



Estimation of the recharging rate of groundwater using random forest technique

Parveen Sihag¹ · Anastasia Angelaki² · Barkha Chaplot³

Received: 28 February 2020 / Accepted: 22 June 2020 / Published online: 3 July 2020
© The Author(s) 2020

Abstract

Accurate knowledge of the recharging rate is essential for several groundwater-related studies and projects mainly in the water scarcity regions. In this study, a comparison between different methods of soft computing-based models was obtained in order to evaluate and select the most suitable and accurate method for predicting the recharging rate of groundwater, as the natural recharging rate of the groundwater is important in efficient groundwater resource management and aquifer recharge. Experimental data have been used to investigate the improved performance of Gaussian process (GP), M5P and random forest (RF)-based regression method and evaluate the potential of these techniques in the prediction of natural recharging rate. The study also compares the prediction of recharging rate to empirical (Kostiakov model, multilinear regression, multi-nonlinear regression) equations. The RF method was selected for the recharging rate prediction and was compared with the M5P tree, GP and also empirical models. While GP, M5P tree and empirical models provide good quality of prediction performance, RF model showed superiority among them with coefficient of correlation (R) values as 0.98 and 0.91 for training and testing, respectively. Out of 106 observations collected from laboratory experiments, 73 were used for developing different models, whereas rest 33 observations were used for the assessment of the models' performance. Sensitivity analysis recommends that time parameter (t) is the main influencing parameter, which is crucial for the prediction of the recharging rate. RF-based model is suitable for accurate prediction of recharging rate of groundwater.

Keywords Recharging rate · Random forest · Gaussian process regression · M5P tree

Introduction

Groundwater is a gift and a considerable element of the hydrologic cycle, during which water moves vertically toward the center of Earth. Aquifer recharge takes place when water moves either from the land surface, or from

the vadose zone into the saturated zone. Quantitative estimation of the recharge rate is crucial in order to understand large-scale hydrologic processes, and it is important for evaluating the sustainability of groundwater supplies. The extensive availability of fresh groundwater is the main cause for its usage as a source of irrigation and drinking, universally (Alley et al. 2002). However, the large amount of crops is grown by irrigated cultivation, which mainly depends upon the available amounts of groundwater. Groundwater plays a fundamental role in river flow mainly in dry periods and is essential to several lagoons, wetlands and lakes (Rockström et al. 2010). Besides, the life of human, vegetation and aquatic animals rely on the groundwater that moves to rivers, lagoons, ponds and wetlands. Last few years, the level of groundwater gradually decreases due to extensive use in various purposes. The quantity of water that may be collected from the aquifer without causing exhaustion is mainly depended upon the recharge of groundwater (Freeze 1969). Thus, the estimation of recharging rate of the ground is essential for water

✉ Parveen Sihag
parveen12sihag@gmail.com

Anastasia Angelaki
anaggel@agr.uth.gr

Barkha Chaplot
barkhachaplotjain@gmail.com

¹ Department of Civil Engineering, Shoolini University, Solan, Himachal Pradesh 173229, India

² Department of Agriculture, Crop Production and Rural Environment, School of Agricultural Sciences, University of Thessaly, Volos, Greece

³ Department of Geography, M.J.K. College, Babasaheb Bhimrao Ambedkar Bihar University, Bettiah, India

supply and groundwater resource management. It is very necessary for areas where economic development depends on groundwater resources.

Precipitation is the principal source for the recharging of groundwater. The amount of water that will ultimately arrive at the water table is defined as natural groundwater recharge (Sophocleous 2002). The quantity of the recharge depends on the period and intensity of precipitation, flood, soil type, soil moisture conditions etc. As there is spatial and temporal variability of the recharging rate of the soil, it is crucial to be precise to the selection of recharging estimation methods. The suitability of recharging models is site-specific due to spatial variation in recharging rate through the soil. Experimentally estimation of recharging rate is a tedious and time-consuming task (Sihag et al. 2017; Kumar and Sihag 2019). Water storage ability differs at various soil textures and soil physical properties (Angelaki et al. 2013). Sand practical consists of relatively greater pore size than clay and thus has higher recharging rate and very small water-holding ability. The actual rate at which water percolates into the soil at any time is identified as the recharging rate. The significance of the recharging process imposed the researchers to generate several models (Green and Ampt 1911; Richards 1931; Kostikov 1932; Horton 1941; Philip 1957; Holtan 1961; Singh and Yu 1990) as well as Modified Kostikov model, SCS model and Novel model. These models are divided into three groups such as physical models, semi-empirical models and empirical models. The correct determination of the recharging rate is essential for several groundwater-related studies and projects (Singh et al. 2018).

Last few years, data mini-techniques like neural network, support vector machines, adaptive neuro-fuzzy inference system (ANFIS), random forest (RF), Gaussian process regression (GP) and M5P model tree have been successfully implemented in civil engineering and water resources problems (Kisi et al. 2012; Ebtehaj and Bonakdari 2013; Parsaie et al. 2016; Parsaie and Haghiabi 2017a, b, c; Qishlaqi et al. 2017; Parsaie et al. 2018a, b; Sihag 2018; Sihag et al. 2018a, b, 2019; Parsaie et al. 2020). There are several convention models, but these outcomes are not general on different location and conditions. The aim of this study was to develop a new model for the accurate prediction of natural recharging rate of groundwater. GP-, M5P- and RF-based regression methods were selected for the prediction of natural recharging rate, and a comparison between the empirical equations (Kostikov model, multi-linear regression (MLR) and multi-nonlinear regression (MNL)) and soft computing-based models has been done. Most important parameter was selected using sensitivity analysis, and Taylor diagram and predicted error box plot were also used to investigate the accuracy of the applied models.

Methodology and dataset

Experimental procedure

In order to investigate the recharging of water through different soil types, three soil samples of different hydrodynamic parameters were used. Soil samples were collected using core cutter from three different locations (Greece). After drying the soils at 105 °C, granulometric analysis has been done. Each soil sample passed through a certain series of sieves with descending diameters. Bulk density, the moisture of the saturated soil and recharging rates were measured in the laboratory, for all soil samples. Apparatus selected for experimentation is shown in Fig. 1. Each soil sample was packed in a transparent column of Plexiglas. In order to achieve good homogeneity of the soil porosity, the column of Plexiglas was filled with soil using a tube with a double sieve in it. TDR probes were inserted carefully at certain locations of the column, and to avoid water leakage, silicon was used for water proofing. As there was an intention to achieve homogeneous steady rain and in addition to achieve a 2 mm head boundary at the top of the soil column, two volumetric tubes were used. One volumetric tube was

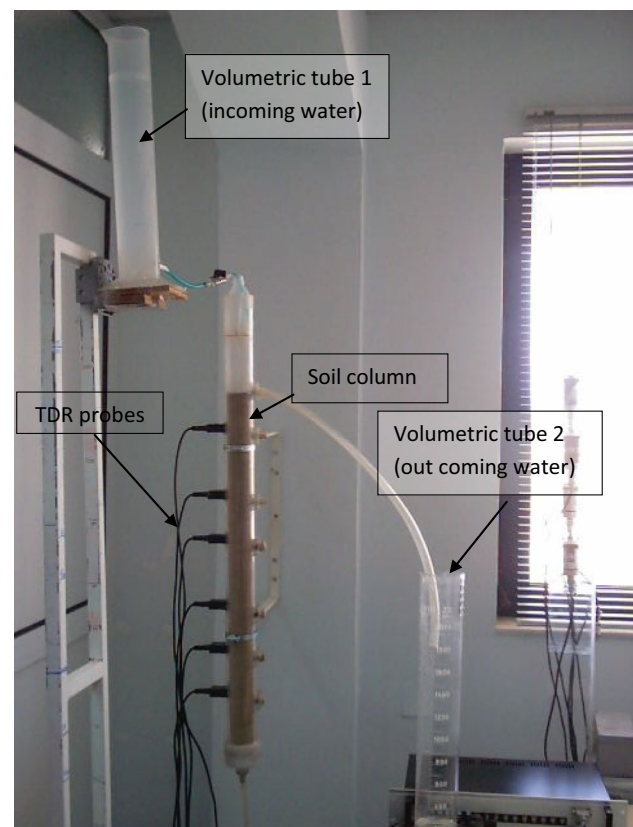


Fig. 1 Experimental procedure

used for pouring water into the column, while the other one was used as an outpouring container. The incoming—into the soil—water volume was calculated by subtracting the volume of water of the second tube (outcoming) from the volume of the first tube (incoming). While the wet profile was moving into the soil, TDR was automatically measuring the moisture of the soil at certain locations and at certain time circles.

Dataset

The entire dataset contains 106 experimental observations from the laboratory. Data were divided into two separate groups, training and testing, respectively. Training data involve 70% of the total data chosen randomly from the whole data set, while testing data involve the remaining 30% of the whole data. The features of the training and testing data sets are represented in Table 1, where time, sand, clay, silt, bulk density and moisture content are input parameters and recharging rate of the soil is the target.

Modeling approaches

Gaussian process regression (GP)

GP regression relies upon the assumption that nearby observation must share the information mutually and it’s an approach of mentioning earlier straight over the function space. The simplification of Gaussian distribution is known as Gaussian regression. The matrix and vector of Gaussian distribution are expressed as covariance and mean in Gaussian process regression. Due to having earlier information of function reliance and data, the validation for generalization

is not essential. The GP regression models are capable to recognize the foresee distribution consequent to the input test data (Rasmussen and Williams 2006).

A GP is the selection of numbers of the random variable, any finite number of them has a collective multivariate Gaussian distribution. Assume p and q are input and output domain respectively, there upon x pairs (g_i, h_i) are drawn freely and equivalently distribution. For regression, it is assumed that $h \subseteq \text{Re}$ than a GP on p is expressed by the mean function $\nu_0 : p \rightarrow \text{Re}$ and covariance function $\mu : p \times p \rightarrow \text{Re}$. The kernels used in present work are radial basis kernel (RBF) and Pearson VII kernel function which is shown below:

1. $\text{RBF} = e^{-\gamma \|x_i - x_j\|^2}$
2. $\text{PUK} = \left(1 / \left[1 + \left(2 \sqrt{\|x_i - x_j\|^2} \sqrt{2^{(1/\omega)} - 1} / \sigma \right)^2 \right]^\omega \right)$

where γ, σ and ω are primary parameters of the kernels.

M5P model (M5P)

M5P tree, initially introduced by Quinlan (1992), is selected to grow a decision tree by engaging the linear regression function method at nodes to build a model which recommend a correlation amid the output value of the preparing cases and value of input attributes. The splitting method is supplied at each node instead to achieve the maximum knowledge with minimum variation in the inter-subset class value down to each branch. The splitting method will be converged when there are diminutive variations among the class values of the instances or left only a few instances or when a tree is pruned back. The fully grown tree demonstrates the very good quality structure and

Table 1 Features of the data set

| Parameters | | Min. | Max. | Mean | Std. Dev. | Kurtosis | Skewness |
|--------------------------|---------|-------|--------|-------|-----------|----------|----------|
| Training data set | | | | | | | |
| t (time) | min | 4.00 | 132.00 | 56.42 | 39.21 | - 1.24 | 0.21 |
| S (sand) | % | 76.00 | 93.00 | 80.45 | 4.96 | 1.66 | 1.29 |
| Si (silt) | % | 3.00 | 13.00 | 9.67 | 3.23 | - 0.60 | - 0.44 |
| C (clay) | % | 4.00 | 11.00 | 9.88 | 1.95 | 6.12 | - 2.65 |
| D (density) | g/cc | 1.48 | 1.56 | 1.52 | 0.03 | - 1.89 | - 0.21 |
| Mc (moisture content) | % | 0.29 | 0.38 | 0.34 | 0.03 | - 1.35 | 0.03 |
| $R(t)$ (recharging rate) | cm/min. | 0.09 | 1.77 | 0.34 | 0.32 | 13.10 | 3.31 |
| Testing data set | | | | | | | |
| t (time) | min | 4.00 | 132.00 | 56.42 | 39.21 | - 1.24 | 0.21 |
| S (sand) | % | 76.00 | 93.00 | 80.45 | 4.96 | 1.66 | 1.29 |
| Si (silt) | % | 3.00 | 13.00 | 9.67 | 3.23 | - 0.60 | - 0.44 |
| C (clay) | % | 4.00 | 11.00 | 9.88 | 1.95 | 6.12 | - 2.65 |
| D (density) | g/cc | 1.48 | 1.56 | 1.52 | 0.03 | - 1.89 | - 0.21 |
| Mc (moisture content) | % | 0.29 | 0.38 | 0.34 | 0.03 | - 1.35 | 0.03 |
| $R(t)$ (recharging rate) | cm/min. | 0.09 | 1.77 | 0.34 | 0.32 | 13.10 | 3.31 |

forecast correctness due to presenting more probable linearity at the leaf node (Singh et al. 2017).

Random forest (RF)

Random forest algorithm is used to generate a model which includes a group of many trees. Each tree illustrates the specific classification and votes the classification. The forest chooses the classification which has the maximum voting in the forest. The tree is fully grown if N is the number of cases at the training set. N cases at random with the substitute from actual data may be the input data set to fully grown the tree. The m variables are chosen arbitrarily out of K input variables for the best split, the value of m should be less than K and constant. The tree is grown without pruning up to the highest extent. RF can work efficiently and exactly with the huge and complex data set.

Empirical models

Kostiakov model

An empirical model was proposed by Kostiakov (1932) in order to estimate the recharging rate:

$$R(t) = at^{-b} \quad (1)$$

$$R(t) = 2.7563t^{-0.6529} \quad (2)$$

where $R(t)$ is the recharging rate at time $t(\text{LT}^{-1})$, t is the recharging time (T), a and b are dimensionless empirical constants.

Multiple linear regression (MLR)

MLR is implemented on more than one predictor parameters. The common structure of the MLR model is:

$$Z = c_0 + x_1^{c_1} + x_2^{c_2} + x_3^{c_3} x_4^{c_4} + \dots + x_n^{c_n} \quad (3)$$

$$R(t) = 0.925 - 0.0012t + 0.0187S + 0.103Si - 0.189C + 0.4089D - 5.173Mc \quad (4)$$

Multiple nonlinear regression (MNLr)

Multiple nonlinear regression (MNLr) is applied on more than one predictor parameters. The common structure of the MNLr model is:

$$Z = c_0 x_1^{c_1} x_2^{c_2} x_3^{c_3} x_4^{c_4} \dots x_n^{c_n} \quad (5)$$

$$R(t) = 0.0648t^{-0.4694} S^{0.438} Si^{-0.839} C^{0.305} D^{4.33} Mc^{0.4047} \quad (6)$$

where Z is the normal value represented as a function of n -number of independent parameters $x_1, x_2, x_3, \dots, x_n$, in which the values of coefficients, $c_0, c_1, c_2, c_3, \dots, c_n$, are unidentified. These values correspond to the local behavior and are evaluated by the least square technique.

Model assessment

Four most popular equations were used to assess the performance of various data mining methods and empirical equations, such as correlation coefficient (R), mean square error (MSE), root mean square error (RMSE) and Nash–Sutcliffe model efficiency (NSE) values (Sihag et al. 2020).

$$R = \frac{a \sum mn - (\sum m)(\sum n)}{\sqrt{a(\sum m^2) - (\sum m)^2} \sqrt{a(\sum n) - (\sum n)^2}} \quad (7)$$

$$\text{MSE} = \frac{1}{a} \sum_{i=1}^a (m - n)^2 \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{a} \left(\sum_{i=1}^a (m - n)^2 \right)} \quad (9)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^a (m - n)^2}{\sum_{i=1}^a (m - \bar{m})^2} \quad (10)$$

where m is the actual value, n is the predicted value, \bar{m} is the mean of actual value and a is the number of values.

Implementation of machine learning methods

Four standard statistical measures: R , MSE, RMSE and NSE were chosen to judge the performance of the data mining methods and empirical equations. Numerous trials were carried out to find the optimum value of the primary parameters. The upper range of R , NSE and a lesser range of MSE, RMSE indicates superior estimation precision of the models. The number of trees to be developed (k) in the forest and the number of features or variables selected (m) at each node to generate a tree are the two standard primary parameters essential for random forest regression. In M5P, calibration of models has been done by means of changing the value of no. of instances allowed at each node (m), while in Gaussian process regression Gaussian noise, γ , σ and ω are the primary parameters. The selected primary parameters of the modeling approaches are presented in Table 2.

Table 2 Primary parameters

| Machine learning approach | Primary parameters | |
|---------------------------|-------------------------|--|
| M5P | $m=4$ | |
| GP | PUK | Gaussian noise = 0.01, $\gamma=2$ |
| | RBF | Gaussian noise = 0.01, $\omega=0.1$, $\sigma=0.1$ |
| RF | $k=1$, $m=1$, $l=100$ | |

Results and discussion

All empirical equations showed good performance when estimating the natural recharging rate of groundwater using the current dataset, except Kostiakov model. Results of each empirical equation were plotted versus the actual data, and the results are shown in Fig. 2. Standard error indices consisting of R , RMSE, MSE and NSE were used to assess the precision of the empirical equations (observe Table 3). The MNLR equation with R value as 0.90, MSE value as 0.02, RMSE value as 0.15, and NSE values as 0.87 is the most accurate among the empirical models, as observing Table 3 and Fig. 4.

Results of M5P tree

Developing of M5P model is a trial-and-error method. The M5P model contains only one user-defined parameter (m). During the M5P development, the optimum value of $m=4$ was found. The agreement diagram of M5P model in both periods of progress is shown in Fig. 3. To assess the performance of this model, performance parameters for both periods are calculated and presented in Table 4. Figure 3 shows that the M5P tree model with R value as 0.82, MSE value as 0.03, RMSE value as 0.18, and NSE value as 0.82 is appropriate for predicting the natural recharging rate of groundwater.

Results of GP

Similar to M5P model preparation, developing of GP model is based on the same dataset. In this study, Gaussian noise (0.01) was fixed for the fair assessment of both the kernel function-based models. The primary parameters for GP models are listed in Table 2. Based on the obtained results (Table 4), the PUK kernel gives a better performance than RBF kernel function-based model. To assess the precision of these models, agreement designs are presented in Fig. 4. The R values of PUK kernel function-based GP model were attained 0.97 and 0.88 for preparing and testing, correspondingly. Assessing Table 4 and Fig. 4 concludes that GP_PUK model is more appropriate than

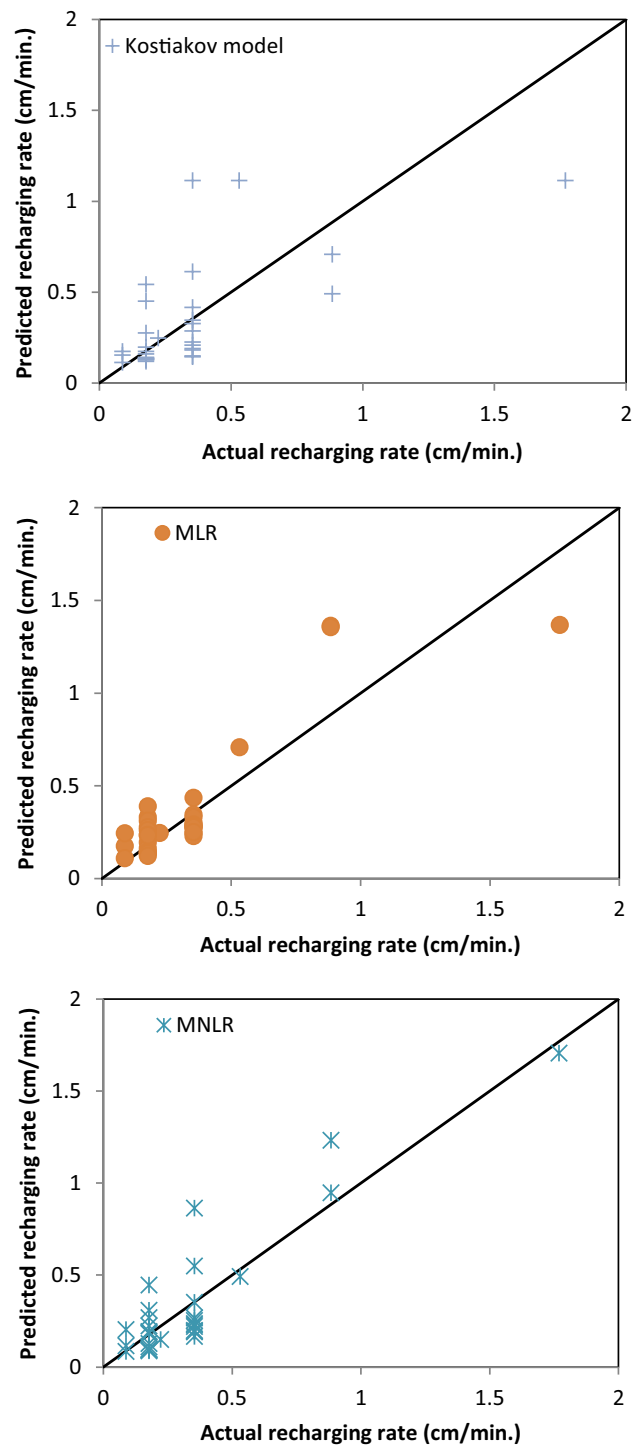


Fig. 2 Performance of empirical models

M5P and GP_RBF models for prediction of the natural recharging rate of the soil. It is remarkable that in these figures the GP_PUK is linked with outcomes of the PUK kernel function-based GP model and GP_RBF is linked to the outcomes of the GP_RBF model.

Table 3 Performance of empirical equations

| Approaches | Training dataset | | | | Testing dataset | | | |
|-----------------|------------------|------|------|------|-----------------|------|------|------|
| | <i>R</i> | MSE | RMSE | NSE | <i>R</i> | MSE | RMSE | NSE |
| MNLR | 0.92 | 0.03 | 0.18 | 0.84 | 0.90 | 0.02 | 0.15 | 0.87 |
| MLR | 0.81 | 0.07 | 0.27 | 0.65 | 0.88 | 0.03 | 0.16 | 0.85 |
| Kostiakov model | 0.73 | 0.10 | 0.32 | 0.52 | 0.66 | 0.06 | 0.25 | 0.65 |

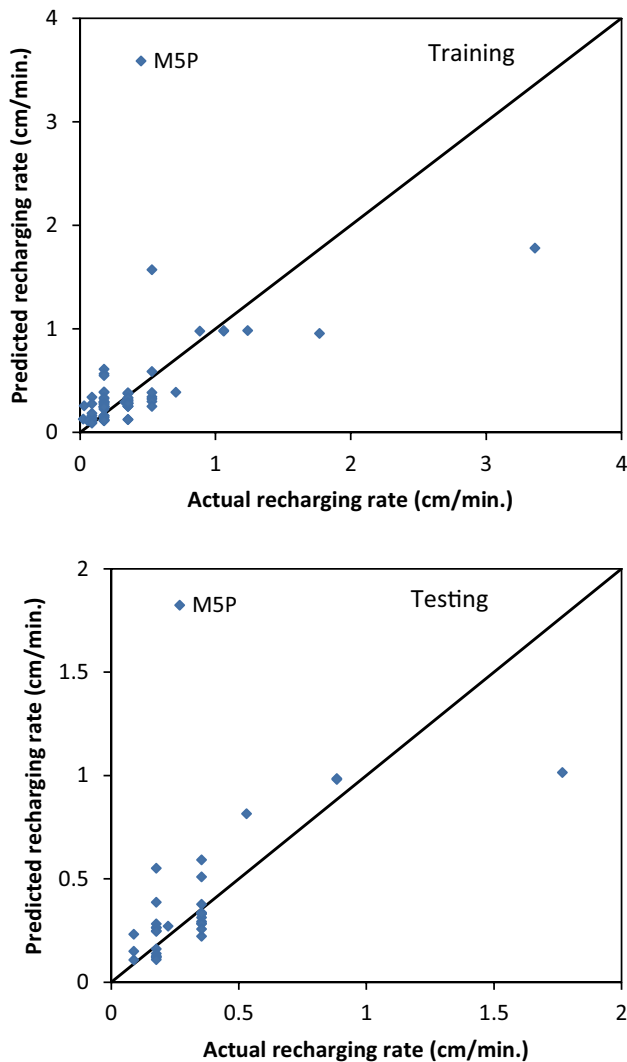


Fig. 3 Performance of M5P tree model

Table 4 Performance of M5P-, GP- and RF-based models

| Approaches | Training dataset | | | | Testing dataset | | | |
|------------|------------------|------|------|------|-----------------|------|------|------|
| | <i>R</i> | MSE | RMSE | NSE | <i>R</i> | MSE | RMSE | NSE |
| M5P | 0.80 | 0.08 | 0.28 | 0.62 | 0.82 | 0.03 | 0.18 | 0.82 |
| GP_PUK | 0.97 | 0.01 | 0.12 | 0.93 | 0.88 | 0.02 | 0.15 | 0.88 |
| GP_RBF | 0.91 | 0.04 | 0.19 | 0.83 | 0.87 | 0.02 | 0.16 | 0.86 |
| RF | 0.98 | 0.02 | 0.12 | 0.93 | 0.91 | 0.02 | 0.13 | 0.91 |

Results of RF

Similarly, the development of the RF model is the same as the M5P and GP model, based on the dataset. The progress of RF includes the number of trees (*k*) and the number of features (*m*). In this study, 1 tree and number of features 1 were selected. Outcomes of the RF model for prediction of the recharging rate of groundwater are presented in Fig. 5. The optimum value of the primary constraint of the RF model is presented in Table 2. Overall, assessing Table 4 and Fig. 5 it is clear that the exactness of the RF model for the prediction of the natural recharging rate of the soil is supreme. The *R* values of the RF model were obtained 0.98 and 0.91 for training and testing, respectively.

Assessment of soft computing and empirical models (Tables 3, 4) states that RF-based model shows better response than other models. Also, the MNLR model shows the better response in the performance of estimating the natural recharging rate of groundwater, than GP, M5P and the empirical models. Finally, the Kostiakov model has the least ability to estimate the natural recharging rate.

Inter-comparison of soft computing and empirical models

Last few years soft computing methods are successfully used in several engineering-related fields. In this, study performance of M5P-, GP- and RF-based models were assessed for the prediction of the recharging rate of the soil. The developed soft computing-based models were compared with Kostiakov model, MLR and MNLR. The performances of all discussed models are listed in Tables 3 and 4 for both training and testing stages. Agreement plot among actual and predicted values with applied models using the testing

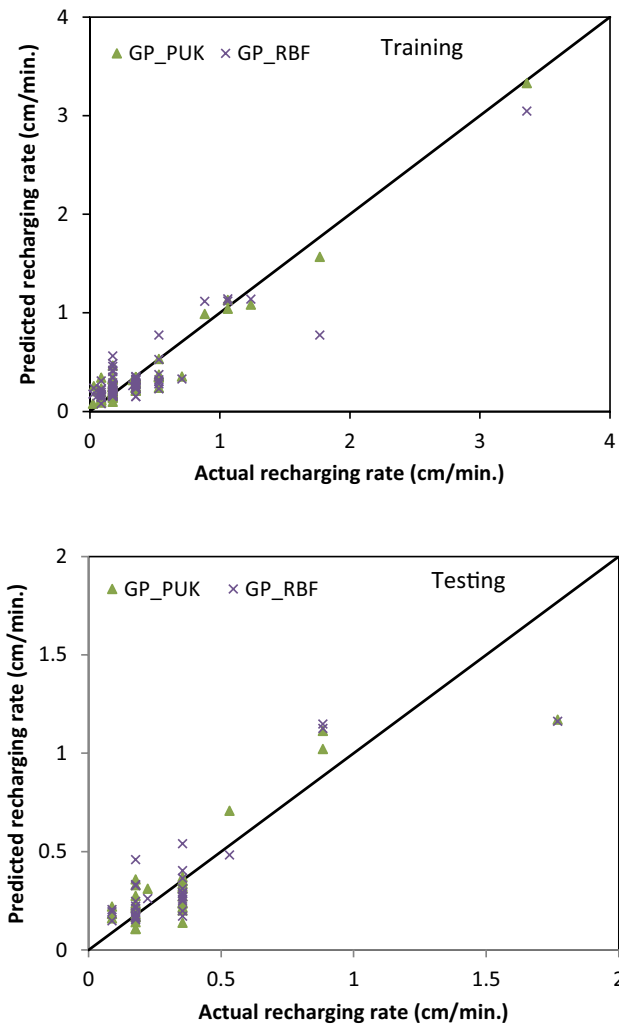


Fig. 4 Performance of GP models

stage is drawn in Fig. 6. Figure 6 and Tables 3 and 4 confirm that the RF model is outperforming than other applied soft computing and empirical models. Box plot (Fig. 7) was plotted, in which overall error distribution was shown. As a result, the negative and positive error values correspond to the over-estimation and under-estimation behavior of the models, respectively. Figure 8 also shows Taylor’s diagram for all applied models. Taylor diagram was used to illustrate schematically the performance of the applied models (Taylor 2001). Three statistic parameters including standard deviation, correlation and root mean square error evaluated the degree of compliance of recharging rate of water through soil among actual and predicted values. Figure 8 suggests that RF model achieves higher correlation with minimum standard deviation values. Taylor diagram also confirms that the RF model is performing better than other applied models.

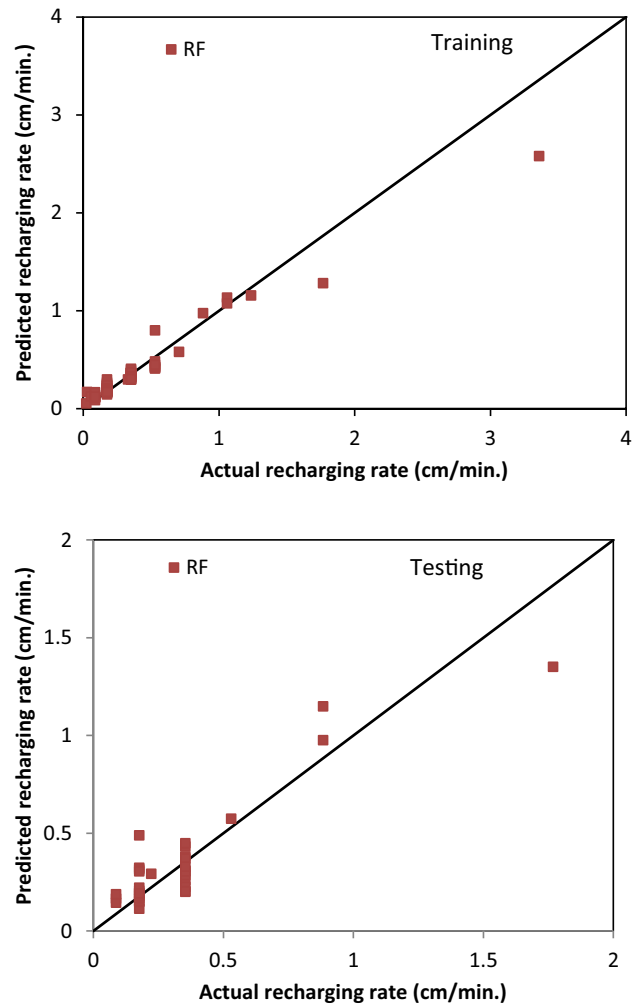


Fig. 5 Performance of RF model

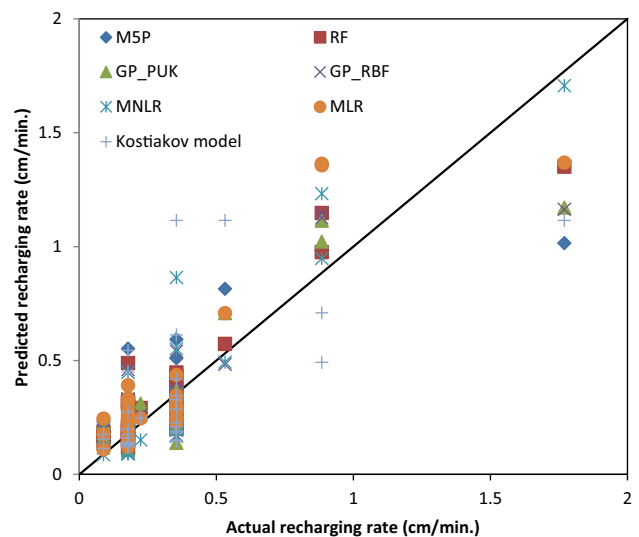


Fig. 6 Agreement plot among actual vs predicted values of recharging rate with various soft computing techniques using testing data set

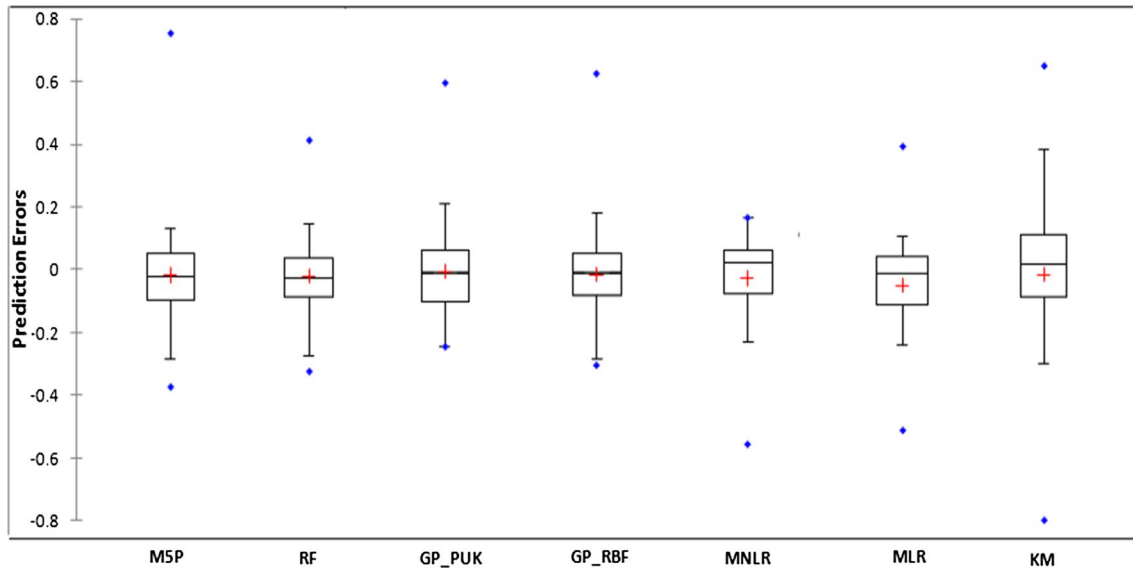


Fig. 7 Box plot for error prediction with various soft computing and empirical model

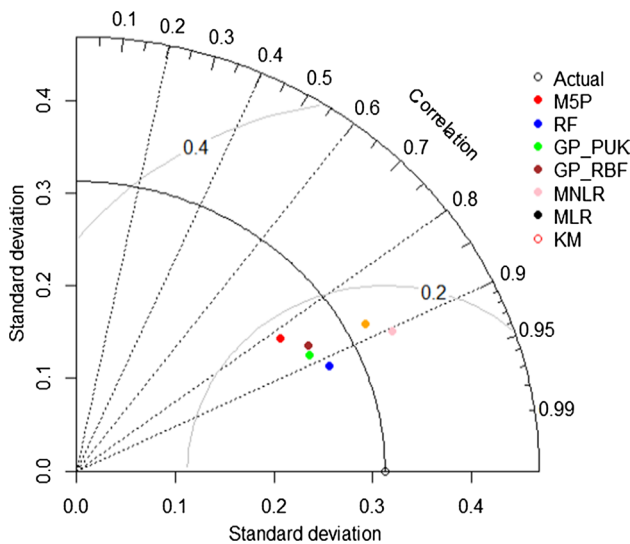


Fig. 8 Taylor diagram for various soft computing and empirical model

Sensitivity investigation using RF

Sensitivity investigation was carried out on the RF model in order to examine the performance of the developed best model in the deficiency of every input. Numerous sets of training data were prepared by removing one input parameter at a time and outcomes were recorded in terms of R and RMSE with the testing dataset. Outcomes of sensitivity investigation on RF are given in Table 5. Table 5 shows that, in comparison with other input parameters, the time has an important role in predicting the recharging rate of the soil.

Table 5 Sensitivity investigation using RF

| Input combination of parameters | Output parameter | Eliminated parameter | RF | |
|---------------------------------|------------------|----------------------|------|---------------|
| | | | R | RMSE (cm/min) |
| <i>t, S, Si, C, D, Mc</i> | <i>R(t)</i> | | 0.91 | 0.13 |
| <i>S, Si, C, D, Mc</i> | <i>R(t)</i> | <i>t</i> | 0.88 | 0.15 |
| <i>t, Si, C, D, Mc</i> | <i>R(t)</i> | <i>S</i> | 0.91 | 0.13 |
| <i>t, S, C, D, Mc</i> | <i>R(t)</i> | <i>Si</i> | 0.90 | 0.14 |
| <i>t, S, Si, D, Mc</i> | <i>R(t)</i> | <i>C</i> | 0.90 | 0.14 |
| <i>t, S, Si, C, Mc</i> | <i>R(t)</i> | <i>D</i> | 0.90 | 0.14 |
| <i>t, S, Si, C, D</i> | <i>R(t)</i> | <i>Mc</i> | 0.93 | 0.12 |

Conclusions

Prediction of the natural recharging rate of the groundwater is essential for efficient use of groundwater resource in agriculture (irrigation) and water supply. In this study, experimental data were used in order to investigate the performance of GP-, M5P- and RF-based regression method and evaluate the potential of these techniques in the prediction of natural recharging rate, while a comparison has been made between the empirical (Kostiakov model, multilinear regression (MLR) and multi-nonlinear regression (MNLR)) equations. Outcomes of this study indicate that the performance of RF-based model has shown a superiority between the other soft computing and empirical models. In particular, based on the attained outcomes, the RF model has an appropriate potential to predict the exact recharging rate of the groundwater with R values as 0.98 and 0.91 for

training and testing stages, respectively, while the MNLr (empirical model) offers better performance than the GP, M5P, MLR and Kostiakov model. Also, the PUK-based GP model is more responsive than the RBF-based GP model, for this data set. In addition, an important conclusion obtained from this study is that sensitivity investigation proposes that the variable of time (*t*) is the most significant when RF-based modeling method is selected for the prediction of recharging rate of the groundwater, as time (*t*) affects strongly the recharging rate. Taylor diagram and Box plot results also confirms that the RF model is performing better than other applied models for the prediction recharging rate of the groundwater.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Data set

| Sr. no. | Input parameters | | | | | | Output |
|---------|------------------|------|------|------|---------|--------------------------|--------------------|
| | TIME | Sand | Silt | Clay | Density | Mois- ture content | Recharging rate |
| 1 | 2 | 93 | 3 | 4 | 1.56 | 0.13 | 3.36 |
| 2 | 4 | 93 | 3 | 4 | 1.56 | 0.29 | 1.77 |
| 3 | 6 | 93 | 3 | 4 | 1.56 | 0.29 | 1.06 |
| 4 | 8 | 93 | 3 | 4 | 1.56 | 0.29 | 0.88 |
| 5 | 10 | 93 | 3 | 4 | 1.56 | 0.29 | 1.24 |
| 6 | 12 | 93 | 3 | 4 | 1.56 | 0.29 | 1.06 |
| 7 | 14 | 93 | 3 | 4 | 1.56 | 0.29 | 0.88 |
| 8 | 16 | 93 | 3 | 4 | 1.56 | 0.29 | 0.88 |

| Sr. no. | Input parameters | | | | | | Output |
|---------|------------------|------|------|------|---------|--------------------------|--------------------|
| | TIME | Sand | Silt | Clay | Density | Mois- ture content | Recharging rate |
| 9 | 2 | 76 | 13 | 11 | 1.48 | 0.14 | 0.53 |
| 10 | 4 | 76 | 13 | 11 | 1.48 | 0.29 | 0.53 |
| 11 | 6 | 76 | 13 | 11 | 1.48 | 0.34 | 0.53 |
| 12 | 8 | 76 | 13 | 11 | 1.48 | 0.35 | 0.18 |
| 13 | 10 | 76 | 13 | 11 | 1.48 | 0.36 | 0.18 |
| 14 | 12 | 76 | 13 | 11 | 1.48 | 0.36 | 0.18 |
| 15 | 14 | 76 | 13 | 11 | 1.48 | 0.36 | 0.35 |
| 16 | 16 | 76 | 13 | 11 | 1.48 | 0.36 | 0.35 |
| 17 | 18 | 76 | 13 | 11 | 1.48 | 0.36 | 0.35 |
| 18 | 20 | 76 | 13 | 11 | 1.48 | 0.37 | 0.53 |
| 19 | 22 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 20 | 24 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 21 | 26 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 22 | 28 | 76 | 13 | 11 | 1.48 | 0.37 | 0.53 |
| 23 | 30 | 76 | 13 | 11 | 1.48 | 0.37 | 0.18 |
| 24 | 32 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 25 | 34 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 26 | 36 | 76 | 13 | 11 | 1.48 | 0.37 | 0.18 |
| 27 | 38 | 76 | 13 | 11 | 1.48 | 0.37 | 0.33 |
| 28 | 40 | 76 | 13 | 11 | 1.48 | 0.37 | 0.22 |
| 29 | 50 | 76 | 13 | 11 | 1.48 | 0.37 | 0.03 |
| 30 | 52 | 76 | 13 | 11 | 1.48 | 0.37 | 0.35 |
| 31 | 54 | 76 | 13 | 11 | 1.48 | 0.37 | 0.53 |
| 32 | 56 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 33 | 58 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 34 | 62 | 76 | 13 | 11 | 1.48 | 0.37 | 0.09 |
| 35 | 64 | 76 | 13 | 11 | 1.48 | 0.37 | 0.18 |
| 36 | 68 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 37 | 70 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 38 | 72 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 39 | 76 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 40 | 78 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 41 | 92 | 76 | 13 | 11 | 1.48 | 0.38 | 0.03 |
| 42 | 96 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 43 | 98 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 44 | 100 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 45 | 104 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 46 | 106 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 47 | 108 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 48 | 110 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 49 | 112 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 50 | 118 | 76 | 13 | 11 | 1.48 | 0.38 | 0.06 |
| 51 | 120 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 52 | 122 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 53 | 124 | 76 | 13 | 11 | 1.48 | 0.38 | 0.35 |
| 54 | 126 | 76 | 13 | 11 | 1.48 | 0.38 | 0.35 |
| 55 | 128 | 76 | 13 | 11 | 1.48 | 0.38 | 0.18 |
| 56 | 132 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |

| Sr. no. | Input parameters | | | | | | Output |
|---------|------------------|------|------|------|---------|--------------------------|--------------------|
| | TIME | Sand | Silt | Clay | Density | Mois- ture content | Recharging rate |
| 57 | 136 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 58 | 140 | 76 | 13 | 11 | 1.48 | 0.38 | 0.09 |
| 59 | 2 | 82 | 8 | 10 | 1.54 | 0.27 | 1.77 |
| 60 | 4 | 82 | 8 | 10 | 1.54 | 0.31 | 0.35 |
| 61 | 6 | 82 | 8 | 10 | 1.54 | 0.30 | 0.18 |
| 62 | 8 | 82 | 8 | 10 | 1.54 | 0.31 | 0.53 |
| 63 | 10 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 64 | 12 | 82 | 8 | 10 | 1.54 | 0.31 | 0.18 |
| 65 | 14 | 82 | 8 | 10 | 1.54 | 0.31 | 0.18 |
| 66 | 16 | 82 | 8 | 10 | 1.54 | 0.31 | 0.18 |
| 67 | 18 | 82 | 8 | 10 | 1.54 | 0.33 | 0.35 |
| 68 | 20 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 69 | 22 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 70 | 26 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 71 | 28 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 72 | 30 | 82 | 8 | 10 | 1.54 | 0.32 | 0.71 |
| 73 | 32 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 74 | 34 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 75 | 36 | 82 | 8 | 10 | 1.54 | 0.32 | 0.53 |
| 76 | 40 | 82 | 8 | 10 | 1.54 | 0.32 | 0.09 |
| 77 | 44 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 78 | 46 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 79 | 48 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 80 | 50 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 81 | 52 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 82 | 54 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 83 | 56 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 84 | 58 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 85 | 60 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 86 | 62 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 87 | 64 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 88 | 66 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 89 | 68 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 90 | 72 | 82 | 8 | 10 | 1.54 | 0.32 | 0.09 |
| 91 | 74 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 92 | 78 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 93 | 82 | 82 | 8 | 10 | 1.54 | 0.32 | 0.09 |
| 94 | 84 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 95 | 86 | 82 | 8 | 10 | 1.54 | 0.33 | 0.35 |
| 96 | 88 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 97 | 90 | 82 | 8 | 10 | 1.54 | 0.32 | 0.35 |
| 98 | 92 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 99 | 94 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 100 | 96 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 101 | 98 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 102 | 100 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 103 | 102 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 104 | 104 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |

| Sr. no. | Input parameters | | | | | | Output |
|---------|------------------|------|------|------|---------|--------------------------|--------------------|
| | TIME | Sand | Silt | Clay | Density | Mois- ture content | Recharging rate |
| 105 | 106 | 82 | 8 | 10 | 1.54 | 0.32 | 0.18 |
| 106 | 110 | 82 | 8 | 10 | 1.54 | 0.33 | 0.09 |

References

- Alley WM, Healy RW, LaBaugh JW, Reilly TE (2002) Flow and storage in groundwater systems. *Science* 296(5575):1985–1990
- Angelaki A, Sakellariou-Makrantonaki M, Tzimopoulos C (2013) Theoretical and experimental research of cumulative infiltration. *Transp Porous Media* 100(2):247–257
- Ebtehaj I, Bonakdari H (2013) Evaluation of sediment transport in sewer using artificial neural network. *Eng Appl Comput Fluid Mech* 7(3):382–392
- Freeze RA (1969) The mechanism of natural ground-water recharge and discharge: 1 One-dimensional, vertical, unsteady, unsaturated flow above a recharging or discharging ground-water flow system. *Water Resour Res* 5(1):153–171
- Green WH, Ampt GA (1911) Studies on soil physics. *J Agric Sci* 4(1):1–24
- Holtan HN (1961) A concept for infiltration estimates in watershed engineering. *Agricultural research service*, vol 41–51. USDA, Washington, DC
- Horton RE (1941) An approach toward a physical interpretation of infiltration-capacity I. *Soil Sci Soc Am J* 5(1):399–417
- Kisi O, Shiri J, Nikoofar B (2012) Forecasting daily lake levels using artificial intelligence approaches. *Comput Geosci* 41:169–180
- Kostiakov AN (1932) On the dynamics of the coefficient of water percolation in soils and the necessity of studying it from the dynamic point of view for the purposes of amelioration. *Trans Sixth Commun Int Soc Soil Sci* 1:7–21
- Kumar M, Sihag P (2019) Assessment of infiltration rate of soil using empirical and machine learning-based models. *Irrig Drain* 68(3):588–601
- Parsaie A, Haghiabi AH (2017a) Mathematical expression of discharge capacity of compound open channels using MARS technique. *J Earth Syst Sci* 126(2):20
- Parsaie A, Haghiabi AH (2017b) Numerical routing of tracer concentrations in rivers with stagnant zones. *Water Sci Technol Water Supply* 17(3):825–834
- Parsaie A, Haghiabi AH (2017c) Computational modeling of pollution transmission in rivers. *Appl Water Sci* 7(3):1213–1222
- Parsaie A, Najafian S, Shamsi Z (2016) Predictive modeling of discharge of flow in compound open channel using radial basis neural network. *Model Earth Syst Environ* 2(3):150
- Parsaie A, Ememgholizadeh S, Haghiabi AH, Moradinejad A (2018a) Investigation of trap efficiency of retention dams. *Water Sci Technol Water Supply* 18(2):450–459
- Parsaie A, Haghiabi AH, Saneie M, Torabi H (2018b) Prediction of energy dissipation of flow over stepped spillways using data-driven models. *Iran J Sci Technol Trans Civil Eng* 42(1):39–53
- Parsaie A, Azamathulla HM, Haghiabi AH (2020) Physical and numerical modeling of performance of detention dams. *J Hydrol* 581:121757
- Philip JR (1957) The theory of infiltration: 1. The infiltration equation and its solution. *Soil Sci* 83(5):345–358

- Qishlaqi A, Kordian S, Parsaie A (2017) Hydrochemical evaluation of river water quality—a case study. *Appl Water Sci* 7(5):2337–2342
- Quinlan JR (1992) Learning with continuous classes. In: 5th Australian joint conference on artificial intelligence, vol 92, pp 343–348
- Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning, vol 1. MIT Press, Cambridge, p 248
- Richards LA (1931) Capillary conduction of liquids through porous mediums. *Physics* 1(5):318–333
- Rockström J, Karlberg L, Wani SP, Barran J, Hatibu N, Oweis T, Bruggeman A, Farahani J, Qiang Z (2010) Managing water in rainfed agriculture—the need for a paradigm shift. *Agric Water Manag* 97(4):543–550
- Sihag P (2018) Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Model Earth Syst Environ* 4(1):189–198
- Sihag P, Tiwari NK, Ranjan S (2017) Estimation and inter-comparison of infiltration models. *Water Sci* 31(1):34–43
- Sihag P, Jain P, Kumar M (2018) Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. *Model Earth Syst Environ* 4(1):61–68
- Sihag P, Singh VP, Angelaki A, Kumar V, Sepahvand A, Golia E (2019) Modelling of infiltration using artificial intelligence techniques in semi-arid Iran. *Hydrol Sci J* 64(13):1647–1658
- Sihag P, Kumar M, Singh B (2020) Assessment of infiltration models developed using soft computing techniques. *Geol Ecol Landsc*. <https://doi.org/10.1080/24749508.2020.1720475>
- Singh VP, Yu FX (1990) Derivation of infiltration equation using systems approach. *J Irrig Drain Eng* 116(6):837–858
- Singh B, Sihag P, Singh K (2017) Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Model Earth Syst Environ* 3(3):999–1004
- Singh B, Sihag P, Singh K (2018) Comparison of infiltration models in NIT Kurukshetra campus. *Appl Water Sci* 8(2):63
- Sophocleous M (2002) Interactions between groundwater and surface water: the state of the science. *Hydrogeol J* 10(1):52–67
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106(D7):7183–7192

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.