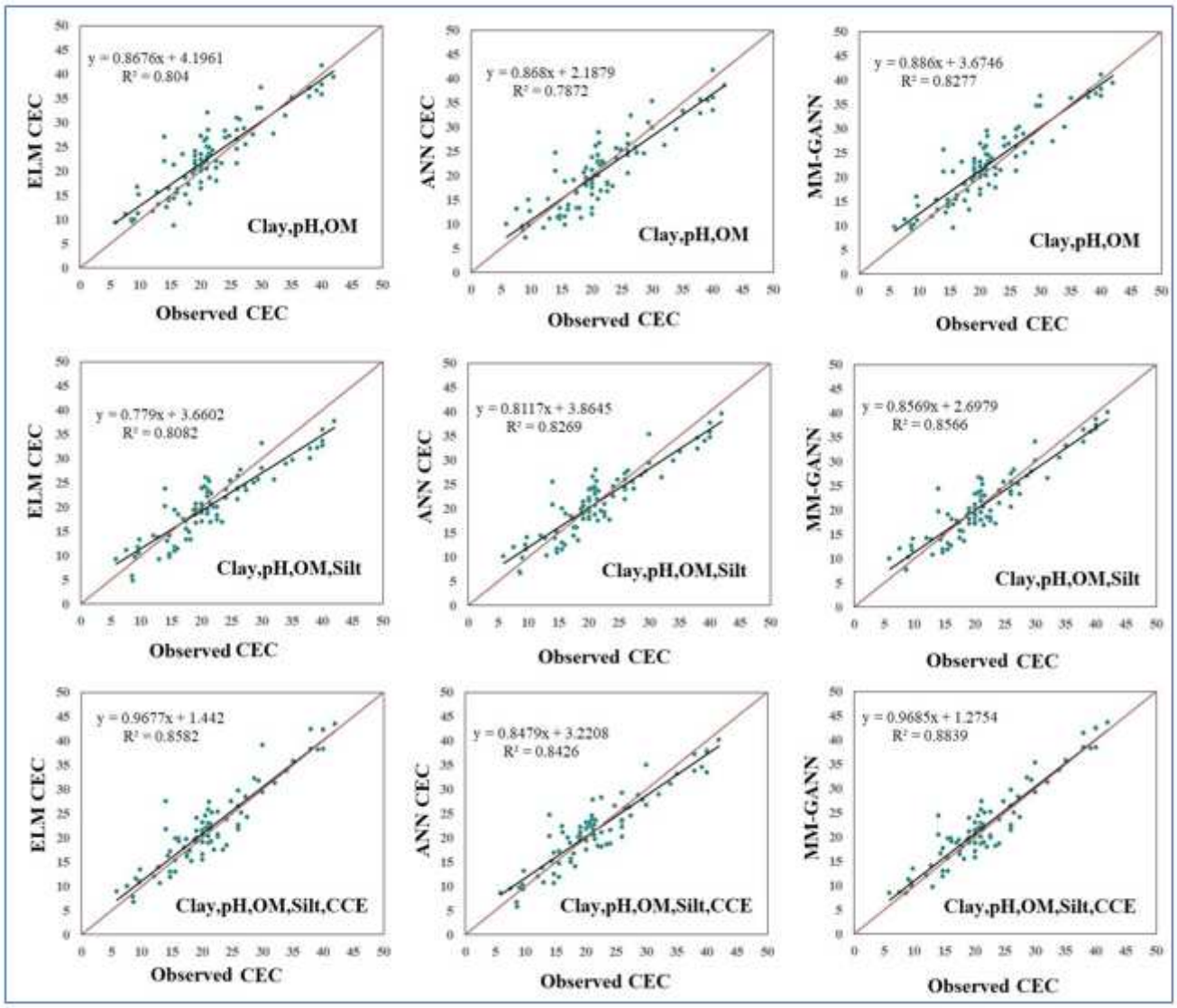


Figure 9

The scatter plot between the observed and estimated soil CEC over the testing test stage using ELM, ANN and MM-GANN models.

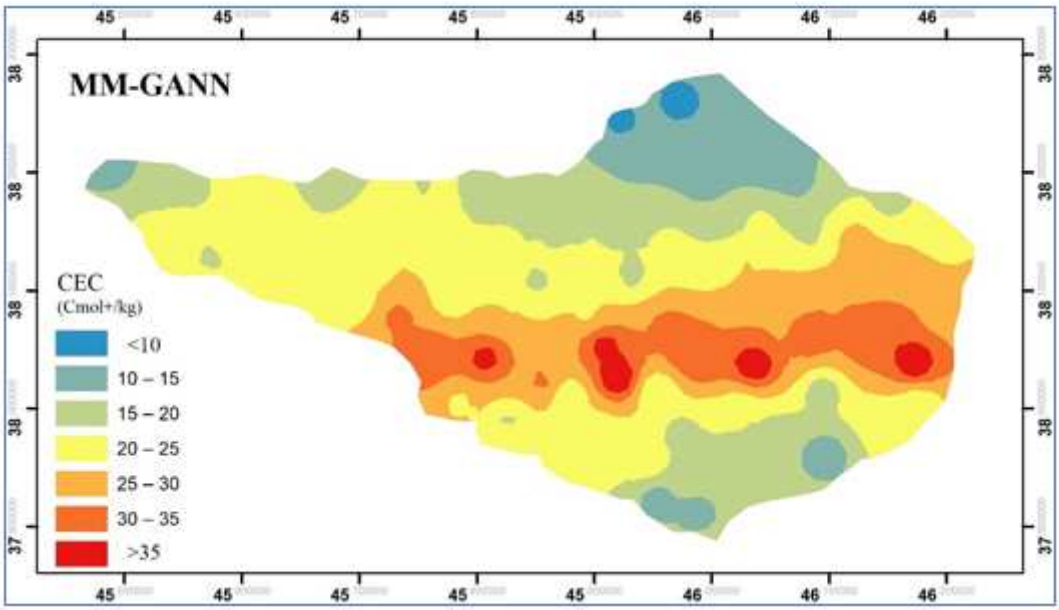
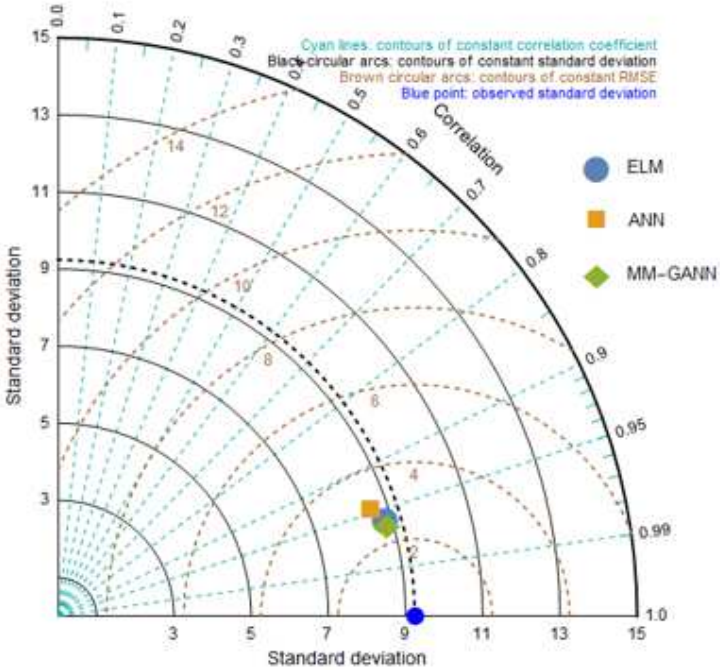
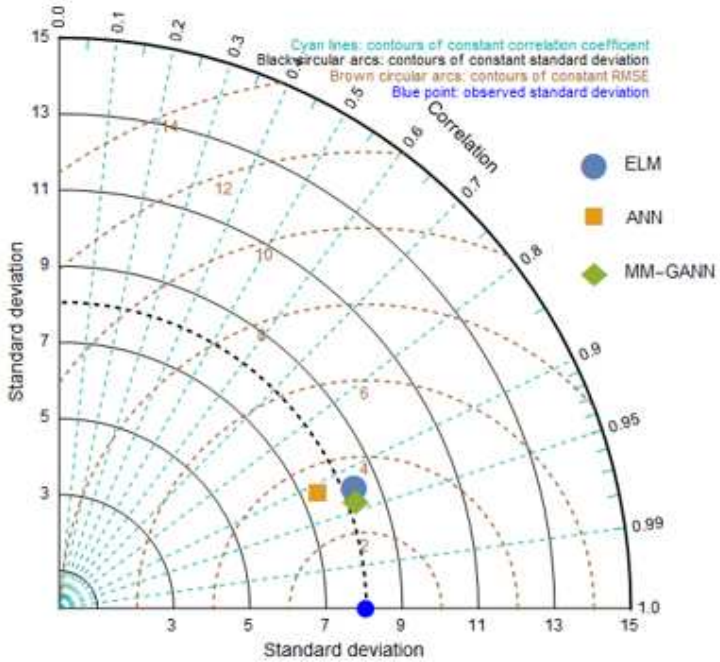


Figure 10

The spatial distribution map of the applied MM-GANN predicting soil CEC (MM-GANN model calibrated with CEC***(ELM) and CEC***(ANN) as inputs). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



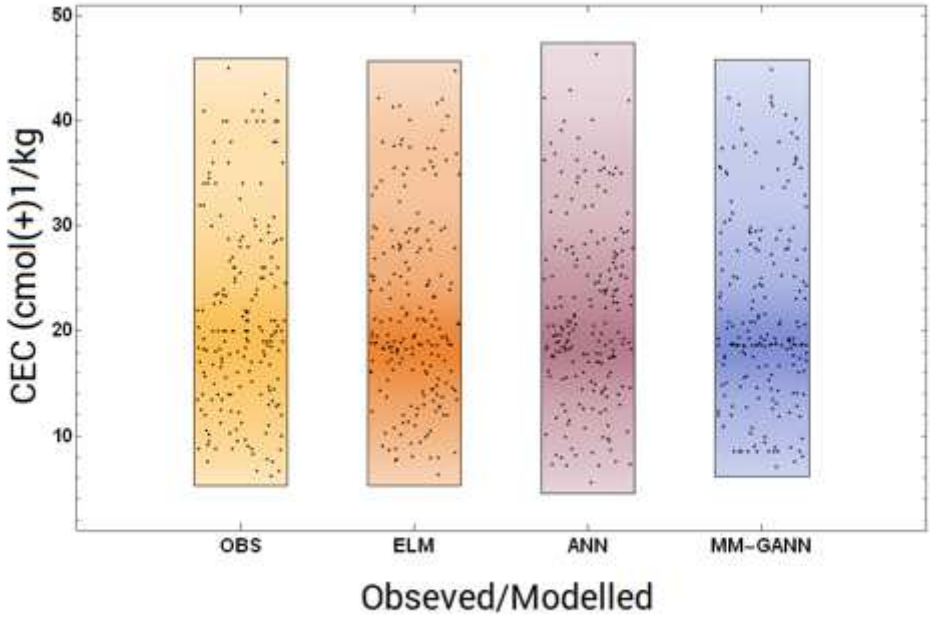
a. Taylor Diagram for Training Dataset



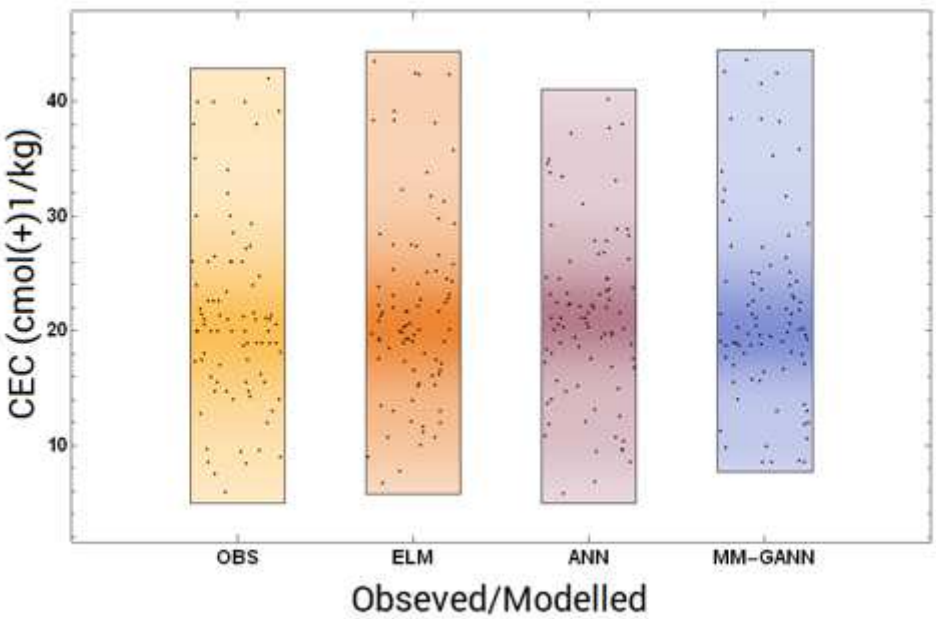
b. Taylor Diagram for Testing Dataset

Figure 11

Taylor diagrams revealing the performance of the applied predictive models for the soil CEC modeling.



a. Point Density Plot for Training Dataset



b. Point Density Plot for Testing Dataset

Figure 12

Point density plots for performance evaluation of soil CEC models

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.pdf