ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

# Modeling the infiltration process with soft computing techniques

[1]**RACHITA RAUL,** Gandhi Institute of Excellent Technocrats, Bhubaneswar, India
[2]**PRIYANKA MOHANTY,** Indus College of Engineering, Bhubaneswar, Odisha, India

**ABSTRACT**

Planning and developing irrigation networks benefit greatly from understanding the infiltration process. In this study, the soil infiltration rate was calculated using the Artificial Neural Network (ANN) technique. The performance of ANN was compared to that of other artificial intelligence techniques as the generalised neural network, the gene expression programming, and the gaussian process (GP) (GRNN). The ANN model surpasses the GP, GRNN, and GEP models, which all provide good estimate performance. A data set of 155 field observations was examined for this study. Out of 155 data, 105 were chosen to create various algorithms, while the remaining 50 were chosen to test the models. One hidden layer was used to create the best ANN model.

**KEYWORDS**

Infiltration rate; Gaussian process regression; gene expression programming; generalized neural network; artificial neural network

## 1. Introduction

Infiltration is defined as a physical phenomenon, in which water penetrates into the soil from surface sources such as snowfall, precipitation, and irrigation. Information of infiltration is essential in hydrologic design, watershed management, irrigation, and agriculture. It is, therefore, necessary to have a thorough understanding of infiltration characteristics for a given land use complex. Infiltration is an important component process of the hydrologic cycle. It is one of the main abstractions accounted for in the rainfall–runoff modeling. In the hydrological process, infiltration separates the water into two parts: surface flow and groundwater flow. Soils of different types have different infiltration characteristics. Infiltration rates are affected by a number of factors, some of which are antecedent soil moisture, texture of the soil, density, and behavior of the soil (Angelaki et al. 2013). Knowledge of infiltration is essential for any beneficial durable study of hydrological evaluations (Pedretti et al. 2012; Shiri et al. 2017a; Shiri et al. 2017b). Because of an elementary role in the hydrological process, it has received a great deal of attention from soil and water researchers. Several models [Philip, Kostiakov, Horton, Modified Kostiakov, Holton, Novel, etc.] have evolved to assess infiltration.

Many researchers developed various conventional models for estimating infiltration rate (Kostiakov 1932; Mishra et al. 2003; Philips 1957; Richards 1931; Sihag et al. 2017a; Singh and Yu 1990). Some researchers used soft-computing in estimating infiltration process (Sihag et al. 2017b; ; Sihag et al. 2018a; Singh et al. 2017; Sy 2006; Tiwari et al. 2017).

In the last decades, Gaussian Process regression, gene expression programming, generalized neural network, Support vector machine, Fuzzy Logic, Adaptive neuro fuzzy inference system, and Artificial neural network have been used as dominant tools in solving water resources problems (Azamathulla et al. 2016a; Azamathulla et al. 2016b; Baba et al. 2013; Haghiabi et al., 2017a; Haghiabi et al. 2018; Karimi et al. 2016; Keshavarzi et al. 2017; Kisi et al. 2015; Kisi et al. 2017; Kumar et al. 2018a; Mehdipour and Memarianfard 2017; Najafzadeh et al. 2018a; Najafzadeh et al. 2018b; Parsaie and Haghiabi 2014; Parsaie and Haghiabi 2015a; Parsaie and Haghiabi 2017a; Parsaie et al. 2017a; Parsaie et al. 2017b; Parsaie et al. 2017c; Roushangar et al. 2014; Rosushangar et al. 2017; Shiri and Kisi 2012; Shiri et al. 2016; Shiri et al. 2017c; Sihag et al. 2017c; Sihag et al. 2018b; Tiwari et al. 2018; Yavari et al. 2017). Conventional models are site specific and require model parameters, whereas soft computing-based models are general for the study area. The advantage of using any soft computing-based model is that these techniques require few user-defined parameters. Keeping in view the improved performance by GP-, GEP-, GRNN-, and ANN-based approaches in water engineering problems, this study compares its performances with conventional models (Kostiakov model and Philips model) of infiltration rate of the soil.

## 2. Soft computing techniques and conventional models

### Overview of Gaussian process regression (GP)

Rasmussen and Williams (2006) assumed for the working of GP regression model that the adjoining observations give knowledge to each other. It is a process to specify a prior immediately over function space. The mean and covariance of Gaussian distribution are vector and matrix. The Gaussian process is over function. GP regression model is capable of recognizing the predictive distribution analogous to check input information.

A GP is a collection of random variables in which any finite number has a joint multivariate Gaussian distribution. Assuming $xXy$ represents the domains of inputs and outputs, respectively, in which $n$ pairs $(x_i, y_i)$ are drawn independently and identically distributed, for regression, let $y \subseteq \Re$; then, a GP on $\chi$ is defined by a mean function $\mu : 3\ \Re$ and a covariance function k: $3 \times 3 \rightarrow \Re$.

There are many kernel functions in GP, so how to select a better kernel function is also a research concern. However, for general purpose, there are two common kernel functions.

1. Radial basis kernel (RBF) = $e^{-|x_i - x_j|^2}$
2. Pearson VII function kernel (PUK) = $\frac{1}{\left[1+\left(\frac{2\sqrt{\|x_i - x_j\|^2}\sqrt{2^{(1/\omega)}-1}}{\sigma}\right)^2\right]^\omega}$

Here, Gaussian noise, $\gamma$, $\sigma$, and $\omega$ are kernel parameters. It is well known that GP generalization performance (prediction precision) depends on a good setting of meta-parameters, parameters Gaussian noise, $\gamma$, $\sigma$, and ω. The choices of Gaussian noise, $\gamma$, $\sigma$, and ω control the prediction (regression) model's complexity. In this study, a physical method (carrying out several trials using different combinations of user-defined parameters) was implemented to select user-defined parameters (i.e. Gaussian noise, $\gamma$, $\sigma$, and $\omega$). For further explanation about GP in detail, readers are referred to Kuss (2006).

### *Overview of gene-expression programming*

GEP proposed by Ferreira (2002) is a search technique that involves computer programs. It is a developed method with the base of genetic algorithms (GA) and has been widely implemented in the current studies. The computer programs of GEP are all encoded in linear chromosomes, which are then articulated or translated into expression trees (ETs). A concise flowchart of GEP is shown in Figure 1. The first step of this program to solve any problem is to produce the initial population, which happens with arbitrary births of chromosomes and in the later, the chromosomes convert to expression trees (ETs) that are examined by performance criteria to represent the solubility

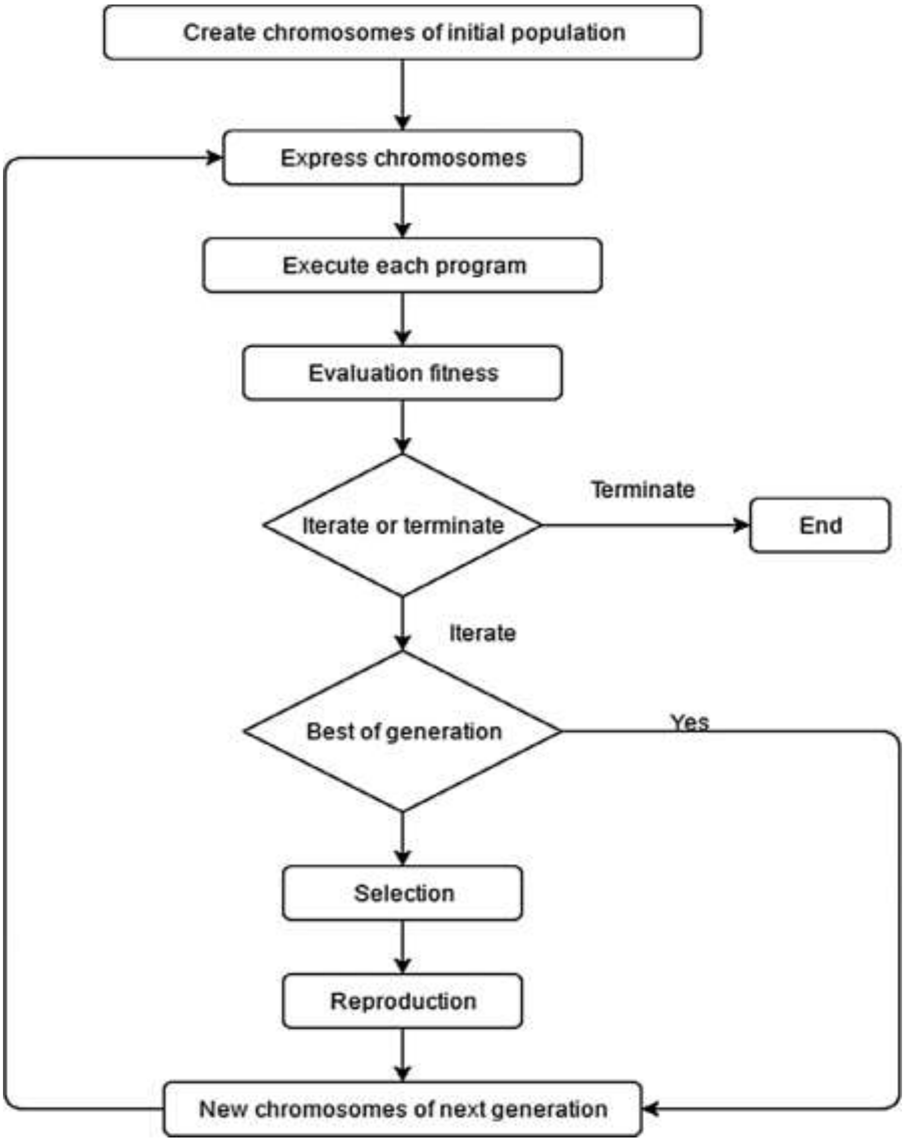



**Figure 1.** Brief algorithm of gene expression programming.

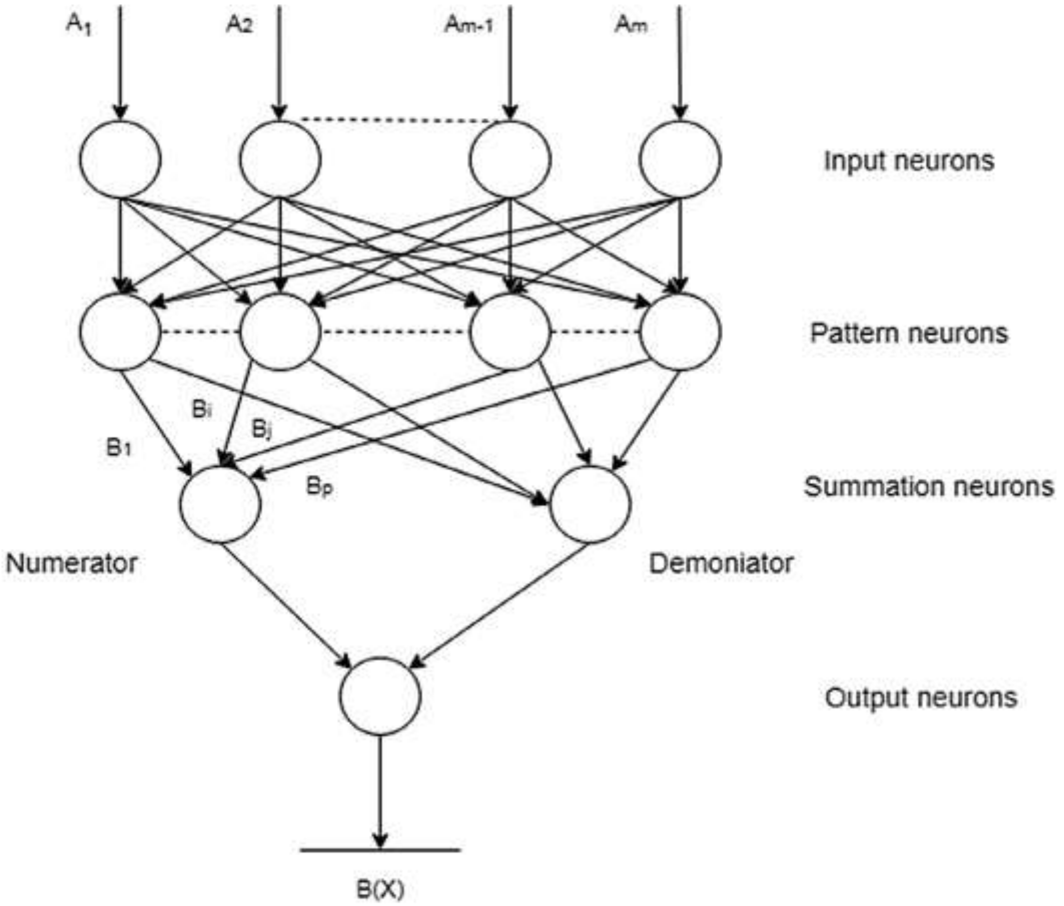



**Figure 2.** Structure of the GRNN model.

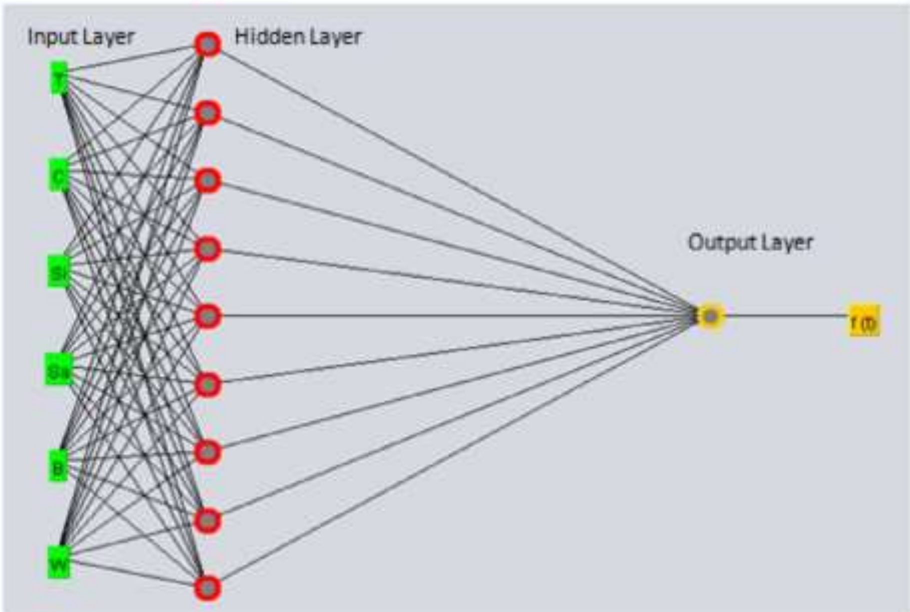



**Figure 3.** Structure of the ANN model.

of produced ETs. If the outcomes convince the performance criteria, population generating stops, and if the results are not satisfactory, system regenerates with some improvement to make new generation with improved value and this process occurs until best results are achieved. For further explanation about GEP, readers are referred to Ferreira (2006), Mehdipour et al. (2017), Parsaie and Haghiabi (2017b), and Shiri (2017).

### *Overview of generalized regression neural network*

GRNN, first introduced by Specht (1991), is a normalized RBF network in which there is a hidden component centered at each training example. These RBF components are known as 'kernels' and are usually probability density functions, for example, support vector machine and Gaussian Process. The

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

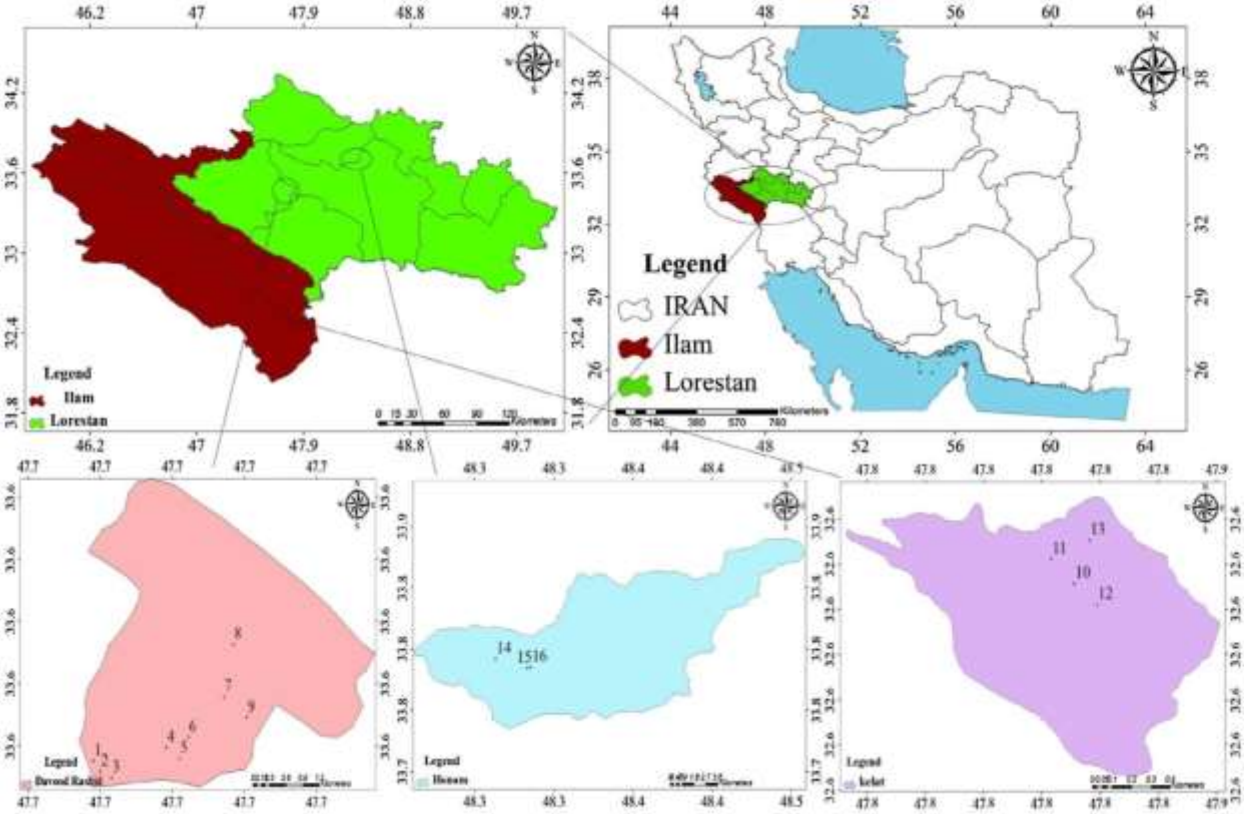



**Figure 4.** Location of the study area. Source: Author.

**Table 1.** The soil texture and coordinate system of all locations.

| Sr. No. | Site No. | Location | Latitude | Longitude |
|---|---|---|---|---|
| 1 | 1 | Kelat | 32°38′33.64″N | 47°50′45.86″E |
| 2 | 2 | Kelat | 32°38′37.16″N | 47°50′42.02″E |
| 3 | 3 | Kelat | 32°38′30.64″N | 47°50′49.64″E |
| 4 | 4 | Kelat | 32°38′39.88″N | 47°50′48.39″E |
| 5 | 5 | Davood Rashid | 33°33′54.13″N | 47°41′33.06″E |
| 6 | 6 | Davood Rashid | 33°33′49.14″N | 47°41′40.52″E |
| 7 | 7 | Davood Rashid | 33°33′58.93″N | 47°41′45.94″E |
| 8 | 8 | Davood Rashid | 33°34′17.67″N | 47°42′6.56″E |
| 9 | 9 | Davood Rashid | 33°34′8.07″N | 47°42′20.06″E |
| 10 | 10 | Davood Rashid | 33°34′41.89″N | 47°42′12.19″E |
| 11 | 11 | Davood Rashid | 33°33′48.08″N | 47°40′50.51″E |
| 12 | 12 | Davood Rashid | 33°33′43.24″N | 47°40′54.40″E |
| 13 | 13 | Davood Rashid | 33°33′40.29″N | 47°41′1.05″E |
| 14 | 14 | Honam | 33°47′25.37″N | 48°15′55.80″E |
| 15 | 15 | Honam | 33°47′4.12″N | 48°17′15.25″E |
| 16 | 16 | Honam | 33°47′6.80″N | 48°17′23.76″E |

**Figure 5.** Double ring infiltrometer.

hidden-to-output weights are just the target values, so the output is basically a weighted average of the target values of training bags near the given input bags. The only weights that need to be studied are the widths of the RBF components. A GRNN arrangement contains only four levels. The input elements are in the initial level, the second level has the pattern elements, the outputs of this level are crossed on to the summation elements in the third level, and the final level covers the output elements. The first level is completely linked to the second, pattern level, where each element shows a training pattern and its output is a measure of the distance of the input from the stored patterns. The structure of the GRNN is shown in Figure 2. The optimal value of the user-defined parameter is determined experimentally, which is known as a spread (*s*). For more information about GRNN, readers are referred to Specht (1991) and Wasserman (1993).

### *Overview of artificial neural network*

ANN is a machine learning technique widely implemented for the numerical forecast of problems (Haghiabi et al. 2017b; Kumar et al. 2018b; Parsaie and Haghiabi 2015b; Sihag 2018). It is inspired by the functioning of the neurons system and brain architecture. ANN has one input, one or more hidden, and one output layers. Each layer consists of the number of nodes and the weighted connection among these layers represents the connection among the nodes. Input layer having nodes equal to the number of input parameters distributes the data accessible to the network and does not help in processing. The target layer is the final processing unit (Figure 3). When an input layer is subjected to an input value which passes through the interconnections between the nodes, these values are multiplied by the corresponding weights and summed up to obtain the net output ($P_j$) to the unit.

$$P_j = \sum_{i} x_{ij} \times y_i \qquad (1)$$

where $X_{ij}$ is the weight of interconnection from unit $i$ to $j$, $y_i$ is the input value at the input layer, $P_j$ is target obtained by activation function to produce a target for unit $j$. The full discussion about ANN is given by Haykin (1999).

## Conventional models

### Kostiakov model

Kostiakov (1932) proposed an empirical model in order to estimate the infiltration rate which is mentioned as follows:

$$f(t) = mt^{-n} \qquad (2)$$

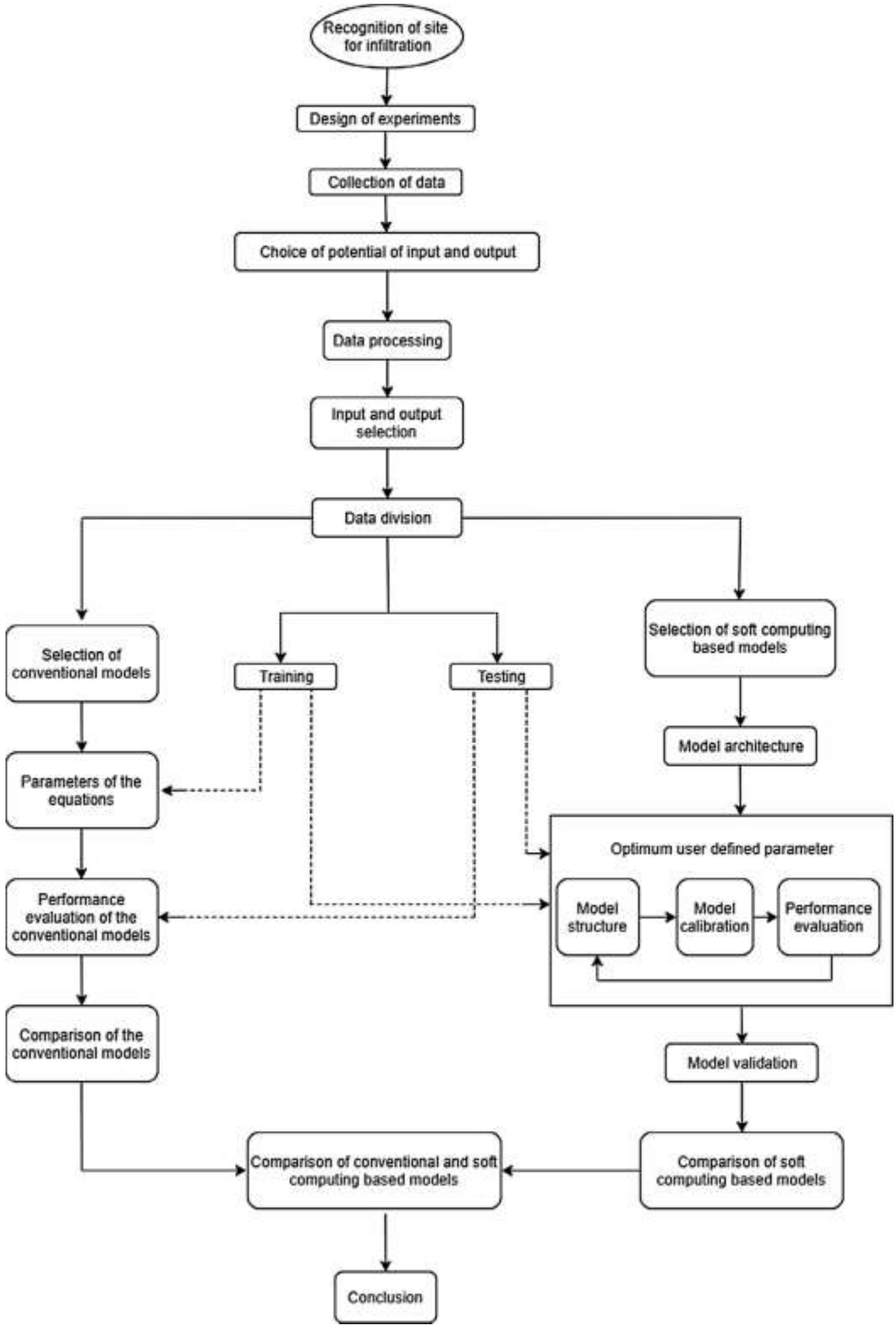



**Figure 6.** Flow chart of research.

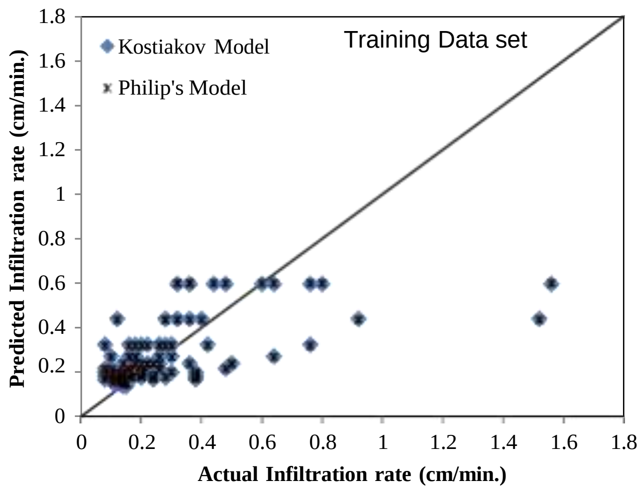
ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

**Table 2.** Details of the initial and final infiltration rates and soil properties of the study area.

| Site No. | Initial infiltration rate at t = 2.5 min. (cm/min.) | Final/steady infiltration rate (cm/min.) | Clay (%) | Silt (%) | Sand (%) | Bulk density (g/cc) | Moisture content (%) |
|---|---|---|---|---|---|---|---|
| 1 | 0.32 | 0.15 | 10 | 50 | 38 | 1.423 | 2.73 |
| 2 | 0.32 | 0.11 | 12 | 58 | 30 | 1.42 | 3.08 |
| 3 | 0.36 | 0.13 | 12 | 50 | 38 | 1.79 | 2.58 |
| 4 | 0.36 | 0.14 | 14 | 50 | 36 | 1.63 | 2.49 |
| 5 | 0.64 | 0.28 | 26 | 44 | 30 | 1.44 | 2.37 |
| 6 | 0.56 | 0.20 | 25 | 49 | 26 | 1.448 | 2.37 |
| 7 | 0.60 | 0.20 | 16 | 51 | 33 | 1.4 | 2.48 |
| 8 | 0.12 | 0.08 | 18 | 62 | 20 | 1.08 | 3.84 |
| 9 | 0.8 | 0.24 | 28 | 55 | 17 | 1.3 | 2.24 |
| 10 | 0.44 | 0.1 | 16 | 65 | 19 | 1.27 | 1.66 |
| 11 | 0.44 | 0.14 | 18 | 53 | 29 | 1.4 | 2.18 |
| 12 | 0.76 | 0.14 | 20 | 51 | 29 | 1.24 | 1.71 |
| 13 | 0.48 | 0.16 | 24 | 49 | 27 | 1.32 | 1.95 |
| 14 | 0.36 | 0.12 | 52 | 37 | 11 | 1.56 | 2.42 |
| 15 | 1.48 | 0.38 | 50 | 44 | 6 | 1.46 | 2.3 |
| 16 | 1.56 | 0.38 | 42 | 37 | 21 | 1.48 | 2.3 |

**Table 3.** Characteristics of the training and testing data used in this study.

| Input parameter | Data | Min· | Max· | Mean· | St. dev. |
|---|---|---|---|---|---|
| T (min.) | Train | 2.5 | 70 | 24.14286 | 16.84141 |
| | Test | 2.5 | 70 | 22.9 | 16.67792 |
| C (%) | Train | 10 | 52 | 23.61905 | 13.04901 |
| | Test | 10 | 52 | 24.92 | 12.90758 |
| Si (%) | Train | 37 | 65 | 50.57143 | 7.623165 |
| | Test | 37 | 65 | 49.84 | 7.399283 |
| Sa (%) | Train | 6 | 38 | 25.65714 | 9.247483 |
| | Test | 6 | 38 | 25.16 | 9.051688 |
| B (g/cc) | Train | 1.08 | 1.79 | 1.417771 | 0.166352 |
| | Test | 1.08 | 1.79 | 1.41888 | 0.150154 |
| W (%) | Train | 1.66 | 3.84 | 2.434952 | 0.51357 |
| | Test | 1.66 | 3.84 | 2.4128 | 0.491578 |



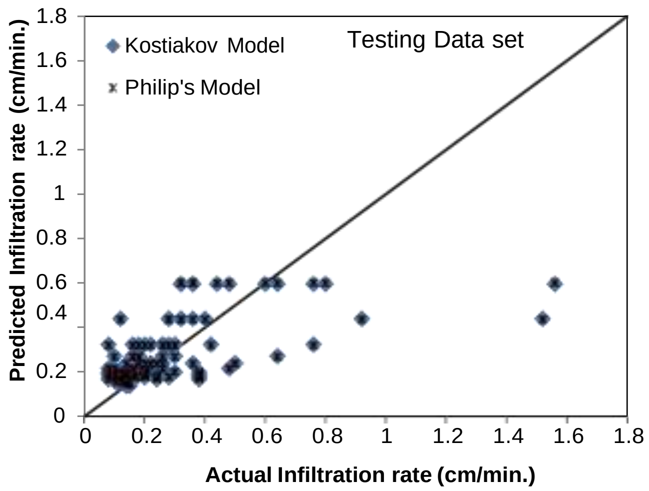




**Figure 7.** Performance of conventional models.

**Table 4.** Performance of conventional models with training and testing data-set.

| Sr. No. | Models | Performance evaluation parameters | | | | | |
|---|---|---|---|---|---|---|---|
| | | Training | | | Testing | | |
| | | CC | RMSE | NSE | CC | RMSE | NSE |
| 1 | Kostiakov model | 0.5689 | 0.2786 | 0.3237 | 0.5017 | 0.2687 | 0.2463 |
| 2 | Philips model | 0.5689 | 0.1979 | 0.3236 | 0.5038 | 0.1915 | 0.2491 |

**Table 5.** Optimum user-defined parameters.

| Kernel Functions | GP |
|---|---|
| RBF kernel | Gaussian noise = 0.1, γ = 12 |
| PUK kernel | Gaussian noise = 0.1, = 0.1, σ = 1 |

### *Philip's model*

Philips (1957) suggested the following theoretical infiltration equation based on physical properties of soil and analysis of water penetration into a uniform soil.

$$f(t) = \frac{1}{2} St^{-0.5} + A \tag{3}$$

where f(*t*) is the infiltration rate ($LT^{-1}$) as a function of time, *m* and *n* are the equation's parameters and *t* is time (T), *S* is the Sorptivity parameter that is function of soil matrix forces ($LT^{-0.5}$), and *A* is the soil parameter related to transmission of water through soil or gravity force($LT^{-1}$).

### *Model performance evaluation criteria*

To analyze the capability of various modeling methods in estimating infiltration rate of soil, correlation coefficient (CC), root mean square error (RMSE), and Nash–Sutcliffe model efficiency (NSE) values were calculated using the training and the testing data-sets.

$$CC = \frac{n\sum HF - (\sum H)(\sum F)}{\sqrt{n(\sum H^2) - (\sum H)^2}\sqrt{n(\sum F^2) - (\sum F)^2}} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(H - F)^2} \tag{5}$$

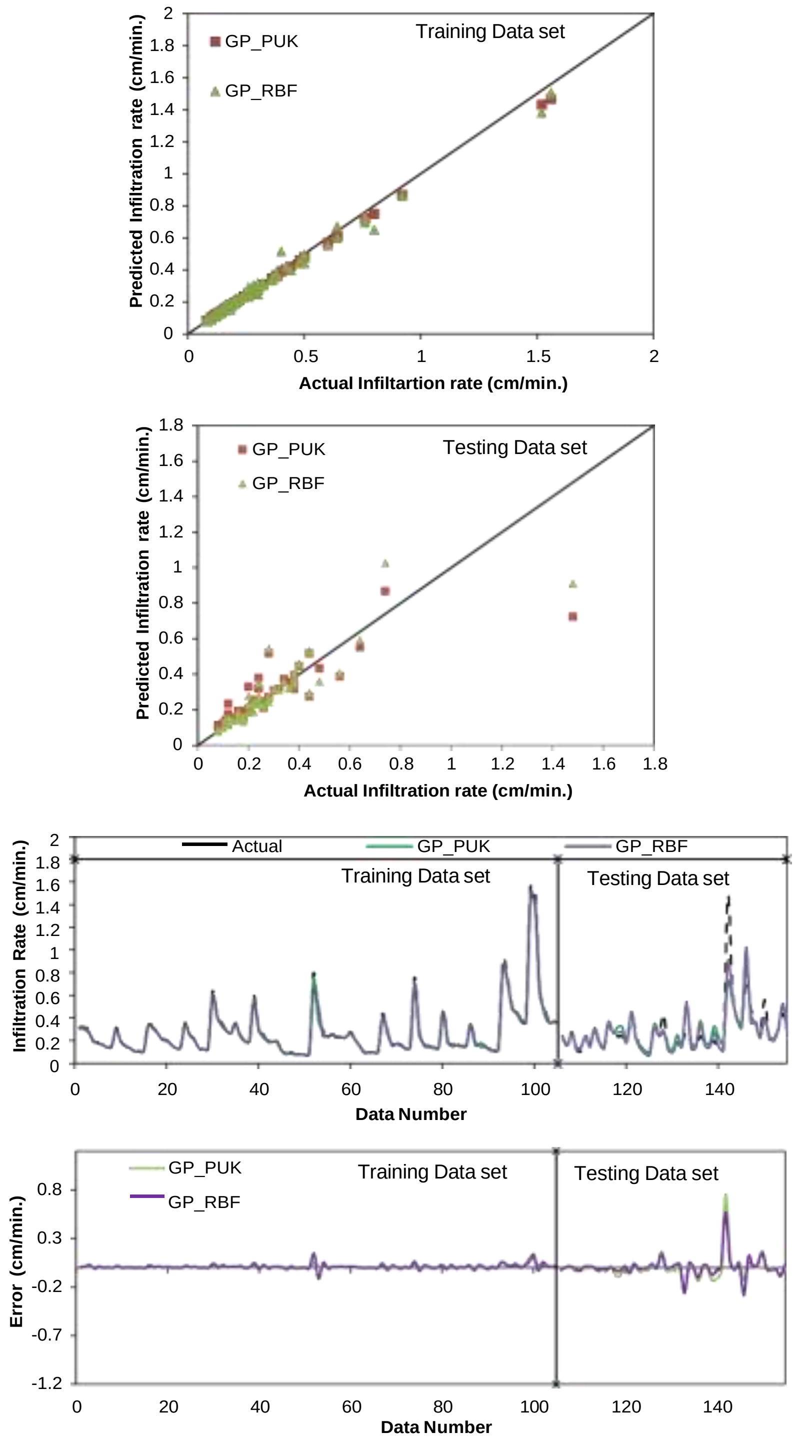



**Figure 8.** Performance of GP models during training and testing stages.

**Table 6.** Genetic operators implemented in GEP.

| Parameter | Description | Value |
|---|---|---|
| A | Mutation rate | 0.00136 |
| B | Inversion rate | 0.00546 |
| C | One-point recombination rate | 0.00277 |
| D | Two-point recombination rate | 0.00277 |
| E | Gene recombination rate | 0.00277 |
| F | Gene transposition rate | 0.00277 |

$$NSE = 1 - \frac{\sum_{i=1}^{n}(H-F)^2}{\sum_{i=1}^{n}(H-\bar{H})^2} \quad (6)$$

where $H$ = observed values, $F$ = predicted values, $\bar{H}$= mean of observed values, $n$ = number of observations.

## 3. Materials and methods

### Study area

Davood Rashid and Honam covering two parts of Lorestan province and Kelat area in Ilam province (Iran) were selected for measurement of infiltration. Davood Rahid area is located at 47°41′34.21″E and 33°33′30.31″N, Honam area is located at 48°16′41.97″E and 33°47′20.01″N, and Kelat area is located at 47°50′38.97″E and 32°38′26.37″N. Figure 4 showed the study area. Table 1 showed the coordinate system of all locations.

### Methodology

Infiltration rates were observed with the help of a cylindrical infiltrometer for all the selected sites. As shown in Figure 5, the

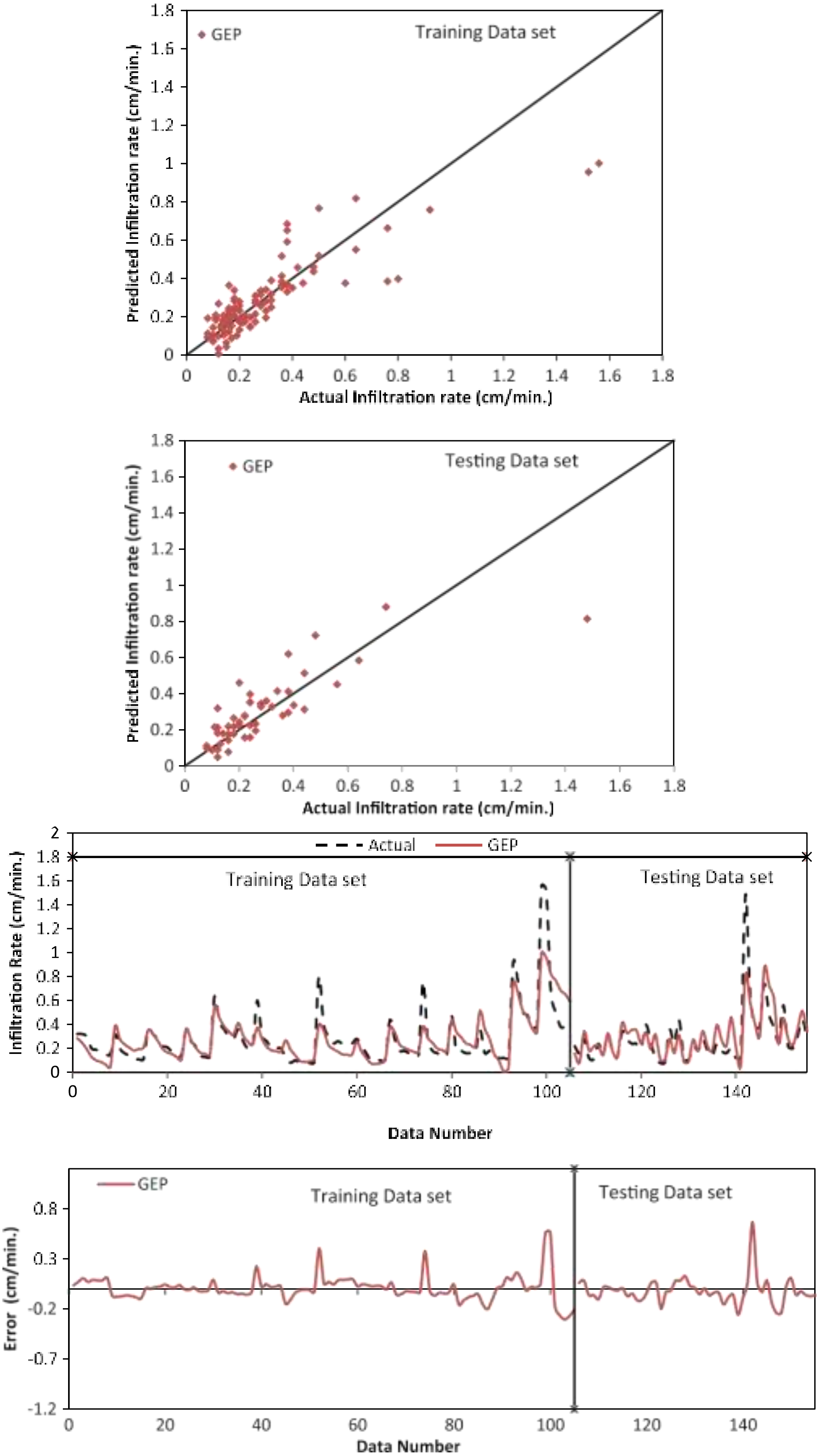



**Figure 9.** Performance of GEP models during training and testing stages.

**Table 7.** Formulation in GEP.

```
G1C3 = - - 8.22604555009296
G8C0 = - 1.47033398526557

y0 = - 0.0
y1 = - tan(sin(G1C3))
y2 = - y1 + tan(tan((((d[Silt]*d[Time])*(log(sin(d[Time]))/log(10)))*d[Silt])))
y3 = - y2 + sin(d[Sand])
y4 = - y3 + sin(sin(sin((sin(sin(log(cos(d[Clay]))))*d[Time]))))
y5 = - y4 + ((log((sin((d[Sand]*cos(d[Silt])))+cos((d[Time]*d[Sand])))))/
  log(10))+d[Silt])
y6 = - y5 + (cos((d[Clay]*cos(d[Silt])))+(d[BD]-(cos(d[Clay])+d[BD])))
y7 = - y6 + tan(tan(tan(tan(((log(sin(sin(d[Time])))/log(10))*d[Time])))))
y8 = - y7 + (d[Clay]/cos(sin((tan(cos(d[Clay]))*(G8C0+sin(d[Silt]))))))
result = Y8
```

cylindrical infiltrometer has two concentric rings. The diameter of the central and outer rings was 30 and 60 cm, respectively, with 30 cm depth. The rings were driven at about 10 cm deep in soil using falling weight-type hammer striking on a wooden plank placed on the top ring uniformly without disturbance of soil surface or minimum disturbance. Both rings were filled with water to equal depth and the initial reading of the water level was noted. The level of water in the central ring of infiltrometer was recorded at a regular interval of 2.5, 5, 10, 15, 20, 25, 30 min. until the rate of infiltration became constant. The infiltration rate was then calculated from the observed cumulative infiltration data. The observations of initial and final infiltration rate and properties of the soil are listed in Table 2 for all the sites. The soil samples for the calculating soil properties were collected from a site close to the location selected for the measurement of infiltration rate. Moisture content was determined by gravimetric method.

Flowchart of the research is presented in Figure 6. The initial step was the selection of site and instrument for experimentation. Total data were divided into two randomly selected parts of training (105) and testing (50). Training data-set was selected to calibrate the conventional and soft computing-based infiltration models and testing data-set was selected to validate the models.

The steps (optimum performance evaluation parameters) of flowchart were repeated till optimum user-defined parameters were achieved. The performance evaluation parameters such as CC, RMSE, and NSE were used to compare the performance of conventional and soft computing-based infiltration models.

### Data-set

To assess the effectiveness of GP, GEP, GRNN, and ANN to estimate the infiltration rate of the soil, data used in the present study were taken from field data. In this study, the data-set contains 155 experimental observations. Out of the 155 observations, arbitrarily selected 105 readings were selected for preparing, whereas the rest 50 were selected for testing the models. To estimate the infiltration rate of soils, six input parameters, time (T) in percent, moisture content (W) in percent, percentage of clay (C) silt (Si) and sand (Sa), and bulk density (B) in g/cc, were used, whereas infiltration rate f(t) in cm/min was the output. Table 3 provides the characteristics of the training and testing data-sets.

## 4. Result and discussion

Kostiakov model

$$f(t) = 0.8924t^{-0.44} \tag{7}$$

Philip's model

$$f(t) = \frac{1}{2}1.7787\, t^{-0.5} + 0.0381 \tag{8}$$

Results of Kostiakov model and Philip's model were plotted versus actual infiltration rate and are shown in Figure 7. Standard error indices including CC, RMSE, and NSE were implemented to access the accuracy of the conventional models. (see Table 4). As shown in Table 4 and Figure 7, the outcomes of Kostiakov and Philip's models are not satisfactory in predicting the infiltration rate of soils for this data-set.

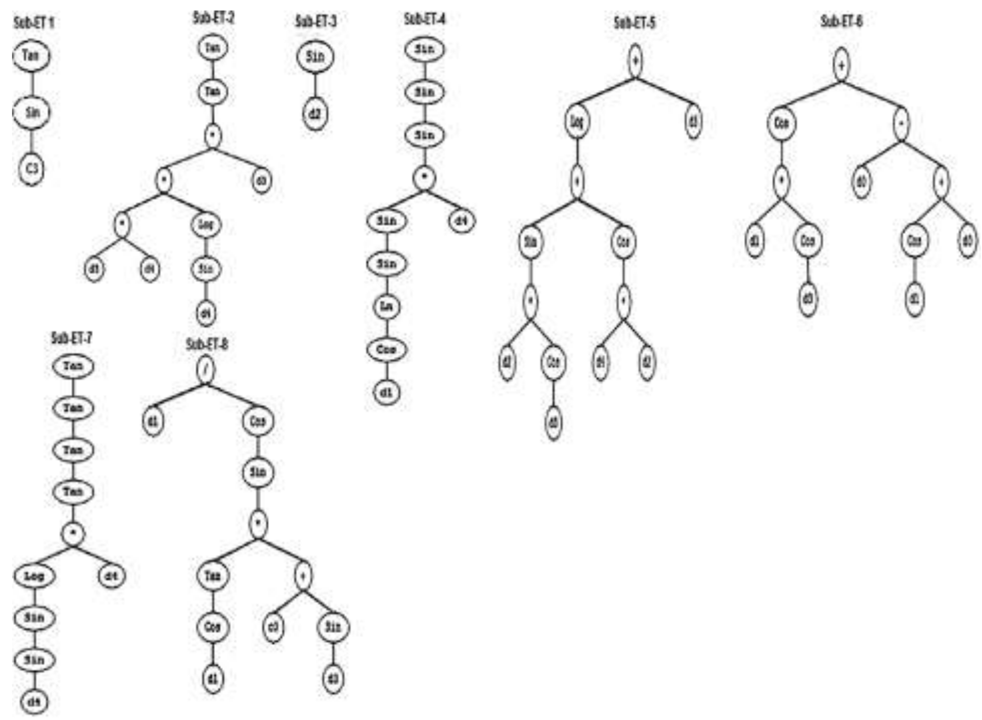



**Figure 10.** Expression of the GEP models.

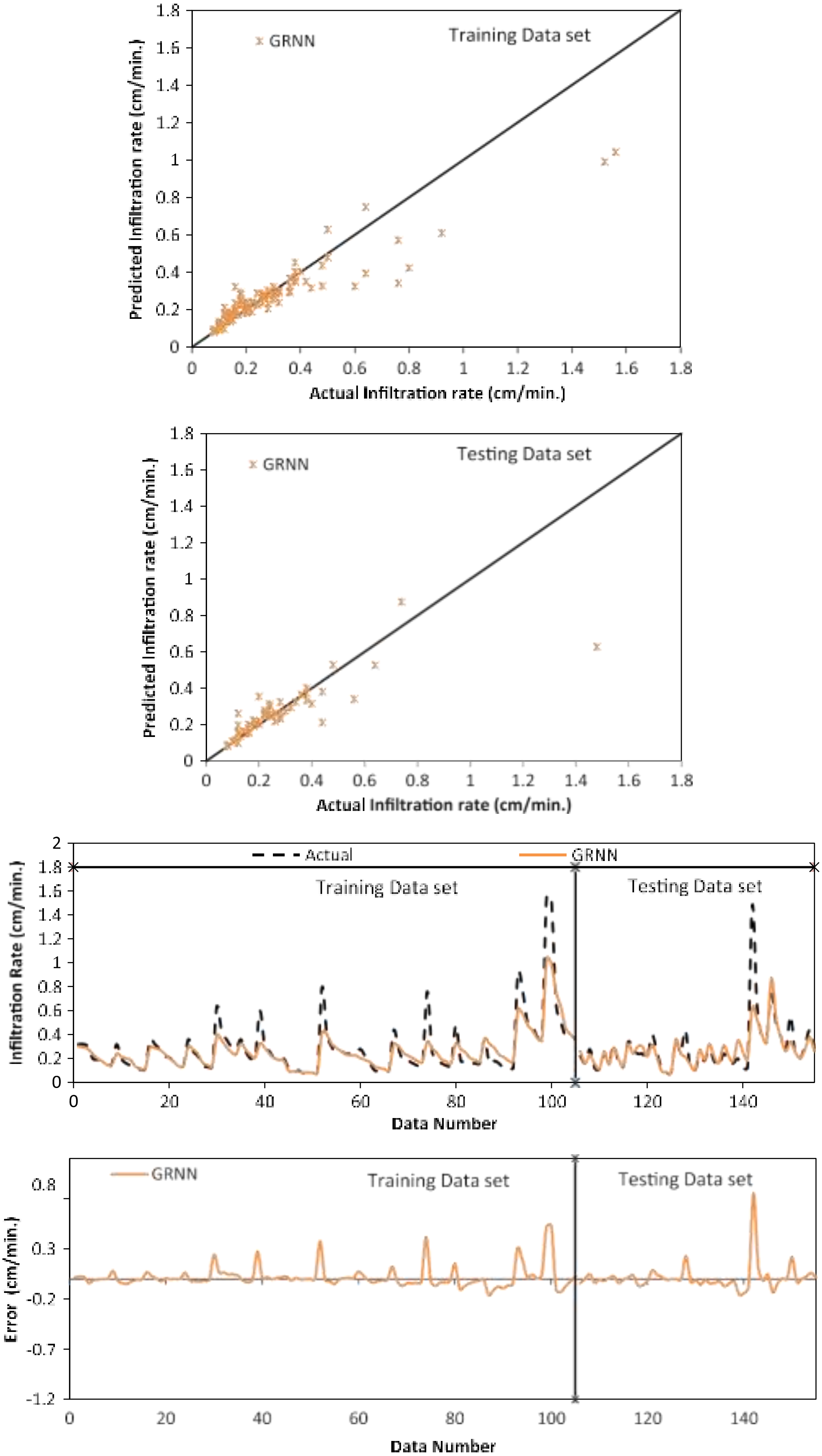



**Figure 11.** Performance of GRNN models during training and testing stages.

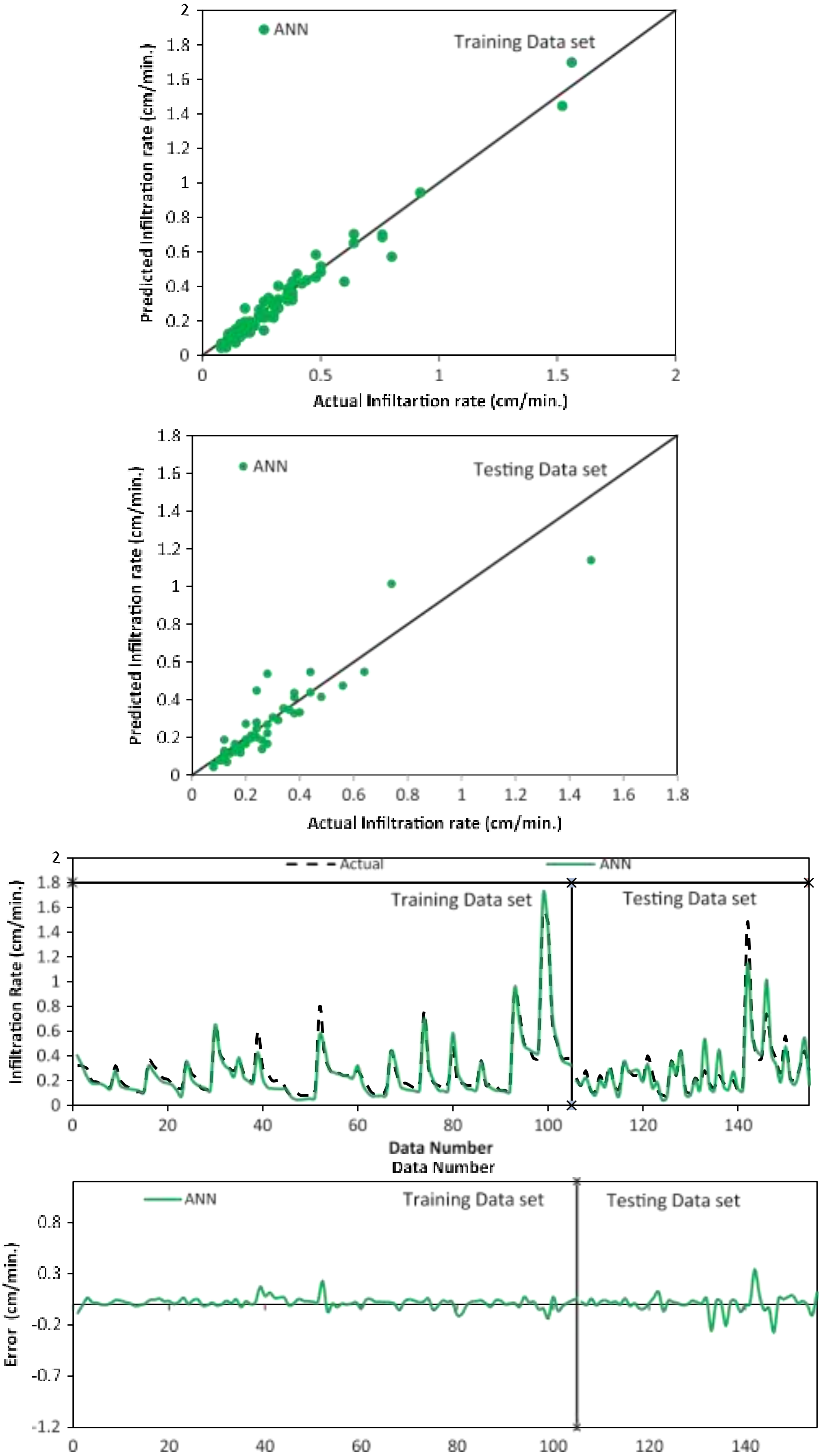



**Figure 12.** Performance of GRNN models during training and testing stages.

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

### Results of Gaussian process (GP)

Developing Gaussian process regression-based models (Gaussian noise, $\gamma$, $\sigma$, and ω) is a trial and error process. Two kernel functions (PUK and RBF) were used to develop the models. Gaussian noise (0.1) was kept constant for both kernels for the fair comparison of models. The optimum user-defined parameters are shown in Table 5. During the GP model development (Table 8), it was found that the RBF kernel has a better performance compared with PUK kernel function. The performance of GP models in every stage of development (training and testing) is presented in Figure 8. To examine the accuracy of these models, error indices for every stage of preparation were estimated and shown in this figure. The CC values of RBF kernel function-based GP model were obtained as 0.9950 and 0.8723 for training and testing, respectively. Overall, assessing Figure 8 shows that the exactness of the RBF kernel function-based GP model is suitable for predicting the infiltration rate of the soil. It is notable that in these figures the actual is associated with actual values and GP_RBF is associated with the results of the RBF kernel function-based GP model.

### Results of GEP

Developing the GEP model is similar to developing the GP model, based on a data-set. Design of GEP includes chromosomes and quantity of genes. In this study, 300 chromosomes and quantity of gene is 8 selected. Genetic operations used in GEP modeling are shown in Table 6. Results of GEP model to estimate infiltration rate of soil are shown in Figure 9. The optimum parameters of GEP models are shown in Table 6. As shown in Figure 9, the performance and error are plotted as well as assessing the performance of the GEP model in training and testing periods. The CC of GEP model was 0.8481, 0.8020, RMSE values 0.1807, 0.2158, and NSE values 0.7153, 0.6343 for the training and testing periods, respectively. The formulation and expression for the GEP were tabulated in Table 7 and Figure 10, respectively.

### Results of GRNN

Developing the GRNN model is similar to GP and GEP models. Development of GRNN is also based on the data-set. For preparing the GRNN model, spread (s) needs to be selected. Choosing the value of spread is a trial and error process. In this study, the optimum value of spread was achieved at 0.2. During the GRNN training, the obtained value for CC is 0.9238, for RMSE 0.1139, for NSE 0.7759 and when testing the model, the CC was 0.7984, the RMSE 0.1386, and the NSE 0.6071. To give more information about the GRNN model performance, the agreement, performance, and error distribution were plotted during both stages of model development (Figure 11).

### Results of ANN

Developing the ANN model (e.g. number of neurons in hidden layer, number of hidden layers, momentum, learning rate, iteration, etc.) is a trial and error process. ANN model contains a single hidden layer with 9 neurons, momentum = 0.2, learning rate = 0.1, and iteration = 1500. The performance of the ANN model is shown in Figure 12. As shown in Table 8, ANN model obtained CC = 0.9133, RMSE = 0.0911, and NSE = 08302 for the testing stage. Overall, assessing Table 8 and Figure 12 shows

**Table 8.** Performance of GP, GEP, GRNN, and ANN models.

| | Performance evaluation parameters | | | | | |
|---|---|---|---|---|---|---|
| | Training | | | Testing | | |
| Approaches | CC | RMSE (cm/min.) | NSE | CC | RMSE (cm/min.) | NSE |
| GP_PUK | 1.0000 | 0.0190 | 0.9938 | 0.8323 | 0.1269 | 0.6706 |
| GP_RBF | 0.9950 | 0.0305 | 0.9839 | 0.8723 | 0.1086 | 0.7589 |
| GEP | 0.8481 | 0.1807 | 0.7153 | 0.8020 | 0.2158 | 0.6343 |
| GRNN | 0.9238 | 0.1139 | 0.7759 | 0.7984 | 0.1386 | 0.6071 |
| ANN | 0.9816 | 0.0502 | 0.9564 | 0.9133 | 0.0911 | 0.8302 |

**Table 9.** Result of Single-Factor ANOVA test for GP, GEP, GRNN, and ANN approaches.

| Approaches | *F* | *P-value* | *F critical* |
|---|---|---|---|
| GP_PUK | 0.001559 | 0.968581 | 3.938111 |
| GP_RBF | 0.022204 | 0.881853 | 3.938111 |
| GEP | 0.263606 | 0.608809 | 3.938111 |
| GRNN | 0.103051 | 0.748882 | 3.938111 |
| ANN | 0.041871 | 0.838288 | 3.938111 |

**Table 10.** Sensitivity analysis using ANN.

| | | ANN | | |
|---|---|---|---|---|
| Input combination | Input parameter removed | Coefficient of correlation | Root mean square error (cm/min.) | NSE |
| T, C, Si, Sa, B, W | | 0.9133 | 0.0911 | 0.8302 |
| C, Si, Sa, B, W | T | 0.6347 | 0.1716 | 0.3977 |
| T, Si, Sa, B, W | C | 0.9068 | 0.0933 | 0.8218 |
| T, C, Sa, B, W | Si | 0.8679 | 0.111 | 0.7481 |
| T, C, Si, B, W | Sa | 0.8982 | 0.0975 | 0.8056 |
| T, C, Si, Sa, W | B | 0.92 | 0.0874 | 0.8439 |
| T, C, Si, Sa, B | W | 0.8944 | 0.1074 | 0.7639 |

that the accuracy of ANN model was more suitable for the estimation of the infiltration rate.

### Comparison of models

Comparison of the soft computing-based models with the conventional models shows that these models perform better than conventional models (Table 4 and Table 7). Comparison of soft computing models indicates that ANN models work well than other soft computing-based models. RBF and PUK kernel function-based GP models work well than GEP and GRNN models. The comparison of RBF kernel-based GP model with PUK kernel function-based GP model in Table 8 indicates that RBF kernel-based GP model works better than PUK kernel function-based GP model. Single-factor ANOVA results (Table 9) show that *F*-values were less than *f*-critical and *P*-values were greater than 0.05, suggesting that the difference in the estimated values of GP, GEP, GRNN, ANN and actual values is insignificant. To compare the performance of GP, GEP, GRNN, and ANN models, agreement, performance, and error were plotted in Figure 13 for both training and testing stages. It can be observed from the figure that the estimated values produced by ANN model were in extreme proximity to the actual infiltration rate and the estimated infiltration rate is found to chase a similar pattern as that of the actual infiltration rate of the soil.

### Sensitivity study

A sensitivity study was performed to find the main significant input parameters in the estimation of infiltration

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

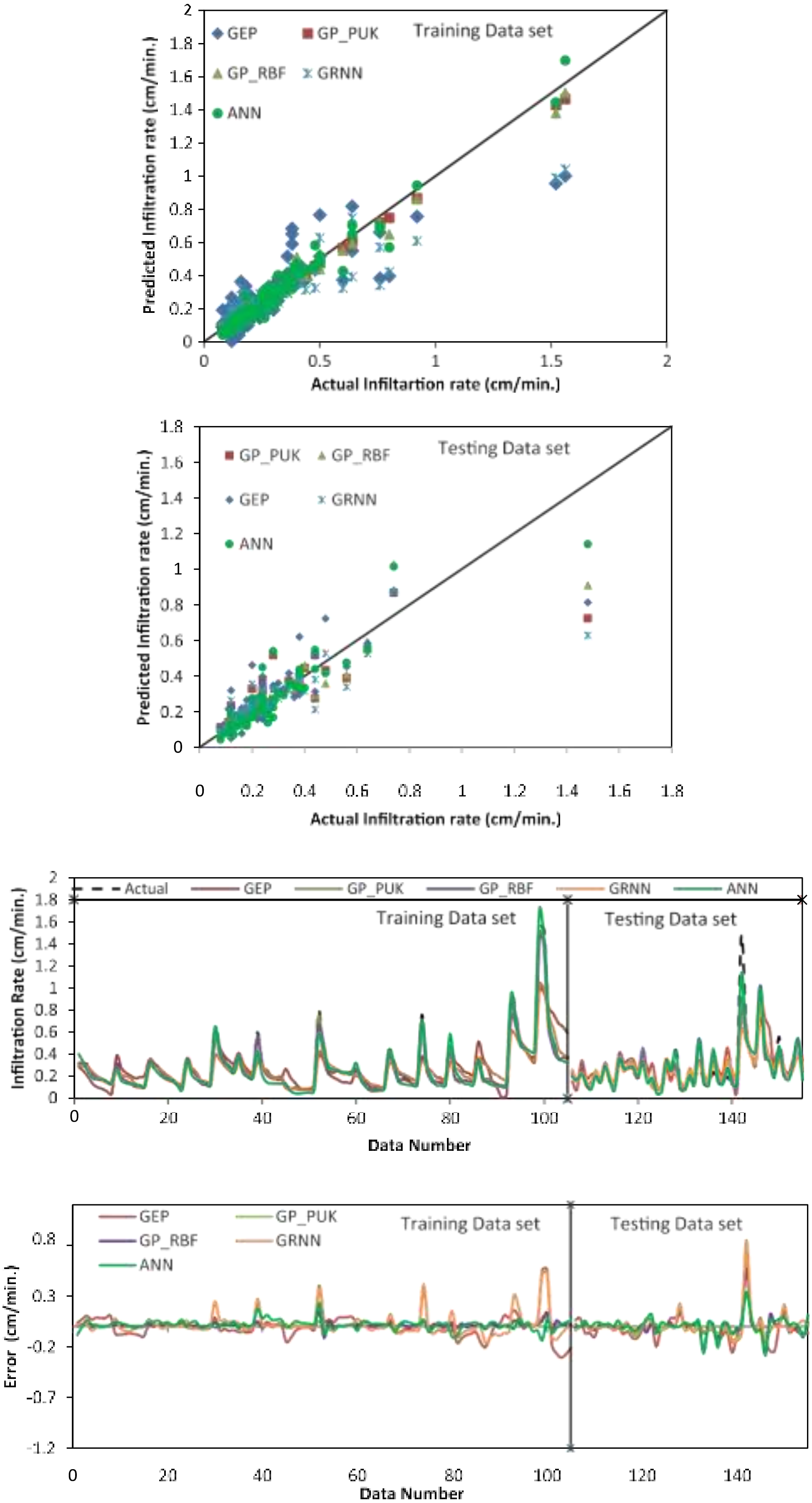



**Figure 13.** Performance of GP, GEP, and GRNN models.

rate of the soil. For this, ANN model performing best with this data-set was selected. Several sets of training data were generated by eliminating single input parameter at a time and outcomes were recorded in terms of CC, RMSE, and NSE with the testing data-set. Results from Table 10 suggest that time has an important role in predicting/estimating the infiltration rate of soil in comparison to other input parameters.

## 5. Conclusion

Prediction of the infiltration rate is an essential element of hydrologic design, watershed management, irrigation, and agriculture studies. Results of this study showed that estimating the infiltration rate of soil with conventional formulae leads to apparently incredible errors in calculation. Based on the obtained results, the ANN model has a suitable capability to predict the infiltration rate of the soil. The ANN model also provides better performance than the GP, GEP, and GRNN models. Sensitivity results suggest that time is the most important parameter when ANN-based modeling approach is used for predicting the infiltration rate of soil for this data-set.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Angelaki, A., Sakellariou-Makrantonaki, M., and Tzimopoulos, C. (2013). "Theoretical and experimental research of cumulative infiltration." *Transp. Porous Media*, 100(2), 247–257.

Azamathulla, H.M., Haghiabi, A.H., and Parsaie, A. (2016a). "Prediction of side weir discharge coefficient by support vector machine technique." *Water Sci. Technol.*, 16(4), 1002–1016.

Azamathulla, H.M., Haghiabi, A.H., and Parsaie, A. (2016b). "Prediction of side weir discharge coefficient by support vector machine technique." *Water Sci. Technol.*, 16(4), 1002–1016.

Baba, A.P.A., Shiri, J., Kisi, O., Fard, A.F., Kim, S., and Amini, R. (2013). "Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN)." *Hydrol. Res.*, 44(1), 131–146.

Ferreira, C. (2002). "Gene expression programming in problem-solving." In *Soft Computing and Industry*, Springer, London, 635–653.

Ferreira, C. (2006). *Gene expression programming: mathematical modeling by an artificial intelligence* (Vol. 21), Springer.

Haghiabi, A.H., Parsaie, A., and Ememgholizadeh, S. (2017a). "Prediction of discharge coefficient of triangular labyrinth weirs using Adaptive Neuro Fuzzy inference system." *Alexandria Eng. J.* doi:10.1016/j.aej.2017.05.005.

Haghiabi, A.H., Azamathulla, H.M., and Parsaie, A. (2017b). "Prediction of head loss on cascade weir using ANN and SVM." *ISH J. Hydraulic Eng.*, 23(1), 102–110.

Haghiabi, A.H., Nasrolahi, A.H., and Parsaie, A. (2018). "Water quality prediction using machine learning methods." *Water Qual. Res. J.*, 53(1), 3–13.

Haykin, S. (1999). *Neural networks a comprehensive foundation*. International editions series, Publisher, Prentice Hall.

Karimi, S., Shiri, J., Kisi, O., and Shiri, A.A. (2016). "Short-term and long-term streamflow prediction by using 'wavelet–gene expression' programming approach." *ISH J. Hydraulic Eng.*, 22(2), 148–162.

Keshavarzi, A., Sarmadian, F., Shiri, J., Iqbal, M., Tirado-Corbalá, R., and Omran, E.E. (2017). "Application of ANFIS-based subtractive clustering algorithm in soil cation exchange capacity estimation using soil and remotely sensed data." *Measurement*, 95, 173–180.

Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petkovic, D., and Hashim, R. (2015). "A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm." *Appl. Math. Comput.*, 270, 731–743.

Kisi, O., Keshavarzi, A., Shiri, J., Zounemat-Kermani, M., and Omran, E.E. (2017). "Groundwater quality modeling using neuro-particle swarm optimization and neuro-differential evolution techniques." *Hydrol. Res.*, 48(6), 1508–1519.

Kostiakov, A.N. (1932). "On the dynamics of the coefficient of water-percolation in soils and on the necessity for studying it from a dynamic point of view for purposes of amelioration." *Trans*, 6, 17–21.

Kumar, M., Ranjan, S., Tiwari, N.K., and Gupta, R. (2018a). "Plunging hollow jet aerators-oxygen transfer and modelling." *ISH J. Hydraulic Eng.*, 24(1), 61–67.

Kumar, M., Tiwari, N.K., and Ranjan, S. (2018b). "Prediction of oxygen mass transfer of plunging hollow jets using regression models." *ISH J. Hydraulic Eng.*, 1–8.

Kuss, M. (2006). "Gaussian process models for robust regression, classification, and reinforcement learning." Doctoral dissertation, Ph.D. thesis, TechnischeUniversität, Darmstadt, 189.

Mehdipour, V., and Memarianfard, M. (2017). "Application of Support Vector Machine and Gene Expression Programming on Tropospheric ozone Prognosticating for Tehran Metropolitan." *Civ. Eng. J.*, 3(8), 557–567.

Mehdipour, V., Memarianfard, M., and Homayounfar, F. (2017). "Application of Gene Expression Programming to water dissolved oxygen concentration prediction." *Int. J. Human Capital Urban Manage.*, 2(1), 39–48.

Mishra, S.K., Tyagi, J.V., and Singh, V.P. (2003). "Comparison of infiltration models." *Hydrol. Processes*, 17(13), 2629–2652.

Najafzadeh, M., Shiri, J., Sadeghi, G., and Ghaemi, A. (2018a). "Prediction of the friction factor in pipes using model tree." *ISH J. Hydraulic Eng.*, 24(1), 9–15.

Najafzadeh, M., Shiri, J., and Rezaie Balf, M. (2018b). "New expression-based models to estimate scour depth at clear water conditions in rectangular channels." *Mar. Georesour. Geotechnol.*, 36(2), 227–235.

Parsaie, A., and Haghiabi, A. (2014). "Predicting the side weir discharge coefficient using the optimized neural network by genetic algorithm." *Sci. J. Pure Appl. Sci.*, 3(3), 103–112.

Parsaie, A., and Haghiabi, A. (2015a). "The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir." *Water Resour. Manage.*, 29(4), 973–985.

Parsaie, A., and Haghiabi, A.H. (2015b). "Predicting the longitudinal dispersion coefficient by radial basis function neural network." *Model. Earth Syst. Environ.*, 1(4), 1–8.

Parsaie, A., and Haghiabi, A.H. (2017a). "Improving modelling of discharge coefficient of triangular labyrinth lateral weirs using SVM, GMDH and MARS techniques." *Irrigation Drainage*, 66(4), 636–654.

Parsaie, A., and Haghiabi, A.H. (2017b). "Prediction of side weir discharge coefficient by genetic programming technique." *Jordan J. Civ. Eng.*, 11 (1): 132–141.

Parsaie, A., Azamathulla, H.M., and Haghiabi, A.H. (2017a). "Prediction of discharge coefficient of cylindrical weir–gate using GMDH-PSO." *ISH J. Hydraulic Eng.*, 1–8.

Parsaie, A., Najafian, S., Omid, M.H., and Yonesi, H. (2017b). "Stage discharge prediction in heterogeneous compound open channel roughness." *ISH J. Hydraulic Eng.*, 23(1), 49–56.

Parsaie, A., Yonesi, H., and Najafian, S. (2017c). "Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method." *Flow Meas. Instrum.*, 54, 288–297.

Pedretti, D., Barahona-Palomo, M., Bolster, D., Sanchez-Vila, X., and Fernandez-Garcia, D. (2012). "A quick and inexpensive method to quantify spatially variable infiltration capacity for artificial recharge ponds using photographic images." *J. Hydrol.*, 430, 118–126.

Philips, J.R. (1957). "The theory of infiltration: the infiltration equation and its solution." *Soil Sci.*, 83(5), 345–357.

Rasmussen, C.E., and Williams, C.K. (2006). *Gaussian processes for machine learning* (Vol. 1), Cambridge: MIT Press, 248.

Richards, L.A. (1931). "Capillary conduction of liquids through porous mediums." *J. Appl. Phys.*, 1(5), 318–333.

Roushangar, K., Mouaze, D., and Shiri, J. (2014). "Evaluation of genetic programming-based models for simulating friction factor in alluvial channels." *J. Hydrol.*, 517, 1154–1161.

Rosushangar, K., Mirheidarian, S., and Shiri, J. (2017). "Modeling local pier scour with bed effect implications: Heuristic vs. empirical strategies." *ISH J. Hydraulic Eng.*, 23(1), 13–22.

Shiri, J. (2017). "Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran." *Agric. Water Manage.*, 188, 101–114.

Shiri, J., and Kisi, O. (2012). "Estimation of daily suspended sediment load by using wavelet conjunction models." *ASCE J. Hydrol. Eng.*, 17(9), 986–1000.

Shiri, J., Shamshirband, S., Kisi, O., Karimi, S., Bateni, S.M., HosseiniNazhad, S.H., and Hashemi, A. (2016). "Prediction of water-level in the Urmia lake using the extreme learning machine approach." *Water Resour. Manage.*, 30, 5217–5229.

Shiri, J., Keshavarzi, A., Kisi, O., and Karimi, S. (2017a). "Using soil easily measured parameters for estimating soil water capacity: Soft computing approaches." *Comput. Electron. Agric.*, 327–339.

Shiri, J., Keshavarzi, A., Kisi, O., Iturraran-Viveros, U., Bagherzadeh, A., Mousavi, R., and Karimi, S. (2017b). "Modeling soil cation exchange capacity using soil parameters: Assessing the heuristic models." *Comput. Electron. Agric.*, 135, 242–251.

Shiri, J., Keshavarzi, A., Kisi, O., Karimi, S., and Iturraran-Viveros, U. (2017c). "Modeling soil bulk density through a complete data scanning procedure: Heuristic alternatives." *J. Hydrol.*, 549, 592–602.

Sihag, P. (2018). "Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network." *Model. Earth Syst. Environ.*, 1–10.

Sihag, P., Tiwari, N.K., and Ranjan, S. (2017a). "Estimation and inter-comparison of infiltration models." *Water Sci.*, 31(1), 34–43.

Sihag, P., Tiwari, N.K., and Ranjan, S. (2017b). "Modelling of infiltration of sandy soil using gaussian process regression." *Model. Earth Syst. Environ.*, 3(3), 1091–1100.

Sihag, P., Tiwari, N.K., and Ranjan, S. (2017c). "Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS)." *ISH J. Hydraulic Eng.*, 1–11.

Sihag, P., Jain, P., and Kumar, M. (2018a). "Modeling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression." *Model. Earth Syst. Environ.*, 1–8.

Sihag, P., Tiwari, N.K., and Ranjan, S. (2018b). "Support vector regression-based modeling of cumulative infiltration of sandy soil." *ISH J. Hydraulic Eng.*, 1–7.

Singh, V.P., and Yu, F.X. (1990). "Derivation of infiltration equation using systems approach." *J. Irrigation Drainage Eng.*, 116(6), 837–858.

Singh, B., Sihag, P., and 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.

Specht, D.F. (1991). "A general regression neural network." *IEEE Trans. Neural Networks*, 2(6), 568–576.

Sy, N.L. (2006). "Modelling the infiltration process with a multi-layer perceptron artificial neural network." *Hydrol. Sci. J.*, 51(1), 3–20.

Tiwari, N.K., Sihag, P., and Ranjan, S. (2017). "Modeling of infiltration of soil using Adaptive Neuro-fuzzy Inference System (ANFIS)." *J. Eng. Technol. Educ.*, 11(1), 13–21.

Tiwari, N.K., Sihag, P., Kumar, S., and Ranjan, S. (2018). "Prediction of trapping efficiency of vortex tube ejector." *ISH J. Hydraulic Eng.*, 1–9.

Wasserman, P.D. (1993). *Advanced methods in neural computing*, Wiley, New York.

Yavari, S., Maroufpoor, S., and Shiri, J. (2017). "Modeling soil erosion by data-driven methods using limited input variables." *Hydrol. Res.*, 49(1), 1–14.