

Research Article

A Comparative Analysis of Data-Driven Empirical and Artificial Intelligence Models for Estimating Infiltration Rates

Mohammad Zakwan ¹ and Majid Niazkar ²

¹Civil Engineering Department, Maulana Azad National Urdu University, Hyderabad, India

²Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

Correspondence should be addressed to Majid Niazkar; mniazkar@shirazu.ac.ir

Received 16 March 2021; Revised 8 April 2021; Accepted 24 April 2021; Published 5 May 2021

Academic Editor: Teddy Craciunescu

Copyright © 2021 Mohammad Zakwan and Majid Niazkar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Infiltration is a vital phenomenon in the water cycle, and consequently, estimation of infiltration rate is important for many hydrologic studies. In the present paper, different data-driven models including Multiple Linear Regression (MLR), Generalized Reduced Gradient (GRG), two Artificial Intelligence (AI) techniques (Artificial Neural Network (ANN) and Multigene Genetic Programming (MGGP)), and the hybrid MGGP-GRG have been applied to estimate the infiltration rates. The estimated infiltration rates were compared with those obtained by empirical infiltration models (Horton's model, Philip's model, and modified Kostiakov's model) for the published infiltration data. Among the conventional models considered, Philip's model provided the best estimates of infiltration rate. It was observed that the application of the hybrid MGGP-GRG model and MGGP improved the estimates of infiltration rates as compared to conventional infiltration model, while ANN provided the best prediction of infiltration rates. To be more specific, the application of ANN and the hybrid MGGP-GRG reduced the sum of square of errors by 97.86% and 81.53%, respectively. Finally, based on the comparative analysis, implementation of AI-based models, as a more accurate alternative, is suggested for estimating infiltration rates in hydrological models.

1. Introduction

Infiltration can be defined as the process by which water enters the surface of Earth [1]. It leads to the entrance of water into the soil, thereby catering to groundwater recharge and subsurface runoff. In essence, the infiltration phenomenon is among the most crucial processes of water cycle. Furthermore, estimates of infiltration capacity of soil is required in the design of efficient irrigation systems, estimation of evapotranspiration, groundwater recharge, surface runoff, effective rainfall, crop water requirement, and transport of chemicals in surface and subsurface water [2]. As a result, modelling and prediction of infiltration rates is an inevitable part of hydrological modelling. For instance, Morel-Seytoux [3] reviewed the importance of infiltration in large-scale hydrologic modelling. Furthermore, Šraj et al. [4] pointed towards the impact of the estimation of infiltration rates on the runoff hydrograph, which plays a vital role in

watershed modelling and water management. Similarly, Wen et al. [5] demonstrated the implication of excessive infiltration on watershed models. Finally, these studies demonstrated why an accurate estimation of time-dependent infiltration is important in hydrological modelling.

Owing to the wide applications of the infiltration rate, its estimation has gained significant attention from researchers. Over the years, various infiltration models have been proposed by the researchers for the estimation of infiltration rates. They include models that have physical, semiempirical, and even empirical formulations. Despite the development of several models, no single model exists that outperforms other ones universally. The suitability of infiltration model for a particular site depends on the type of soil and field conditions [1]. In this regard, many comparative studies have been conducted to assess the suitability of various infiltration models for different soil types under varying field conditions. Mishra et al. [6] conducted one of the most

comprehensive analyses on suitability of infiltration models for different soils. Similarly, the methodology used to model infiltration rate has a significant impact on the estimation of infiltration. Deep and Das [7] compared various optimization algorithms to estimate the parameters of infiltration models. Nonetheless, the application of different optimization techniques can only move the solution from local optimum parameters towards global optimum parameters, while they cannot increase the flexibility of infiltration models to mimic actual infiltration rates. Haghiabi et al. [8] employed a dimensionless form of infiltration data to estimate infiltration parameters accurately. However, Zakwan [9] suggested that such transformation may not necessarily improve the accuracy of infiltration equations. Finally, Chen et al. [10] utilized genetic algorithm to improve the estimate of Green-Ampt infiltration model under a rainfall condition.

Recent applications of computational techniques in water resource engineering have widened the scope further [11–21]. With the advancement in the computational method and modelling approaches, the application of these approaches has provided a viable alternative towards the estimation of infiltration rates also. Kumar and Sihag [22] applied Gene Expression Programming (GEP) to model infiltration rates. Moreover, Dewidar et al. [23] proposed the application of fuzzy logic to estimate the infiltration rates. In addition, Patle et al. [24] employed a multiple linear regression model to predict time-dependent infiltration values based on several soil properties such as bulk density, silt, sand percentage, and moisture content. Furthermore, Sihag et al. [25] exploited the support vector machine (SVM) for modelling infiltration rates in sandy soil. Also, Pahlavan-Rad et al. [26] compared the performance of Multiple Linear Regression (MLR) and Random Forest Tree in depicting the spatial variation of infiltration rates and reported the superiority of Random Forest Tree over MLR. Recently, Sepahvand et al. [27] utilized several data-driven models to predict infiltration rates. Their investigation revealed the superiority of neural networks over other data-driven techniques such as model tree, Gaussian process, and regression analysis. According to the recent studies, considering time as the exclusive state variable in empirical models should be revisited in favour of better infiltration predictions, while the AI-based models were found to have a better performance in comparison with the conventional infiltration models. Therefore, despite previous efforts in improving the estimation of infiltration rates, further studies are needed to explore these issues.

The present study aims to compare the performances of different infiltration methods. Additionally, it attempts to assess the capability MGGP and of the novel hybrid MGGP-GRG to model the infiltration process. In a bid to seek for a better time-dependent infiltration model, the performances of the MGGP-based models were compared with those of the conventional models, regression techniques, and commonly used neural network.

2. Materials and Methods

2.1. Data. In the present study, the infiltration data reported by Sihag et al. [28] were utilized. The data were divided into

training and testing data sets. To be more precise, 75% of the data were used for training, while the rest of the data were exploited to test the obtained results. Table 1 summarises the data sets used in the present paper.

Figure 1 shows the observed infiltration data at the same time duration. Also, it illustrates that the infiltration rate may be dependent on other factors (soil properties, such as bulk density and sand percentage) apart from time. The infiltration data set, which was obtained from the literature [28], belongs to the infiltration observations carried out at Davood Rashid and Honam regions in Lorestan Province and the Kelat region in Ilam Province in Western Iran.

2.2. Conventional Infiltration Models. There are a number of infiltration models available in the literature. Brief description of some of the commonly used infiltration models considered in the present study is as follows.

(1) Horton's Model

Horton [29] proposed an empirical equation, which is presented in the following equation, for exponential decay of the infiltration rate after analysing several infiltrometer data sets:

$$f = f_c + (f_0 - f_c)e^{-kt}, \quad (1)$$

where f is the infiltration capacity at any time t from the start; f_c is the final or ultimate infiltration capacity occurring at $t = t_c$; f_0 is the initial infiltration capacity at time $t = 0$; and K is Horton's decay coefficient.

(2) Philip's Model

Philip [30] proposed an infinite series solution of Richard's equation to drive a relationship between the cumulative infiltration (F) and soil properties. It is presented in the following equation:

$$F = st^{0.5} + Kt. \quad (2)$$

By differentiating the above equation, the infiltration rate may be represented as

$$f = \frac{1}{2}st^{-0.5} + K. \quad (3)$$

(3) Modified Kostikov's Model

Kostiakov [31] observed the temporal variation of infiltration into soil and proposed a time-dependent infiltration model, invariantly known as Kostiakov's model. The major limitation of Kostiakov's model is that it approached to zero final infiltration rates rather than toward constant final infiltration rates and infinite infiltration rates at the start. Smith [32] modified Kostiakov's [31] equation to include the constant term f_i . The modified version is shown in the following equation:

$$f = (ab)t^{b-1} + f_i. \quad (4)$$

The parameters of different infiltration models were obtained by minimizing the sum of square of errors using a

TABLE 1: Training and testing data used.

Data	Count	Infiltration rate (cm/min)	Sand (%)	Density (g/cm ³)
Training	116	0.080–1.560	6.00–38	0.08–1.56
Testing	38	0.080–1.480	6.00–38.00	0.08–1.48

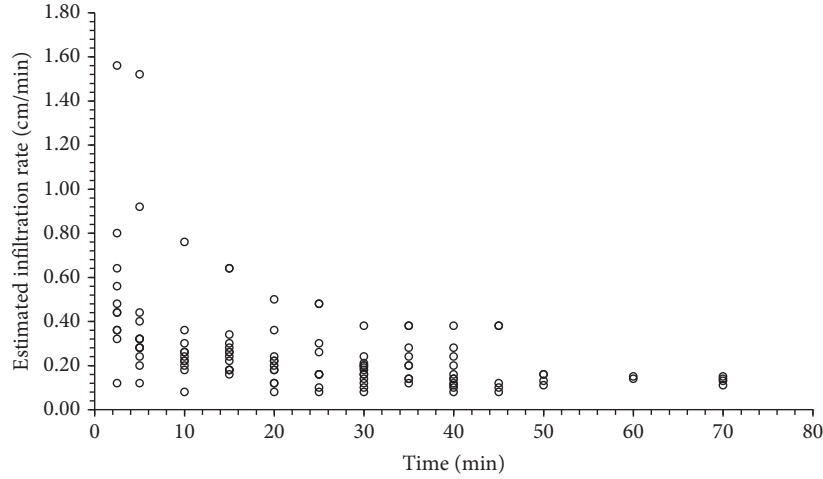


FIGURE 1: Observed infiltration data.

nonlinear optimization tool. Thus, the objective function becomes

$$\text{Min SSE} = \sum_{i=1}^N [f_{obs_i} - f_{est_i}]^2, \quad (5)$$

where f_{obs} is the observed infiltration rate and f_{est} is the estimated infiltration rate at any time t .

2.3. Multiple Linear Regression. MLR has been widely used in water resource engineering [17, 33]. It has also been applied to estimate the infiltration rate [25]. In accordance with MLR, infiltration rate can be expressed as

$$f = c_1 t + c_2 S + c_3 D + c_4, \quad (6)$$

where c_1, c_2, c_3, c_4 , and c_5 are coefficients, f is the infiltration rate in cm/min; t is time in minutes; S is the percentage of sand; D is the density in g/cm³.

2.4. Generalized Reduced Gradient (GRG). GRG is a gradient-based nonlinear optimization technique [34]. Earlier, Zakwan et al. [1] and Muzzammil et al. [35] suggested that GRG technique is superior to the conventional graphical method for estimating infiltration parameters and rating curve parameters. In accordance with GRG, the infiltration rate can be expressed as

$$i = c_5 t^{c_6} S^{c_7} D^{c_8}, \quad (7)$$

where c_5, c_6, c_7 , and c_8 are coefficients.

In the present study, GRG solver embedded in Microsoft Excel was used to estimate the infiltration rate based on minimizing the sum of square of errors. Detailed

explanation on working of GRG technique is available in the literature [17, 36].

2.5. Artificial Neural Network. ANN is one well-documented AI model. It has been used for solving various problems in water resources and hydrological modelling [37, 38]. Generally, ANN has a few layers, whose neurons store data. The neurons in each layer (input, hidden, and output layers) are connected with neurons in the previous and next layers, whereas there is no connection between neurons in a typical layer [39]. The flexible architecture of ANN basically facilitates the estimation of a relationship between input and output data [40]. In this study, a feed-forward ANN was exploited to predict the rate of the infiltration. The controlling parameters of ANN were set as those used in the previous studies [41].

2.6. Multigene Genetic Programming. MGGP is a modified version of genetic programming (GP), which is classified as an AI technique [42]. Not only does it utilize genetic algorithm as its search engine but also it works as a flexible estimator without the need to know the shape of a prediction model under investigation [43]. In essence, MGGP follows a similar solving approach as GP using a tree-like structure, while it enables the use of more than one gene, i.e., tree, in each individual. This characteristic benefits MGGP in the light of developing estimation models when the relation among involved variables is complicated to study. As a result, a typical MGGP solution consists of a set of equations, each associated with one gene, which is algebraically summed up using weighting coefficients. These coefficients are calibrated in MGGP, while a term invariantly called as bias is also added to the final solution. The terms comprising the

final solution of MGGP help in improving its flexibility in capturing the relationship between input and output data.

In this study, an open-access code of MGGP was exploited. This code was adopted from the literature [44], while it was used in previous studies for other purposes [20]. It minimizes the root mean square of errors between the estimated and observed values of the normalized infiltration rates. Additionally, the MGGP parameters were selected from previous studies [20, 43]. Since each run of MGGP may result in a unique equation, more than 100 runs of MGGP were considered for developing the relation between the infiltration rates and other variables involved. The common number of MGGP runs in the literature is 50 [20, 45], while the double number of runs, i.e., 100, was taken into account to make sure that the best relation was achieved.

2.7. Hybrid MGGP-GRG Technique. The hybrid MGGP-GRG was first proposed for developing stage-discharge

relationships in the literature [20]. In this technique, MGGP and GRG are used in two successive steps to find the best-fit model. Figure 2 depicts the flowchart of the hybrid MGGP-GRG for estimating infiltration rates. As shown, MGGP is initially operated to search for the best-fit form of equation to the data, while the GRG technique is utilized to optimize the coefficients of the equation obtained by MGGP. Hence, this hybrid technique not only benefits from the powerful capability of MGGP for seeking an accurate prediction model, but also uses GRG to enhance the performance of the estimation model.

2.8. Performance Evaluation Criteria. The performance of infiltration models and soft computing techniques was compared based on several criteria, which are presented in the following equations [28, 46]:

$$\begin{aligned}
 \text{normalised root mean square (NRMSE)} &= \frac{\sqrt{\sum_{i=1}^N (f_{obs_i} - f_{est_i})^2 / N}}{f_{obs_{max}} - f_{obs_{min}}}, \\
 \text{Wilmott Index (WI)} &= \frac{\sum_{i=1}^N (f_{obs_i} - f_{est_i})^2}{\sum_{i=1}^N (|f_{obs_i} - \bar{f}| + |f_{est_i} - \bar{f}|)^2}, \\
 \text{mean absolute square (MAE)} &= \frac{1}{N} \sum_{i=1}^N |f_{obs_i} - f_{est_i}|, \\
 \text{mean absolute relative square (MARE)} &= \frac{1}{N} \sum_{i=1}^N \left| \frac{f_{obs_i} - f_{est_i}}{f_{obs_i}} \right|, \\
 \text{maximum absolute relative square (MXARE)} &= \max \left(\left| \frac{f_{obs_i} - f_{est_i}}{f_{obs_i}} \right| \right), \quad f \text{ or } i = 1, \dots, N \\
 \text{Nash Sutcliffe criterion (NSE)} &= \left[1 - \frac{\sum_{i=1}^N (f_{obs_i} - f_{est_i})^2}{\sum_{i=1}^N (f_{obs_i} - \bar{f})^2} \right],
 \end{aligned} \tag{8}$$

where f_{obs} is the observed infiltration rate, $f_{obs_{max}}$ and $f_{obs_{min}}$ are the maximum and minimum observed infiltration rates, f_{est} is the estimated infiltration rate at any time, and \bar{f} is the mean of the observed infiltration capacity. Nash criterion has been widely used as an indicator for goodness of fit, while its value ranges from 0.0 to 1.0. The higher values of NSE indicate a better agreement between measured and estimated data. Similarly, WI values close to unity represent the best-fitted model. However, SSE, NRMSE, MAE, MARE, and MXARE should be as low as possible for the model with highest accuracy.

2.9. Sensitivity Analysis. In a bid to determine how much the results achieved by a typical model are sensitive to each input

parameter, a sensitivity analysis can be conducted [47]. In this study, the parentage of the sensitivity analysis (SA) of the infiltration rate in respect of each input parameter (time, sand percentage, and density), which were selected based on Sihag et al.'s [28] study, is computed using [48]

$$SA_i = \frac{IR_{\max}(x_i) - IR_{\min}(x_i)}{\sum_{i=1}^N [IR_{\max}(x_i) - IR_{\min}(x_i)]} \times 100, \tag{9}$$

where $IR_{\max}(x_i)$ and $IR_{\min}(x_i)$ are the minimum and maximum infiltration rate determined by considering the variation of the input parameter (x_i) when each one of other input parameters are set as their average values. The more the SA percentage for a specific input variable, the higher the model is sensitive to that variable.

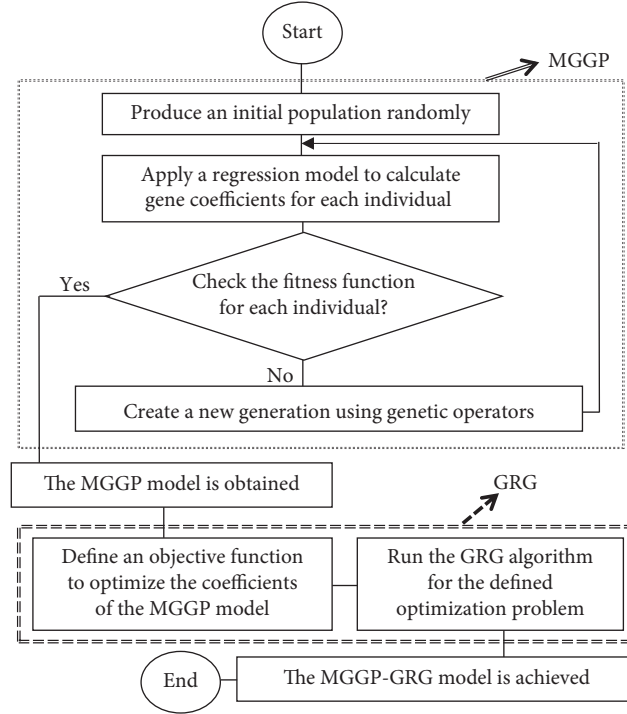


FIGURE 2: Flowchart of the hybrid MGGP-GRG for estimating infiltration rates.

2.10. Reliability Analysis. The reliability analysis is basically conducted to investigate the overall consistency of a prediction model. For this analysis, the relative error for each data point is achieved by the estimation model and compared with a threshold. Then, the number of cases, which have an equal or lower relative error than the threshold specified, is divided by the total number of points. Finally, the aforementioned ratio in the percentage would be the reliability metric, which demonstrates how reliable the prediction model performs in accordance with the desirable threshold. In this study, the reliability analysis was carried out for all methods used for predicting the infiltration rate, while the threshold was selected to be 20% based on the literature [49].

3. Results and Discussion

Accurate estimation of infiltration rate plays a vital role in various aspects of watershed hydrology. The present work focuses on improving the estimates of the infiltration rate through application of different soft computing approaches. The infiltration rates estimated by these techniques were compared with those approximated by the conventional infiltration models (Horton's model, modified Kostiakov's model, and Philip's model). In the conventional infiltration model, the observed infiltration rates and time were used as input data in accordance to model equations to obtain the estimated infiltration rate. On the other hand, in MLR, GRG, ANN, MGGP, and the hybrid MGGP-GRG models, the observed infiltration rates, time, sand percentage, and density were used as the input variables to obtain the estimated infiltration rates.

3.1. Comparison of the Conventional Infiltration Models. Table 2 presents the model parameters obtained in the training phase for the three conventional infiltration models. For the test phase, these parameters were used to estimate the infiltration rate based on equations (1), (3), and (4).

The results of different approaches considered in the present study were compared with respect to four criteria for both train and test data. This comparative analysis is shown in Table 3. In this comparison, the same data divisions were considered for all methods. The metrics used for comparing different infiltration models are given in Table 3. Based on Table 3, it may be observed that the performance of Horton's model was the worst for both training and testing parts of data. The modified Kostiakov's model improved the estimates of the infiltration rate by almost 4% and 10% as compared to those of Horton's model during training and testing, respectively. The performances of Philip's model and modified Kostiakov's model were almost comparable.

3.2. Comparison of the Conventional Models with Soft Computing Approaches. A perusal of Table 3 reveals that the technique used to model infiltration rates influences the estimates of the infiltration rate considerably. It can be observed that MLR provides the worst estimates of infiltration rates, which may depict the nonlinear nature of the infiltration process. The conventional models provide slightly better predictions of infiltration rates as compared to those obtained by MLR. Application of GRG solver further improves the estimate of infiltration as equation (7) involves a higher nonlinearity and more number of parameters as compared to equations (1)-(4). Before the application of

TABLE 2: Parameters obtained for the conventional infiltration models.

Models	Calibrated parameters		
Horton	F_c (cm/min) = 0.080	f_0 (cm/min) = 0.535	$k = 0.046$
Modified Kostiakov	$a = 1.699$	$b = 0.464$	$f_i = 0.081$
Philip	s (cm/min ^{0.5}) = 1.585	K (cm/min) = 0.061	—

TABLE 3: Comparative statistics for fit of model to the observed infiltration data.

Methods		Training phase					Testing phase				
		SSE (cm ² /min ²)	NRMSE	WI	MAE (cm/min)	NSE	SSE (cm ² /min ²)	NRMSE	WI	MAE (cm/min)	NSE
Conventional models	Horton	4.645	0.135	0.610	0.123	0.236	1.650	0.149	0.636	0.127	0.319
	Modified Kostiakov	4.485	0.133	0.613	0.118	0.263	1.474	0.141	0.686	0.119	0.392
	Philip	4.482	0.133	0.616	0.117	0.263	1.474	0.141	0.687	0.119	0.392
MLR		4.710	0.202	0.136	0.123	0.242	1.668	0.210	0.150	0.128	0.312
GRG		3.943	0.184	0.124	0.104	0.352	0.837	0.148	0.106	0.101	0.655
AI-based models	ANN	0.097	0.020	0.996	0.018	0.984	0.380	0.071	0.954	0.051	0.843
	MGGP	0.838	0.057	0.962	0.059	0.862	0.487	0.081	0.938	0.071	0.798
	MGGP-GRG	0.836	0.057	0.962	0.059	0.862	0.483	0.081	0.938	0.070	0.801

MGGP and the hybrid MGGP-GRG model, the observed infiltration rates were normalized as $i = (f_i - f_{\min}) / (f_{\max} - f_{\min})$, where i , f_{\min} , and f_{\max} are the normalized, minimum, and maximum discharges of the

i th observation. The normalized infiltration rate obtained from MGGP and the hybrid MGGP-GRG model are presented by the following equations, respectively:

$$i = \frac{0.01731 \tanh(\text{psqroot}(t))}{\cos(\exp(s))} - 3.651 \sin(\sin(s + d)) - 0.3395 \text{psqroot}\left|\frac{\text{psqroot}(t)}{s + d}\right| - \frac{0.02371 \tanh(\cos(d))}{\cos(\exp(s))} - 2.778 \cos(\sin(s + d)) + 4.863, \quad (10)$$

$$i = \frac{0.01723 \tanh(t^{0.4595})}{\cos(\exp(s))} - 3.624 \sin(\sin(s + d)) - 0.3395 \text{psqroot}\left|\frac{\text{psqroot}(t)}{s + d}\right| - \frac{0.02399 \tanh(\cos(d))}{\cos(\exp(s))} - 2.7507 \cos(\sin(s + d)) + 4.8255, \quad (11)$$

$$\text{where } \text{psqroot}(t) = \begin{cases} \sqrt{t} & t > 0 \\ 0 & t \leq 0 \end{cases}.$$

Figures 3–10 present the relative error plots obtained from different conventional infiltration models and computational techniques during training and testing. These figures also compare different methods based on MARE and MXARE for both train and test data. Although the relative error plots of the conventional infiltration models and other computational techniques (MLR, GRG, MGGP, and the hybrid MGGP-GRG) followed a similar sequence, the nature of relative error plots of ANN followed a different pattern during both training and testing. It may also be observed from Figures 3–10 that the relative errors achieved by ANN are the least as compared to others. On the other hand, relative errors obtained by Horton's infiltration model were the highest as compared to others. Furthermore, the AI-based models (ANN, MGGP, and the hybrid MGGP-GRG), which consider three independent variables (t , s , and d) instead of one variable (t), achieved much better MARE and

MXARE in comparison with the empirical models during the training and testing phases. According to Figures 3–10, ANN and the hybrid MGGP-GRG resulted in the first and second best MARE and MXARE values, whereas MLR and Horton's model yielded to the first and second worst MARE and MXARE values for the train and test data.

Figures 11 and 12 depict the comparison between the observed and estimated infiltration rates obtained by the best-fit model (ANN and the hybrid MGGP-GRG) and the worst-fit model (Horton's model). It may be observed from Figure 11 that the infiltration rates estimated by ANN almost fit the observed data during training phase. On the other hand, the infiltration rates predicted by Horton's model deviated significantly from the observed data. The performance of the hybrid MGGP-GRG was better than that of Horton's model but poorer than that of ANN. During the testing phase, the estimates of the hybrid MGGP-GRG and ANN were almost identical as shown in Figure 12. The

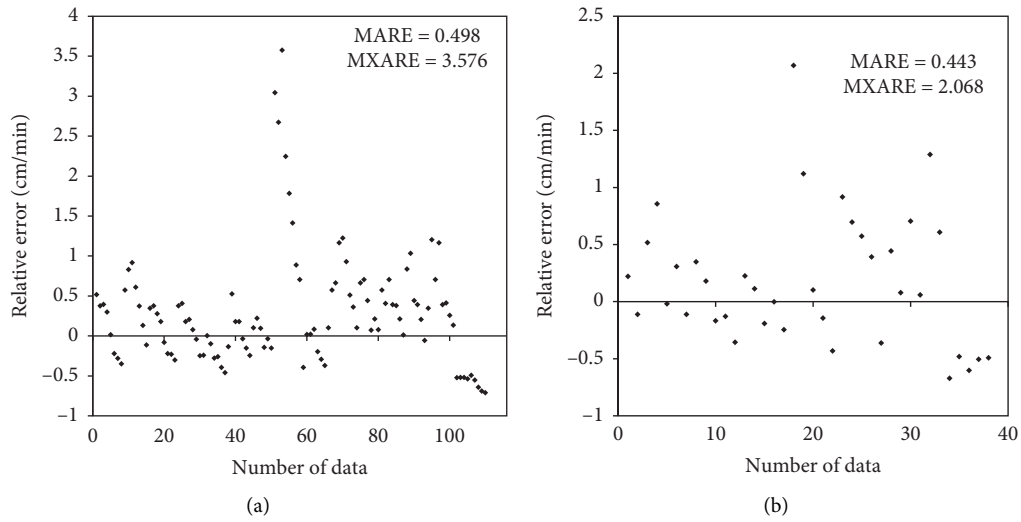


FIGURE 3: Relative error plots for Horton's model for the training and testing data.

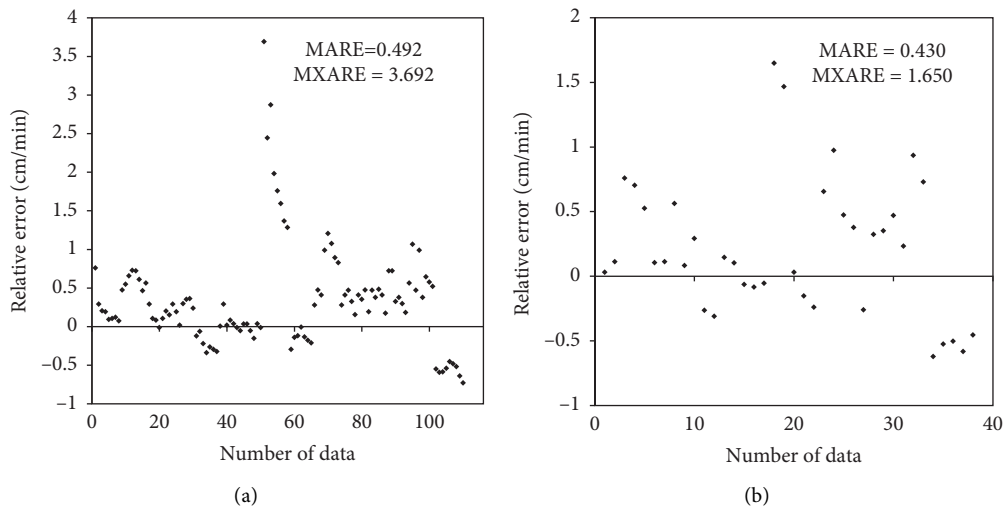


FIGURE 4: Relative error plots for modified Kostiakov's model for the training and testing data.

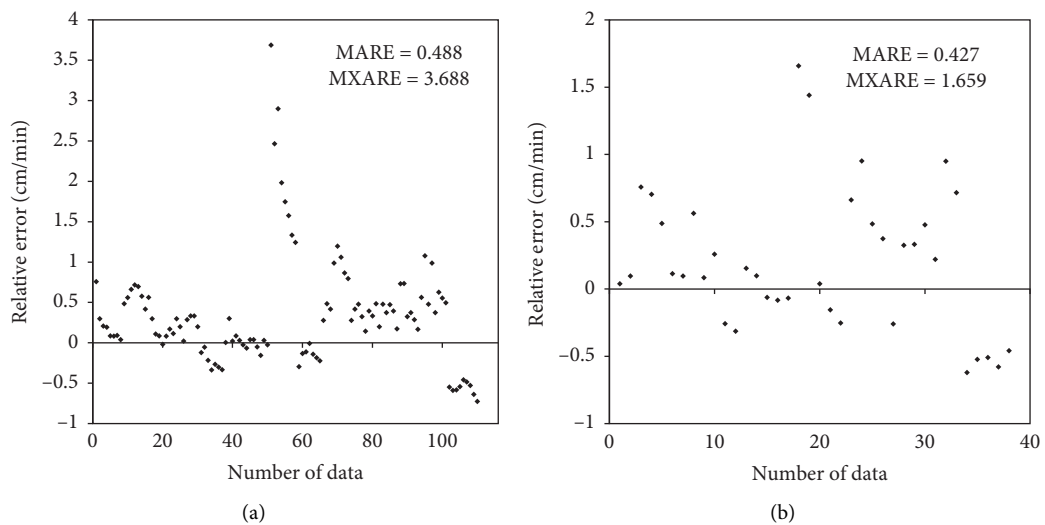


FIGURE 5: Relative error plots for Philip's model for the training and testing data.

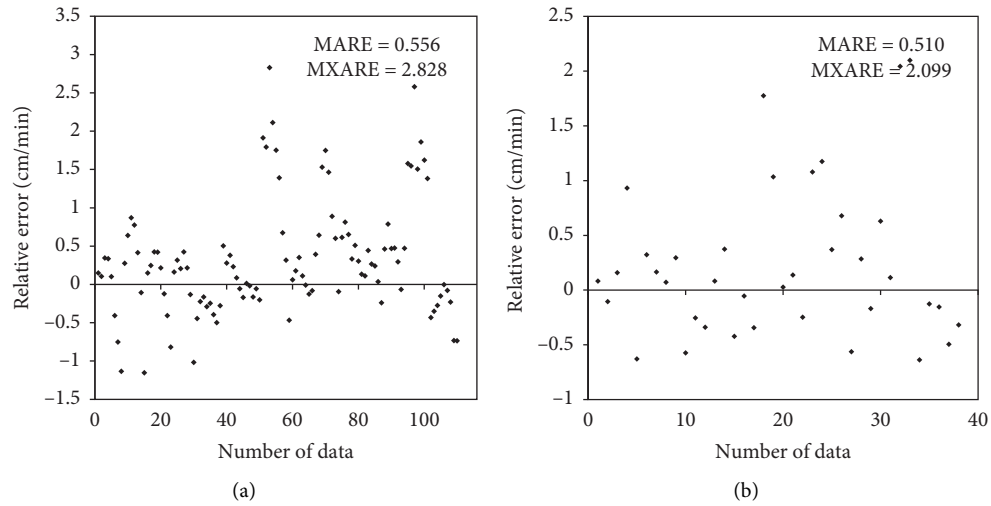


FIGURE 6: Relative error plots for the MLR model for the training and testing data.

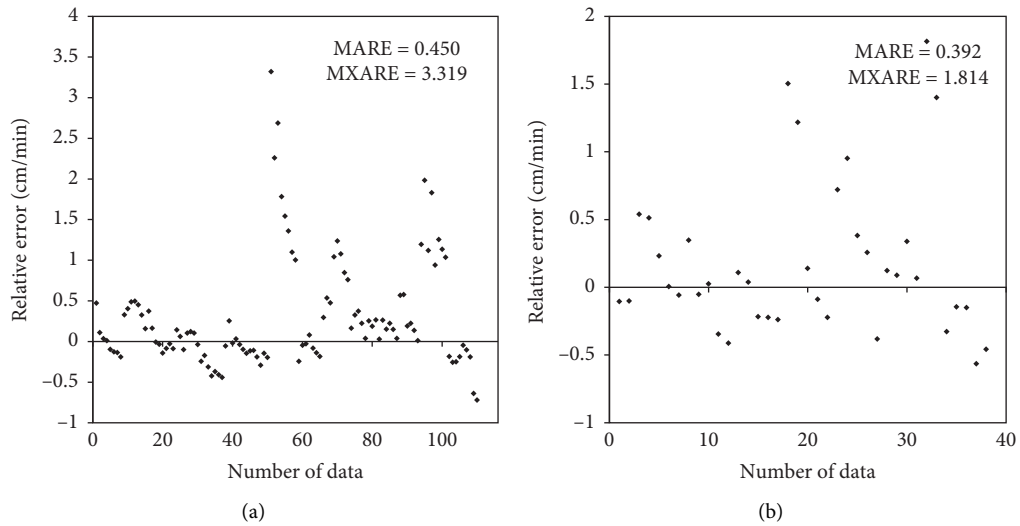


FIGURE 7: Relative Error plots for the GRG model for the training and testing data.

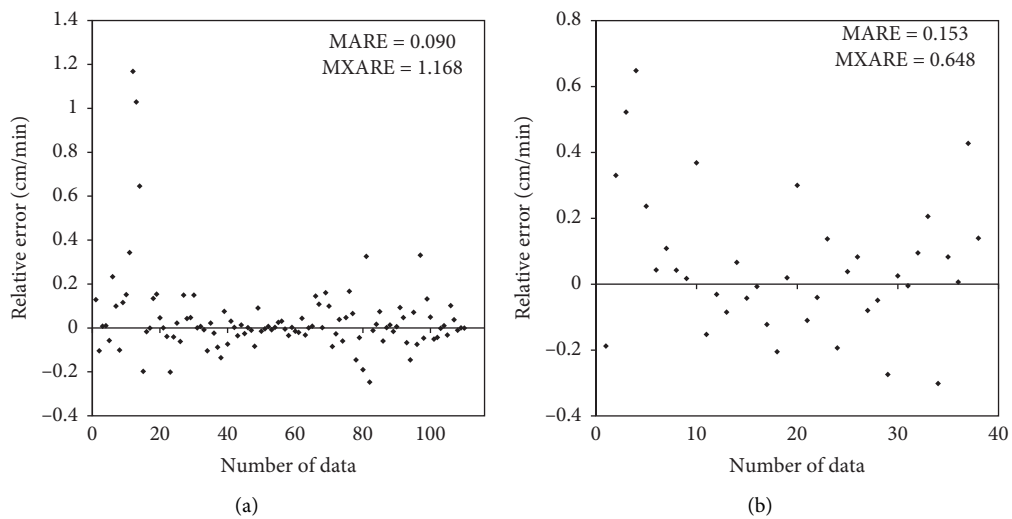


FIGURE 8: Relative error plots for the ANN model for the training and testing data.

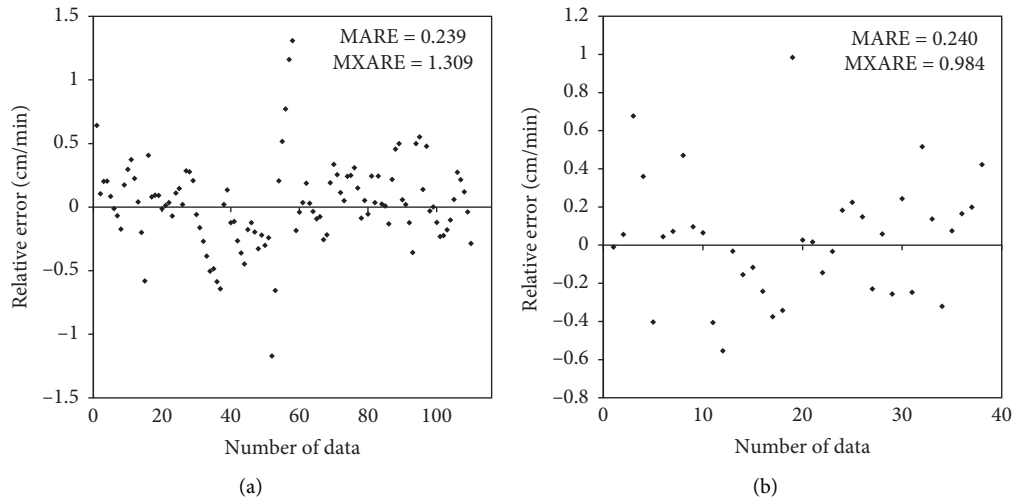


FIGURE 9: Relative error plots for the MGGP model for the training and testing data.

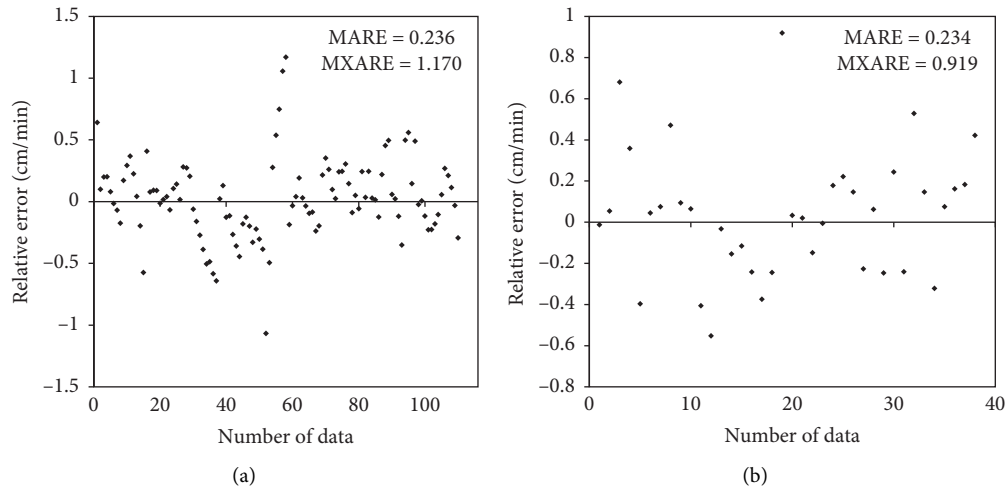


FIGURE 10: Relative error plots for the hybrid MGGP-GRG model for the training and testing data.

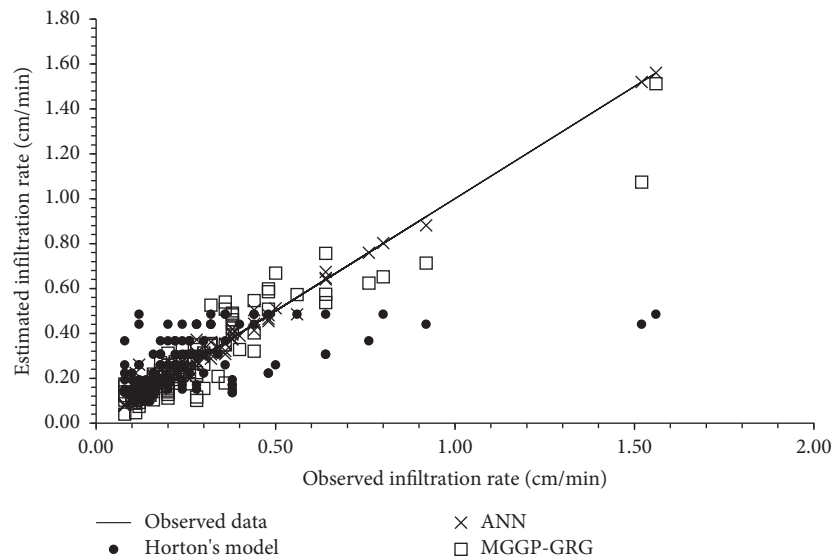


FIGURE 11: Estimated versus observed infiltration rates during the training phase.

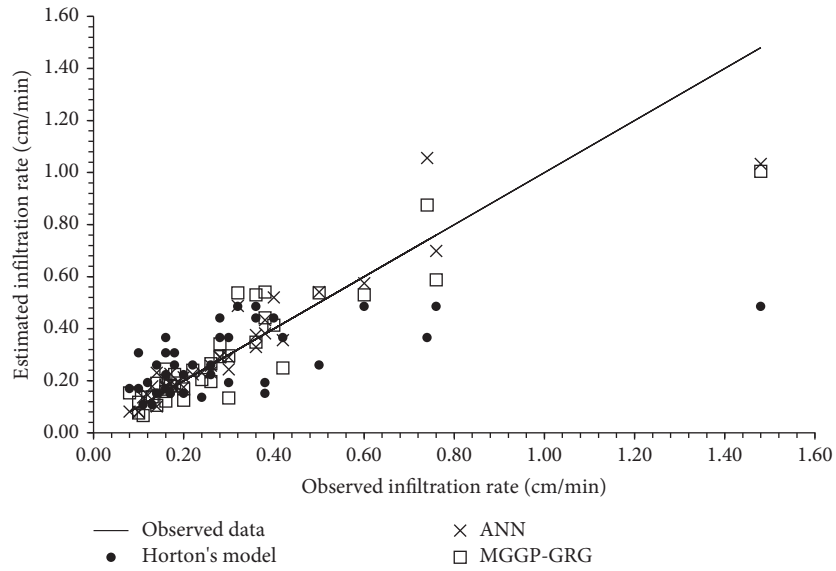
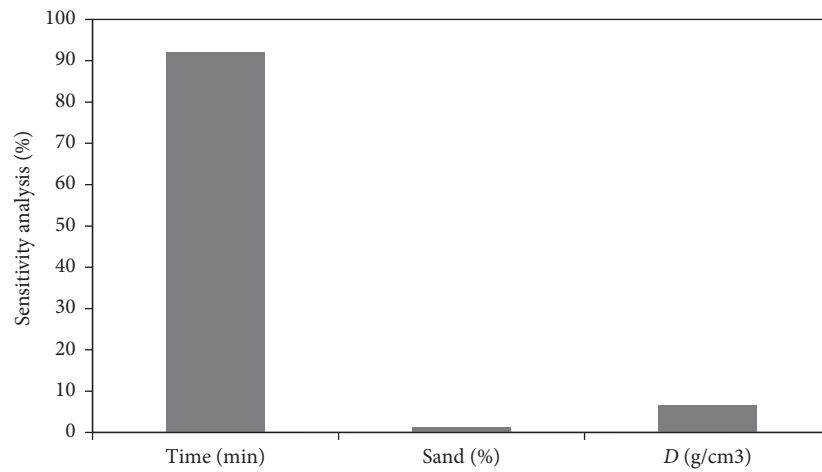
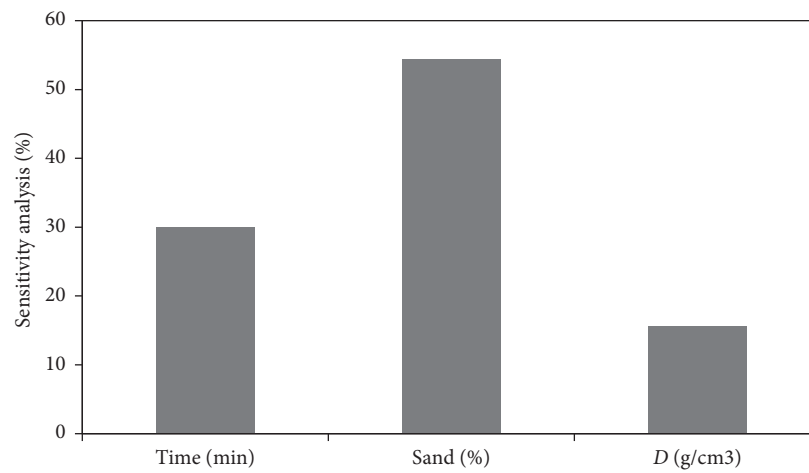


FIGURE 12: Estimated versus observed infiltration rates during the testing phase.



(a)



(b)

FIGURE 13: Results of the sensitivity analysis based on (a) ANN and (b) MGGP.

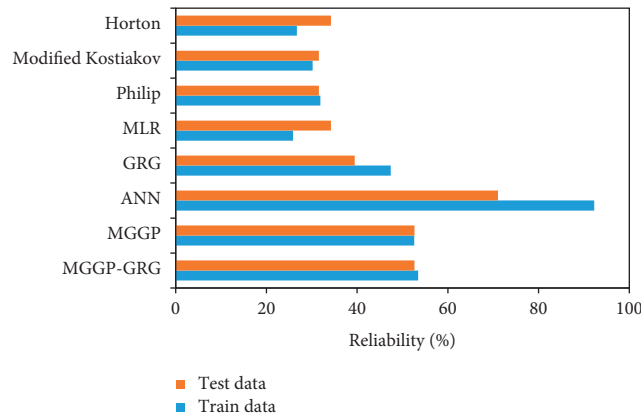


FIGURE 14: Results of the reliability analysis for train and test data.

estimates obtained by Horton's model during the testing phase were again significantly different from the corresponding observed values. Hence, Figures 11 and 12 obviously demonstrate how much the infiltration estimations can be enhanced by considering other variables involved in the process in addition to time, while they clearly indicate the better performances of AI-based models in comparison with those of the available empirical equations.

Figure 13 depicts the results of the sensitivity analysis, which was conducted for ANN and MGGP. As shown, time has the highest SA percentage ($SA = 91.98\%$) for ANN, which implies that the infiltration rates predicted by ANN are mostly sensitive to time in comparison with other two input variables (sand percentage and density). This achievement is in agreement with the fact that the empirical models (such as Horton's and modified Kostiakov) used for estimating infiltration rates rely only on time. On the other hand, MGGP-based model, which yielded a lower accuracy for predicting infiltration rates than ANN, was found to be more sensitive to sand percentage than to time. Therefore, as infiltration rates may be affected by time based on the physical background of the problem statement, the results of the sensitivity analysis also indicate that ANN estimated infiltration rates better than the MGGP-based model.

The reliability analysis was carried out for the train and test data separately. The results of this analysis are presented in Figure 14. As shown, ANN achieved the highest percentages of reliability for both the train and test data. Furthermore, the reliability percentages obtained by MGGP and the hybrid MGGP-GRG were higher than those of empirical model, MLR, and GRG. Finally, the reliability analysis conducted in this study reveals the improvement made by the AI models over other data-driven methods available in the literature for predicting infiltration rates.

The structure of the equations developed by the conventional infiltration models, MLR and GRG, are known in advance of applying these methods. On the other hand, ANN, MGGP, and the hybrid MGGP-GRG are highly nonlinear techniques with greater degrees of freedom and complexity and, therefore, provide better estimates of the infiltration rate.

However, more precise results are obtained by ANN, MGGP, and the hybrid MGGP-GRG at the expense of higher computational efforts. These machine learning tools require a considerable number of runs, unlike the conventional models and MLR in which a single attempt is sufficient for determining the model output. Based on the comparative analysis conducted in this study, ANN certainly yielded to the best estimates of infiltration rates. However, the estimates obtained from the hybrid MGGP-GRG were also comparable, especially, for the test data. Furthermore, unlike ANN, the hybrid MGGP-GRG model provided explicit equations for predicting infiltration rates, which can be implemented in a typical hydrological modelling or preferred in practice by engineers, which may be counted as an advantage of this AI-based technique.

4. Conclusions

In the present study, published infiltration data was used to assess the performances of MGGP and the hybrid MGGP-GRG technique in modelling the infiltration rates of soil. The estimated infiltration rates were compared with those obtained by the conventional models (Horton's model, Philip's model, and modified Kostiakov's model). It was observed that application of the hybrid MGGP-GRG and MGGP improved the estimates of infiltration rates as compared to the conventional infiltration model by over 80%. On the other hand, ANN provided the best estimates of infiltration rates. In addition to the accuracy improvement, the application of ANN, MGGP, and the hybrid MGGP-GRG increased the complexity of modelling equations. Future studies may focus on the comparison of the hybrid MGGP-based models with the other machine learning approaches, while applying the explicit infiltration models developed by either MGGP or the hybrid MGGP-GRG in hydrological models is anticipated in favor of assessing their performances in practice.

Data Availability

The data used in this study are available in the related literature.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

- [1] M. Zakwan, M. Muzzammil, and J. Alam, "Application of spreadsheet to estimate infiltration parameters," *Perspectives in Science*, vol. 8, pp. 702–704, 2016.
- [2] M. Zakwan, "Comparative analysis of the novel infiltration model with other infiltration models," *Water and Environment Journal*, vol. 33, no. 4, pp. 620–632, 2019.
- [3] H. J. Morel-Seytoux, "Infiltration modeling component in a large-scale hydrologic model," *Journal of Hydraulic Research*, vol. 42, no. 6, pp. 666–667, 2004.
- [4] M. Šraj, L. Dirnbek, and M. Brilly, "The influence of effective rainfall on modeled runoff hydrograph," *Journal of Hydrology and Hydromechanics*, vol. 58, no. 1, pp. 3–14, 2010.
- [5] Y. Wen, C. Hu, G. Zhang, and S. Jian, "Response of the parameters of excess infiltration and excess storage model to land use cover change," *Journal of Hydrology and Hydromechanics*, vol. 68, no. 2, pp. 99–110, 2020.
- [6] S. K. Mishra, J. V. Tyagi, and V. P. Singh, "Comparison of infiltration models," *Hydrological Processes*, vol. 17, no. 13, pp. 2629–2652, 2003.
- [7] K. Deep and K. N. Das, "Optimization of Infiltration parameters in hydrology," *World Journal of Modelling and Simulation*, vol. 4, pp. 120–130, 2008.
- [8] A. Haghiabi, J. Abedi-koupai, M. Heidarpour, and J. Mohammadzadeh-habili, "A new method for estimating parameters of Kostiakov and Modified Kostiakov infiltration equations," *World Applied Science Journal*, vol. 15, pp. 129–135, 2011.
- [9] M. Zakwan, "Assessment of dimensionless form of Kostiakov model," *Aquademia: Water, Environment and. Technology*, vol. 1, no. 1, p. 01, 2017.
- [10] L. Chen, L. Xiang, M. H. Young, J. Yin, Z. Yu, and M. T. v. Genuchten, "Optimal parameters for the Green-Ampt infiltration model under rainfall conditions," *Journal of Hydrology and Hydromechanics*, vol. 63, no. 2, pp. 93–101, 2015.
- [11] G. Tayfur, "Modern optimization methods in water resources planning, engineering and management," *Water Resources Management*, vol. 31, no. 10, pp. 3205–3233, 2017.
- [12] A. D. Mehr, V. Nourani, E. Kahya, B. Hrnjica, A. M. Sattar, and Z. M. Yaseen, "Genetic programming in water resources engineering: a state-of-the-art review," *Journal of Hydrology*, vol. 566, pp. 643–667, 2018.
- [13] C. Shen, "A transdisciplinary review of deep learning research and its relevance for water resources scientists," *Water Resources Research*, vol. 54, no. 11, pp. 8558–8593, 2018.
- [14] H. Tyralis, G. Papacharalampous, and A. Langousis, "A brief review of random forests for water scientists and practitioners and their recent history in water resources," *Water*, vol. 11, no. 5, p. 910, 2019.
- [15] S. Kambalimath and P. C. Deka, "A basic review of fuzzy logic applications in hydrology and water resources," *Applied Water Science*, vol. 10, no. 8, pp. 1–14, 2020.
- [16] M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," *Water Science and Technology*, vol. 82, no. 12, pp. 2635–2670, 2020.
- [17] M. Pandey, M. Zakwan, M. A. Khan, and S. Bhawe, "Development of scour around a circular pier and its modelling using genetic algorithm," *Water Supply*, vol. 20, no. 8, pp. 3358–3367, 2020.
- [18] M. Najafzadeh and G. Oliveto, "Riprap incipient motion for overtopping flows with machine learning models," *Journal of Hydroinformatics*, vol. 22, no. 4, pp. 749–767, 2020.
- [19] M. Najafzadeh and G. Oliveto, "More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency," *Soft Computing*, vol. 25, pp. 5723–5746, 2021.
- [20] M. Niazkar and M. Zakwan, "Assessment of artificial intelligence models for developing single-value and loop rating curves," *Complexity*, vol. 2021, Article ID 6627011, 21 pages, 2021.
- [21] N. Niazkar, N. Talebbeydokhti, and S. H. Afzali, "One dimensional hydraulic flow routing incorporating a variable grain roughness coefficient," *Water Resources Management*, vol. 33, no. 13, pp. 4599–4620, 2019.
- [22] M. Kumar and P. Sihag, "Assessment of infiltration rate of soil using empirical and machine learning-based models," *Irrigation and Drainage*, vol. 68, no. 3, pp. 588–601, 2019.
- [23] A. Z. Dewidar, H. Al-Ghobari, and A. Alataway, "Developing a fuzzy logic model for predicting soil infiltration rate based on soil texture properties," *Water SA*, vol. 45, no. 3, pp. 400–410, 2019.
- [24] G. T. Patle, T. T. Sikar, K. S. Rawat, and S. K. Singh, "Estimation of infiltration rate from soil properties using regression model for cultivated land," *Geology, Ecology, and Landscapes*, vol. 3, no. 1, pp. 1–13, 2019.
- [25] P. Sihag, N. K. Tiwari, and S. Ranjan, "Support vector regression-based modeling of cumulative infiltration of sandy soil," *ISH Journal of Hydraulic Engineering*, vol. 26, no. 1, pp. 44–50, 2020.
- [26] M. R. Pahlavan-Rad, K. Dahmardeh, M. Hadizadeh et al., "Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran," *Catena*, vol. 194, Article ID 104715, 2020.
- [27] A. Sepahvand, B. Singh, M. Ghobadi, and P. Sihag, "Estimation of infiltration rate using data-driven models," *Arabian Journal of Geosciences*, vol. 14, no. 1, pp. 1–11, 2021.
- [28] P. Sihag, V. P. Singh, A. Angelaki, V. Kumar, A. Sepahvand, and E. Golia, "Modelling of infiltration using artificial intelligence techniques in semi-arid Iran," *Hydrological Sciences Journal*, vol. 64, no. 13, pp. 1647–1658, 2019.
- [29] R. I. Horton, "The interpretation and application of runoff plot experiments with reference to soil erosion problems," *Soil Science Society of America Proceedings*, vol. 3, pp. 340–349, 1938.
- [30] J. R. Philip, "Theory of infiltration," *Soil Science*, vol. 83, pp. 345–357, 1957.
- [31] A. N. Kostiakov, "On the dynamics of the coefficients of water percolation in soils," in *Sixth Commission*, pp. 15–21, International Society of Soil Science, Part A, Vienna, AT, USA, 1932.
- [32] R. E. Smith, "The infiltration envelope: results from a theoretical infiltrometer," *Journal of Hydrology*, vol. 17, pp. 1–21, 1972.
- [33] M. Pandey, M. Zakwan, P. K. Sharma, and Z. Ahmad, "Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment," *ISH Journal of Hydraulic Engineering*, pp. 1–8, 2018.
- [34] M. Zakwan, M. Muzzammil, and J. Alam, "Estimation of soil properties using infiltration data," in *Proceedings of the*

- National Conference of Advanced Geotechnological Engineering*, pp. 198–201, Aligarh, India, December 2016.
- [35] M. Muzzammil, J. Alam, and M. Zakwan, “A spreadsheet approach for prediction of rating curve parameters,” in *Hydrologic Modeling. Water Science and Technology Library*, V. Singh, S. Yadav, and R. Yadava, Eds., Vol. 81, Springer, Singapore, 2018.
 - [36] A. R. Nawaz, M. Zakwan, I. Khan, and Z. A. Rahim, “Comparative analysis of variants of Muskingum model,” *Water and Energy International*, vol. 63, no. 7, pp. 64–73, 2020.
 - [37] M. Niazkar and S. H. Afzali, “Developing a new accuracy-improved model for estimating scour depth around piers using a hybrid method,” *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, no. 2, pp. 179–189, 2018.
 - [38] M. Niazkar, N. Talebbeydokhti, and S.-H. Afzali, “Bridge backwater estimation: a Comparison between artificial intelligence models and explicit equations,” *Scientia Iranica*, vol. 28, pp. 573–585, 2020.
 - [39] M. Niazkar, F. Hajizadeh mishi, and G. Eryılmaz Türkkkan, “Assessment of artificial intelligence models for estimating lengths of gradually varied flow profiles,” *Complexity*, vol. 2021, Article ID 5547889, 11 pages, 2021.
 - [40] M. Niazkar, “Revisiting the estimation of colebrook friction factor: a comparison between artificial intelligence models and C-W based explicit equations,” *KSCE Journal of Civil Engineering*, vol. 23, no. 10, pp. 4311–4326, 2019.
 - [41] M. Niazkar, “Assessment of artificial intelligence models for calculating optimum properties of lined channels,” *Journal of Hydroinformatics*, vol. 22, no. 5, pp. 1410–1423, 2020.
 - [42] M. Niazkar, N. Talebbeydokhti, and S. H. Afzali, “Novel grain and form roughness estimator scheme incorporating artificial intelligence models,” *Water Resources Management*, vol. 33, no. 2, pp. 757–773, 2019.
 - [43] H. R. Niazkar and M. Niazkar, “COVID-19 outbreak: application of multi-gene genetic programming to country-based prediction models,” *Electronic Journal of General Medicine*, vol. 17, no. 5, p. em247, 2020.
 - [44] D. Searson, *GPTIPS: Genetic Programming and Symbolic Regression for MATLAB*, https://www.researchgate.net/publication/277297934_GPTIPS_Genetic_Programming_Symbolic_Regression_for_MATLAB_User_Guide, 2009.
 - [45] J. S. Lee and K. D. Suh, “Development of stability formulas for rock armor and tetrapods using multigene genetic programming,” *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 146, no. 1, Article ID 04019027, 2020.
 - [46] H. R. Niazkar and M. Niazkar, “Application of artificial neural networks to predict the COVID-19 outbreak,” *Global Health Research and Policy*, vol. 5, no. 1, p. 50, 2020.
 - [47] M. Niazkar and S. H. Afzali, “Assessment of modified honey bee mating optimization for parameter estimation of non-linear muskingum models,” *Journal of Hydrologic Engineering*, vol. 20, no. 4, Article ID 04014055, 2015.
 - [48] A. Garg, K. Tai, V. Vijayaraghavan, and P. M. Singru, “Mathematical modelling of burr height of the drilling process using a statistical-based multi-gene genetic programming approach,” *The International Journal of Advanced Manufacturing Technology*, vol. 73, no. 1–4, pp. 113–126, 2014.
 - [49] F. Saberi-Movahed, M. Najafzadeh, and A. Mehrpooya, “Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions,” *Water Resources Management*, vol. 34, no. 2, pp. 529–561, 2020.