Water Science & Technology

Water Science & Technology Vol 84 No 9, 2366 doi: 10.2166/wst.2021.444

Rain garden infiltration rate modeling using gradient boosting machine and deep learning techniques

Sandeep Kumar and K. K. Singh *

Department of Civil Engineering, NIT Kurukshetra, Kurukshetra, India
*Corresponding author. E-mail: kksinghunitech@gmail.com; kksingh@nitkkr.ac.in

KKS, 0000-0003-4374-7713

ABSTRACT

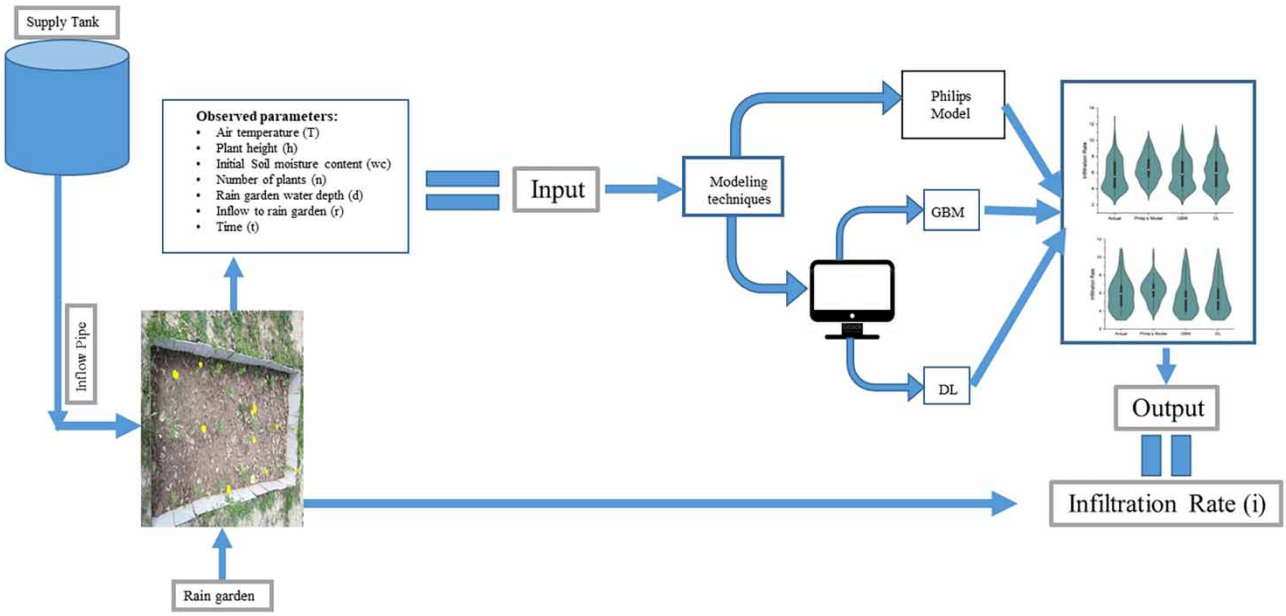
Rain garden is effective in reducing storm water runoff, whose efficiency depends upon several parameters such as soil type, vegetation and meteorological factors. Evaluation of rain gardens has been done by various researchers. However, knowledge for sound design of rain gardens is still very limited, particularly the accurate modeling of infiltration rate and how much it differs from infiltration of natural ground surface. The present study uses experimentally observed infiltration rate of rain gardens with different types of vegetation (grass, candytuft, marigold and daisy with different plant densities) and flow conditions. After that, modeling has been done by the popular infiltration model i.e. Philip's model (which is valid for natural ground surface) and soft computing tools viz. Gradient Boosting Machine (GBM) and Deep Learning (DL). Results suggest a promising performance (in terms of CC, RMSE, MAE, MSE and NSE) by GBM and DL in comparison to the relation proposed by Philip's model (1957). Most of the values predicted by both GBM and DL are within scatter limits of $\pm 5\%$, whereas the values by Philips model are within the range of $\pm 25\%$ error lines and even outside. GBM performs better than DL as the values of the correlation coefficients and Nash-Sutcliffe model efficiency (NSE) coefficient are the highest and the root mean square error is the lowest. The results of the study will be useful in selection of plant type and its density in the rain garden of the urban area.

Key words: deep learning, gradient boosting machine, Philip's model, rain garden, simulation of infiltration rate

HIGHLIGHTS

- Performance of rain garden for different vegetation has been studied.
- Soft computing techniques has been used to analyze the performance of rain garden.
- Results suggest that soft computing techniques model better than existing conventional models.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The world's population has increased rapidly in the last few decades, which results in more urbanization and demand for infrastructure (Panagopoulos 2019). Urbanization has posed numerous challenges for the planners, policy makers and administrators, among which urban flooding is the most significant challenge (NIUA 2016). Factors responsible for urban flooding are hydrological, meteorological and human (Nguyen *et al.* 2019). The key causes of the urban flooding are inappropriate planning, inadequate natural drainage, extreme climate events, choking of drainage system, and developments in the flood plains (Rafiq *et al.* 2016).

Separate sewerage system is commonly used for draining the storm water in urban area which has design period of 20–30 years only. However, the impervious surfaces of the city continue to increase with time leading to increase in runoff volume. Thus, the existing storm water sewer becomes inadequate to drain the water conveniently. Improving the capacity of the existing system is highly uneconomical. The storm water management practice can be viable alternative. There are various best management practices (BMPs) which can mitigate the menace of flooding such as permeable paving, swales (vegetated ditches), bio-retention, detention ponds, wetlands, infiltration surfaces, media filters, green-roofs and rain garden (Fletcher *et al.* 2015). Among these, rain garden is emerging practice, whose infiltration rate is more than natural ground surface (Aaron *et al.* 2012).

Rain garden can reduce the impacts of urban flooding by reducing the quantity of runoff and improving the quality of water by bio-retention (Osheen & Singh 2019). Basically, it is a management technique designed to control storm water runoff from its source in a simple and natural way (Bhandari *et al.* 2018; Osheen & Singh 2020). Rain gardens are effective in terms of reducing pollutants, reducing storm water runoff and attenuating peak runoff rate (Muerdter *et al.* 2016; Shafique 2016; Skorobogatov *et al.* 2020; Tirpak *et al.* 2021). These can effectively remove 75–80% of sediments and 80–90% of chemicals and nutrients from the runoff water (Davis *et al.* 2006; Aaron *et al.* 2012). As compared to other usual lawns, rain gardens permit around 30% more water to infiltrate into the ground (Malaviya *et al.* 2019). It can be used in a wide variety of environments (Weerasundara *et al.* 2016). The observed peak runoff time is also delayed by significant factor by rain gardens. Several authors have supported the overall effectiveness of rain gardens in reducing peak runoff when compared with the inflow hydrograph peak and control of urban runoff (Li *et al.* 2016; Shuster *et al.* 2017; Yuan *et al.* 2017; Zhang *et al.* 2019). Efficiency of rain garden depends upon several factors such as soil media, vegetation, runoff rate etc. Wide range of recommendations have been made to relate soil content with the performance of rain gardens as permeability of the planting mix (soil media) varies with change in soil content (Zhang *et al.* 2018). The investigation of the effect of the percentage of various soil components, including sand, topsoil and mulch (compost), on the pollutant removal efficiency and infiltration rate has been done in the laboratory. Clay greatly reduces

permeability and enhances filtration and absorption of pollutants (Muerdter *et al.* 2016; Yang *et al.* 2020). Ponding and drawdown of water in rain garden depends on the infiltration capabilities of the soil characteristics of the planting medium. Guidelines suggest that ponding depth of 15–30 cm and drawdown time of 24–72 h are suitable. Drawdown is a key variable in the determination of surface area of bio retention cells. The larger surface area has shorter drawdown time (Mohammed *et al.* 2019). Due to growth of plant roots with time, they penetrate deeper into the soil. These lead to lengthening of soil pits, reversing the soil compaction and the formation of large pores (Yuan *et al.* 2017; Muerdter *et al.* 2018; Skorobogatov *et al.* 2020). The existing infiltration models (Sihag *et al.* 2020) are not able to describe rain garden infiltration rate accurately.

Regression analysis on predicting soil infiltration rate for cultivated land was carried out Patle *et al.* (2019). Worland *et al.* (2018) compared the performance of various machine learning models and base line model to predict the mean stream flow in various districts of USA. Kumar & Sihag (2019) proposed the application of machine learning for predicting the infiltration rate through different soil media. Singh *et al.* (2019) employed regression techniques for predicting effect of water quality on infiltration rate. Angelaki *et al.* (2021) compared the performance of ANN, SVM and FUZZY for estimating the infiltration rate through soil. But there is no literature available that modeled the rain garden infiltrations rate by using soft computing techniques. So, the aim of the present study is to model the rain garden infiltration rate by using soft commuting tools i.e. Deep Learning and Gradient Boosting Machine.

In the present work, the infiltration rate has been studied experimentally by using different kinds of vegetation planted in a rain garden. The various models like Philip's model, GBM and DL have been applied over the observed values of infiltration. The comparative study of these different models will be helpful on the applicability of the machine learning techniques over the prediction of infiltration rate for rain garden which is absent in the literature. At last, the extent of applicability of the Philips model with the results obtained using soft computing is also carried out.

2. MATERIALS AND METHODS

2.1. Modeling techniques

In the present study, modeling of rain garden infiltration rate has been explored using conventional method and machine learning techniques viz. Gradient Boosting Machine (GBM) and Deep Learning (DL). Relevant detail of the conventional method has been provided in the subsequent section. For details of GBM and DL techniques, the readers are advised to refer to internet/reference books.

2.1.1. Conventional model

There are many infiltration models available in the literature e.g. Green-Ampt equation (1911), Kostiakov equation (1932), Horton's equation (1933) and Philip's equation (1957) (Subramanya 2020). Survey of previously published work suggest Philips equation is widely accepted to predict infiltration rate.

2.1.1.1. Philips' Model (1957). The Philip's model (Philip 1957) is a two parameter model (Equation (1)) and its parameters can be estimated using soil-water characteristic curves.

Philip's two-term model relates 'i' (infiltration rate) to 't' (time) as

$$i = 0.5 \cdot S \cdot t^{-0.5} + K \tag{1}$$

where S = a function of soil suction potential (sorptivity) (cm/hr$^{-0.5}$) and K = Darcy's hydraulic conductivity (cm/hr). The values S and K can be obtained by plotting 'i' against '$t^{-0.5}$' on arithmetic graph paper. The best fit straight line through plotted points gives value of K as intercept and the slope of the line gives value of 0.5 S.

2.2. Experimental setup and methodology

Four experimental rain gardens were created in the Hydraulics Laboratory of the Civil Engineering Department, National Institute of Technology, Kurukshetra, India. The two rain gardens were of size 2 m × 1 m × 0.15 m (named as RG 1 and RG 2 having higher density of plantation) and the other two (RG 3 and RG 4) were of size 1 m × 1 m × 0.05 m. All four gardens had flat bed profiles. The sides of these rain gardens were enclosed with acrylic sheets to retain the side face of soil in position and prevent lateral infiltration. Particle size analysis and Atterberg limits of the rain gardens soil were performed to understand the type of soil of the rain garden. The garden soil contains sand (75–425 μ) and clay ($\leq$2 μ) 55.41% and 44.59%,

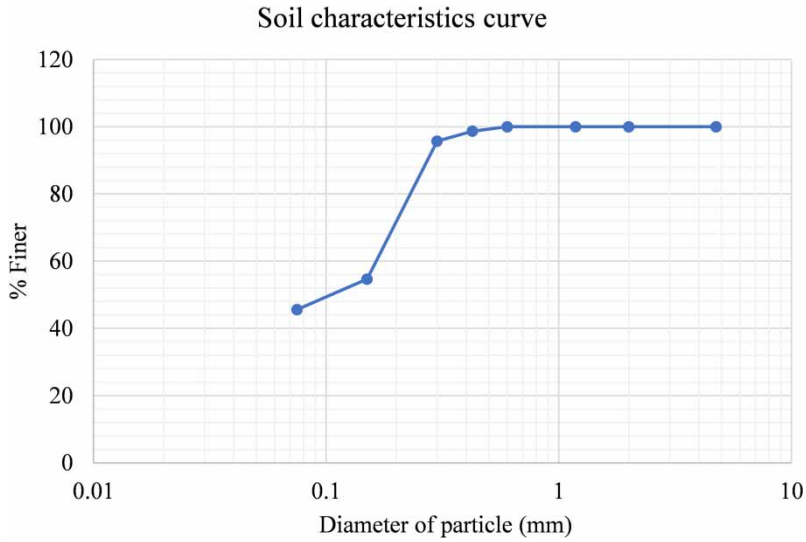


Figure 1 | Soil characteristic curve.

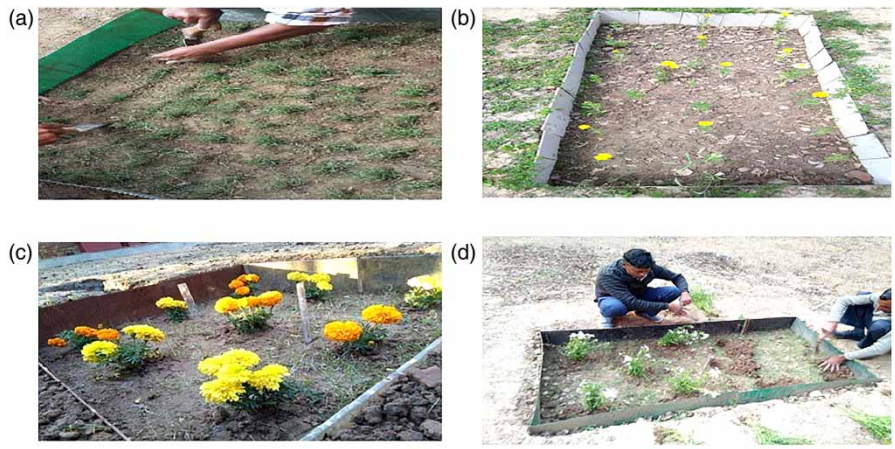


Figure 2 | Photographic view of planting in rain garden (a) scutch grass (b) candytuft flowers (c) marigold flowers (d) daisy flowers.

respectively shown in Figure 1. Thus, the soil in rain gardens is sandy loam of brown colour. Atterberg limits viz. liquid limit, plastic limit and plasticity index, were 24.50%, 16.00% and 8.50% respectively.

For experimental observations, four types of vegetation have been planted in the rain garden. First scutch grass (Cynodon dactylon) was planted in all the four rain gardens as shown in Figure 2(a). Scutch grass is a local vegetation of Kurukshetra region which requires very little irrigation and is most suitable for rain gardens. After this chandani flower (candytuft) plants are planted in RG 1 and RG 2 (Figure 2(b)) and marigold flower plants in RG 3 and RG 4 (Figure 2(c)). Daisy flower plants were grown in RG 1, RG 2, RG 3 and RG 4 (Figure 2(d)). Table 1 shows the number of plants planted in the different RGs.

A cylindrical water storage tank of capacity 170 liters was provided near the rain gardens. Inflow to rain garden was provided through a polyvinyl chloride (PVC) pipe connected to the storage tank. Same volume of water (i.e. 100 liters), in form of inflow was applied to RG1 and RG2. To study the effect of overflow on infiltration rates, in RG 3 a pipe outlet at a height of 5 cm above the raingarden bed was provided. Time elapsed from the start to end of infiltration was noted down. The infiltration rate observations were taken for different plant heights. The overflow was collected in the container from the outlet pipe. Experiments were performed with average inflow rate of 7 lit/min and 5.4 lit/min. Table 2 provides range of various data collected from experimental observations.

Table 1 | Details of rain gardens and vegetation grown

S.No.	Type of vegetation	RG Name	Size	No. of plants
1	Scutch Grass (*Cynodon dactylon*)	RG 1	2 m × 1 m × 0.15 m	300
2		RG 2	2 m × 1 m × 0.15 m	240
3		RG 3	1 m × 1 m × 0.05 m	148
4		RG 4	1 m × 1 m × 0.05 m	140
5	Candytuft flower	RG 1	2 m × 1 m × 0.15 m	220
6		RG 2	2 m × 1 m × 0.15 m	110
7	Marigold flower	RG 3	1 m × 1 m × 0.05 m	9
8		RG 4	1 m × 1 m × 0.05 m	9
9	Daisy flower	RG 1	2 m × 1 m × 0.15 m	24
10		RG 2	2 m × 1 m × 0.15 m	12
11		RG 3	1 m × 1 m × 0.05 m	12
12		RG 4	1 m × 1 m × 0.05 m	12

Table 2 | Range of various data collected from experimental observations

S.No.	Variables	Min	Max	Mean	SD
1	Air temperature (0°C)	11	34	20.41	7.29
2	Rain garden soil water content	4.36	15.1	10.27	3.40
3	Average inflow rate to RG (lit/min)	5.40	7	6.33	0.79
4	Plant height (cm)	3	33	16.81	13.42
5	Number of plants	0	300	75.38	92.80
6	Water depths in RG (cm)	0	5	1.59	1.37
7	Time elapsed in drop of water depth (sec)	366	3,752	1,147.02	693.08
8	Infiltration rate (cm/hr)	0	12.41	5.03	2.65

3. MODEL PERFORMANCE

Performance of GBM and DL were evaluated by statistical parameters i.e. correlation coefficients (CC), root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE) and Nash-Sutcliffe model efficiency coefficient. The relationships used in study are given below:

- **Coefficient of correlation (CC):** It can be calculated as:

$$CC = \frac{N\sum_{i=1}^{N} c_i e_i - \left(\sum_{i=1}^{N} c_i\right)\left(\sum_{i=1}^{N} e_i\right)}{\sqrt{\left[N\left(\sum_{i=1}^{N} c_i^2\right) - \left(\sum_{i=1}^{N} c_i\right)^2\right]\left[N\left(\sum_{i=1}^{N} e_i^2\right) - \left(\sum_{i=1}^{N} e_i\right)^2\right]}} \quad (2)$$

- **Root mean square error (RMSE):** The equation of the RMSE is:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - e_i)^2} \quad (3)$$

- **Mean Absolute Error (MAE):** The formula for MAE:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|c_i - e_i| \tag{4}$$

- **Mean Squared Error (MSE):** The mean squared error (MSE) is given by the following relation:

$$\mathrm{MSE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{c_i} - \mathrm{e_i})^2 \tag{5}$$

The smaller the MSE, the better the model's performance.

- **Nash-Sutcliffe model efficiency coefficient (η):** The Nash-Sutcliffe efficiency (NSE), NSE can be computed as:

$$\eta = 1 - \frac{\sum_{i=1}^{N}(c_i - e_i)^2}{\sum_{i=1}^{N}(e_i - \bar{c}_i)^2} \tag{6}$$

Where $N =$ number of observations of infiltration rate; $\mathrm{c_i} =$ observed infiltration rate; $\mathrm{e_i} =$ predicated infiltration rate and $\bar{c}_\mathrm{i} =$ average infiltration rate

3.1. Data set

A total of 216 observations were used for modeling. The input data set consists of air temperature (t), plant height (h), initial soil moisture content (wc), number of plants (n), water depth (d), inflow to rain garden (r), time (T), and output data is average infiltration rate (i). 75% of data were used for training and the remaining 25% of data were used for testing.

4. RESULTS AND DISCUSSION

4.1. Prediction of infiltration rate using Philip's Model (1957)

The values of S and K were obtained by plotting 'i' against '$t^{-0.5}$' on arithmetic graph paper as 5.31 m·sec$^{-1/2}$ and 1.13 cm/hr, respectively. Philip's model infiltration rates have been compared with observed experimental infiltration rates. Agreement diagram between Philip's model and observed infiltration rate are shown in Figure 3. The plot shows three straight lines passing through the origin. The central straight line corresponds to the perfect agreement line at 45° with the x-axis, which means there is a perfect match between the model calculated value and observed value. The other two upper and lower straight lines are corresponding to +25% and −25% error lines, which mean overestimation and underestimation by the model.

It is seen from Figure 3 that there are wide variations in infiltration rates obtained by Philip's model. Most of the predicted infiltration rates are lying within the ±25% error line, but some are beyond the error lines also. There is mostly under prediction by Philips's model. This might be due to the presence of lesser amount of voids in soil under natural conditions as compared to the voids observed in a rain garden, which is created artificially. This warrants improving the relationships. Hence, machine learning techniques are tried in order to have a better prediction.

4.2. Prediction of infiltration rate using GBM

A fully open source H2O software, distributed in-memory machine learning platform with linear scalability has been used for modeling. Performance of GBM depends on the input parameters for training the model and the primary parameters are given in Table 3. Figure 4 shows the variation of training and testing data set in terms of deviance with respect to number of trees required. It is observed that the deviance in both data set becomes almost asymptomatic as the number of trees reaches 30. The modeling has been carried out for 51 trees.

Figure 5 shows the agreement diagram between predicted and observed infiltration rate. It can be seen that model predicted infiltration rate falls on or very near to the 1:1 slope line, which demonstrates very good modeling.

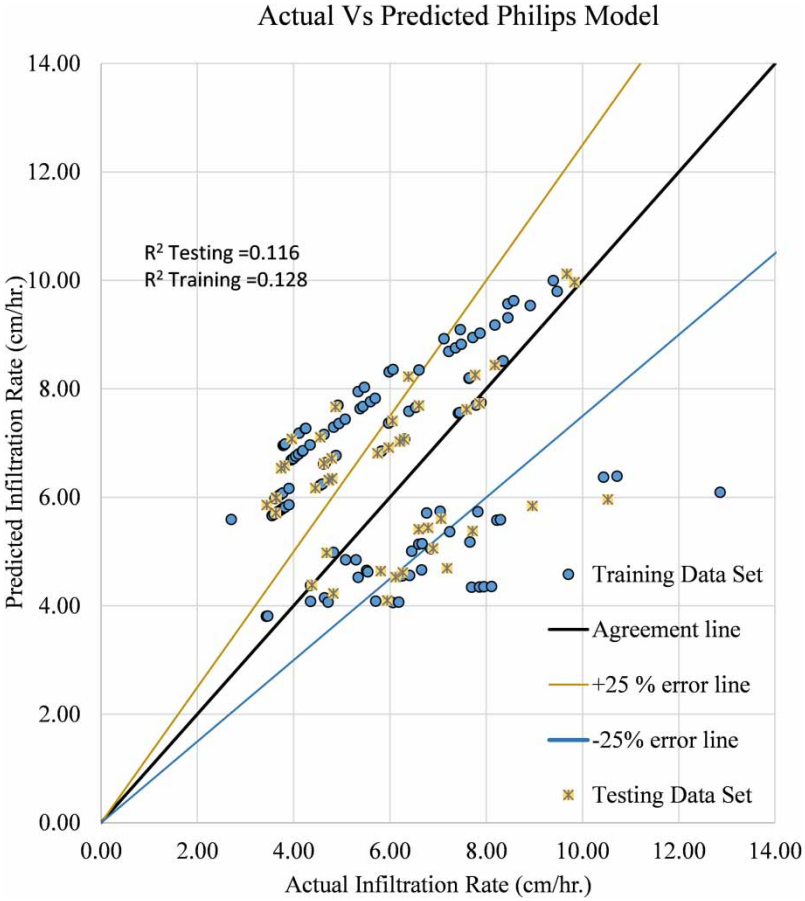


Figure 3 | Agreement diagram between Philip's model (Philip 1957) and observed infiltration rate.

Table 3 | Values/type of model parameters used in GBM

S.No.	GBM parameter name	Value/Type
1	N folds	5
2	Score tree interval	5
3	Fold assignment	Modulo
4	Response Column (output parameter)	I
5	N trees	51
6	Max depth	14
7	Min rows	1
8	Distribution	Gaussian
9	Sample rate	0.8
10	Histogram type	Uniform Adaptive

4.3. Prediction using deep learning

In the deep learning technique, the first step is to identify the number of epochs required for predicting the values with good accuracy along with the minimum computational cost. The primary parameter is given in Table 4. The variation in the deviation of the results for training and testing data sets with respect to the epochs is shown in Figure 6. It is observed that when

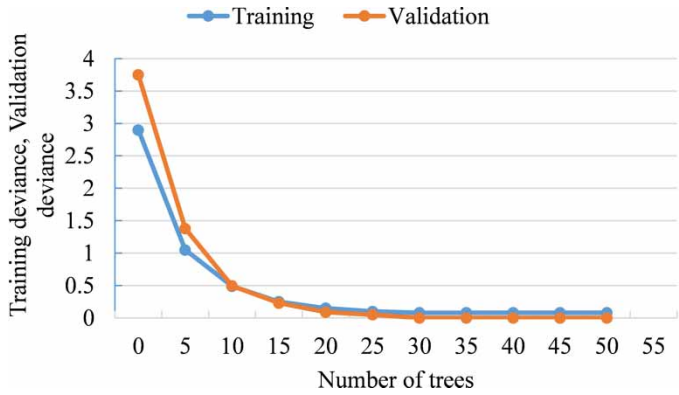


Figure 4 | Scoring deviance of GBM.

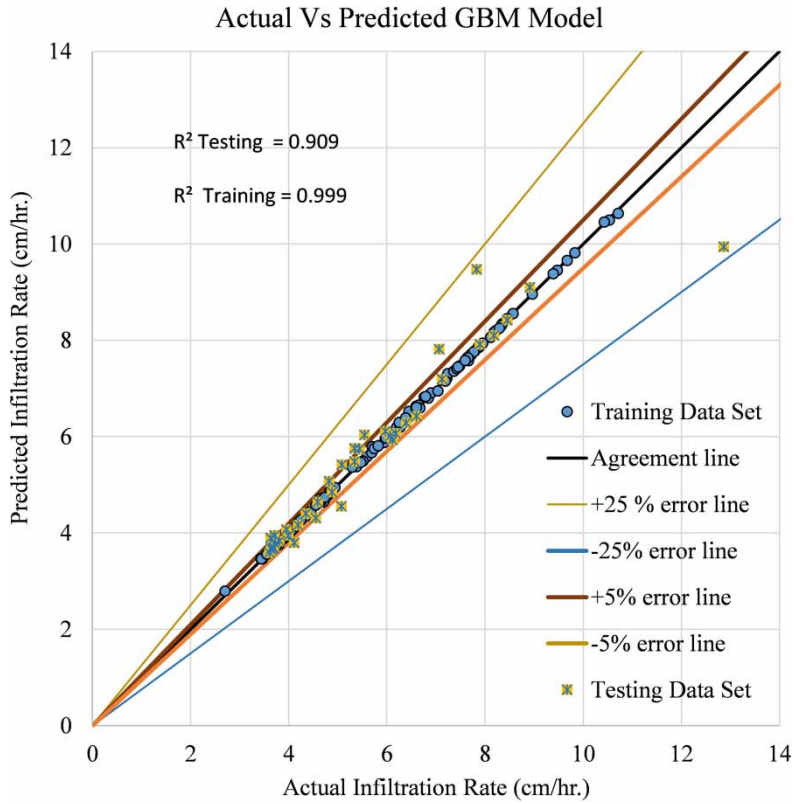


Figure 5 | Agreement diagram between predicted and observed infiltration rate using GBM.

the epochs reaches 9,500, the deviation becomes asymptomatic to the x-axis. Hence a number of epochs equal to 10,000 has been used for modeling.

Figure 7 shows the agreement diagram between predicted and observed infiltration rate. It can be seen that the predicted infiltration rate falls on or very near to the 1:1 slope line, which also demonstrates very good modeling.

4.4. Comparison of performances

From Figures 5 and 7, it can be seen that almost 99% of the value predicted using GBM and DL are close to the perfect agreement line. It is observed that for the GBM technique, most of the predicted values fall near to the trend line. Whereas in the

Table 4 | Values/type of model parameters used in DL

S.No.	DL parameter name	Value/Type
1	N folds	5
2	Fold assignment	Modulo
3	Response Column (output parameter)	I
4	Activation	Rectifier with Dropout
5	Hidden	50, 50
6	Epochs	10,000
7	Rho	0.9
8	Distribution	Gaussian
9	Categorical Encoding	One Hot Internal

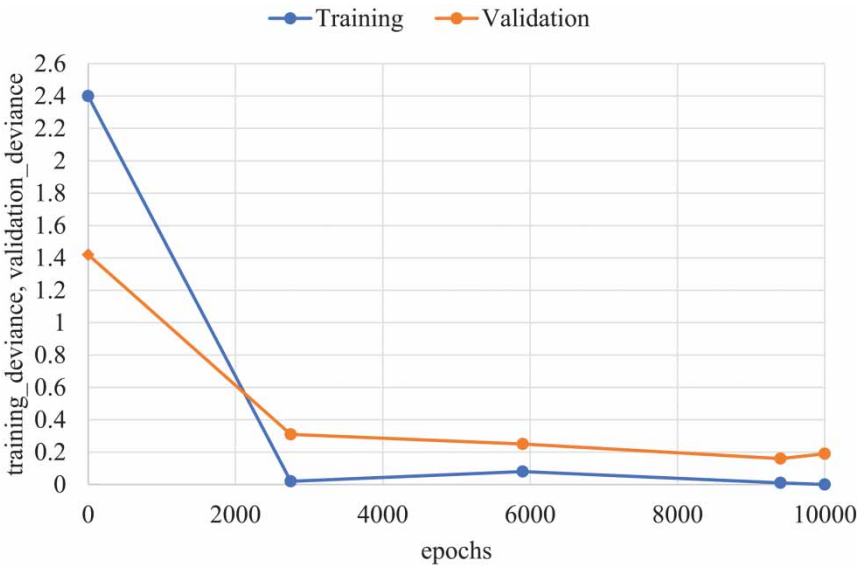


Figure 6 | Scoring deviance of DL.

case of the deep learning technique, a few values lie at notable distance from the trend line. Table 5 shows the performance comparison of different methods.

The GBM performance is best with performance parameters (CC = 0.976, RMSE = 0.238, MSE = 0.056600, MAE = 0.1606 & NSE = 0.996). Deep Learning also outperforms Philip's model (Philip 1957) with performance parameters (CC = 0.933, RMSE = 0.399, MSE = 0.159958, MAE = 0.2092 & NSE = 0.985), while in Philip's model the large amount of error is reflected in the performance parameters (CC = 0.358, RMSE = 1.871, MSE = 3.503028, MAE = 1.5928 & NSE = −0.587).

Singh *et al.* (2019) employed various soft computing techniques (M5P tree, support vector machine and Gaussian process) for predicting infiltration rate. Among these models, the support vector machine PUK kernel (Pearson VII function-based Universal Kernel) was found to perform better for both the training and testing data set. They found CC for training and testing set as 1.000 and 0.429, respectively. In the present study, for GBM and DL, the CC values obtained for the training and testing data set are 0.999, 0.933 and 0.995, 0.976, respectively. Thus, the accuracy of GBM and DL are much better compared to other models used by Singh *et al.* (2019).

Violin cum box plot has been used for the comparison of the soft computing tools. Violin cum box plot is the combination of the violin plot and box plot. In the violin cum box plot, the box plot was built between the violin plots, which gave the benefits of both of the plots. Just like the violin plot and box plot, the violin cum box plot also had multiple layers and marked the mean (horizontal bar in the box plot). The plot was created in the 'R' language environment using the 'dplyr'

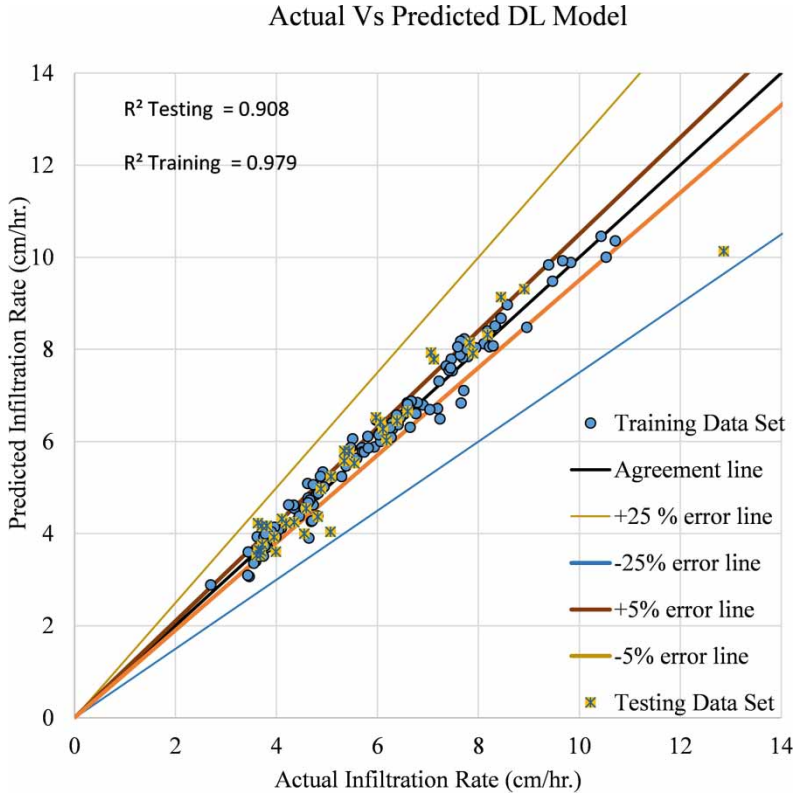


Figure 7 | Agreement diagram between predicted and observed infiltration rate using DL.

Table 5 | Results of the different techniques for training and testing data set

	Training					Testing				
Techniques	**CC**	**RMSE (cm/hr)**	**MAE (cm/hr)**	**MSE (cm/hr)^2**	**NSE**	**CC**	**RMSE (cm/hr)**	**MSE (cm/hr)^2**	**MAE (cm/hr)**	**NSE**
GBM	0.999	0.015	0.010	0.000	1.000	0.976	0.238	0.057	0.161	0.905
DL	0.995	0.135	0.102	0.018	0.979	0.933	0.399	0.160	0.2091	0.908
Philip's model (Philip 1957)	0.370	1.960	1.740	3.840	−0.258	0.144	2.430	5.910	2.090	−0.229

package. Figure 8(a) and 8(b) shows the violin cum box plot for the soft computing techniques for the prediction of infiltration rate for training and testing datasets. Figure 8(a) and 8(b) shows that the violin cum box plot was approximately symmetrical with actual values except for GBM for the training part but for the testing part, all the plots of different soft computing techniques were different, only the plot of GBM was symmetrical with the actual one. Even the mean bar in the GBM plot was the same as the actual infiltration rate. Thus, the violin cum box plot (Figure 8(a) and 8(b)) also gave concluding remarks that GBM had an edge on the DL and Philips model in the prediction of infiltration rate. The conventional model and soft computing tools can predict infiltration rate, but in the case of GBM its reliability and accuracy were higher than that of the other two models. Therefore, this method can be used to predict the infiltration rate of the rain garden in the study area.

4.5. Sensitivity analysis

Figure 9 shows the result of sensitivity analysis of the input parameters. The most sensitive parameters were time (T), water depth (d) (it is in accordance with the results available in literature (Singh *et al.* 2019)) and number of plants (n). The least sensitive parameter was inflow to rain garden (r). It was evident that the third important parameter that governs the infiltration rate was number of plants planted in the rain garden; that is, plant density. From the data observed in the field, it is seen that Scutch Grass (*Cynodon dactylon*) plant performs best out of the chosen plant types. From the observed field data it is

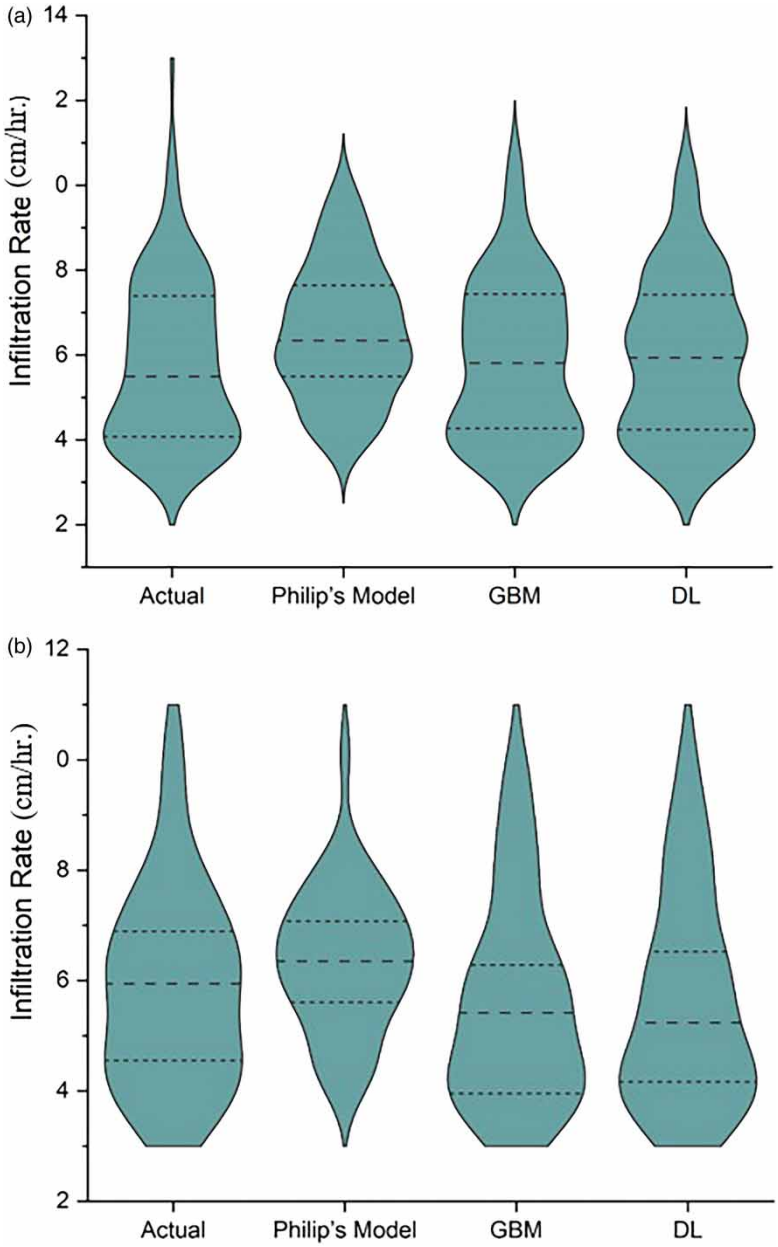


Figure 8 | (a) Violin plot for soft computing techniques training data set. (b) Violin plot for soft computing techniques testing data set.

evident that as plant density increases, the infiltration rate also increases. As far as soft computing is concerned it does not understand text or alphabetical terms, hence an efficient plant type cannot be chosen using soft computing. However, number of plants is the third most sensitive parameter and one of the inputs in the soft computing technique on which output; that is, infiltration rate, is dependent. Hence, it is useful in deciding the plant density of the rain garden.

In the case of DL, several parallel computations can be carried out with the help of graphics processing units (GPUs). These computations can even be scaled for any volume of data set. The architecture provided by DL is very flexible and can be adopted for analyzing data of any type. However, DL requires large computational efforts, which adds to the computational cost and time. No appropriate theory or specific guidelines are available in the literature which helps in selecting the appropriate DL tool. The efficiency of the DL itself depends upon several parameters such as training method, topology and so on; the problems associated with overfitting are also observed in the literature. The over fitted network gives good accuracy of predictions for the training set, but not for the testing data set (Ghasemi *et al.* 2018).

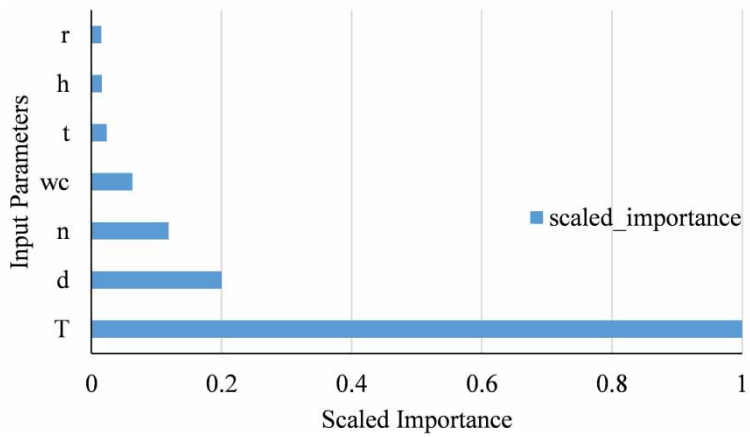


Figure 9 | Sensitivity analysis of input parameters.

As compared to DL, GBM is computationally efficient as no pre-processing is required. Even the GBM can work well with both categorical and numerical data. It has large flexibility as it has power to optimize to itself on loss function, which makes it flexible to fit the function. However, the accuracy of GBM depends upon the number of trees adopted. Hence, the number of trees must be chosen carefully (Natekin & Knoll 2013).

5. CONCLUSION

This study was carried out to model the infiltration rate of a rain garden using GBM and DL algorithms as conventional infiltration models underestimate. Results suggest a promising performance by GBM and DL in comparison to the relation proposed by Philip's model. The estimated values by both GBM and DL lie within scatter limits of $\pm 5\%$, whereas most of the values predicted by Philip's model are within a scatter of the $\pm 25\%$ error lines and even outside. Comparison of the correlation coefficient and root-mean squared error values also suggested GBM and DL algorithms perform better. A possible reason for the better performance by GBM and DL may be that they have a larger number of user-defined parameters. Tuning of these parameters as well as the flexibility of GBM and DL help them to fit the data better than the relation proposed by Philip's model.

GBM performs better than DL as CC (0.976) and NSE (0.996) values are highest and RMSE (0.238), MSE (0.056600) and MAE (0.1606) are lowest. Further, GBM being analogous to piecewise linear functions, have certain advantages over DL as they provide better insight into the created models; that is, parameter selection and assessing their relationships, and hence may be acceptable to decision makers. A lower computational time by GBM compared to the DL suggests that prediction of infiltration rate in GBM is much faster. Accurate modeling of infiltration rate of the rain garden will be useful in improved design of rain gardens.

The present research can be extended for carrying out the study on influence of soil media, plant type and its density. Performance of other soft computing techniques such as distributed random forest, generalized linear model, and so on can also be explored.

AUTHOR CONTRIBUTIONS

Conceptualization: [Sandeep Kumar, K. K. Singh]; Formal analysis and investigation: [Sandeep Kumar, K. K. Singh]; Writing – original draft preparation: [Sandeep Kumar]; Writing – review and editing: [K. K. Singh]; and Supervision: [K. K. Singh].

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

CONSENT TO PARTICIPATE

Not applicable.

CONSENT TO PUBLISH

All the authors have approved the submission and consented for publication.

FUNDING

Present work is financially supported jointly by MHRD, GOI and Director NIT kurukshetra through Ph.D. scholarship grant 2K19/NITK/PHD/61900082.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Aaron, B., Gauri, K., Jessica, P. & Kyle, T. 2012 *Redesigning the Urban Water Cycle A Vision for Redgedale Mall 2030*. Available from: https://conservancy.umn.edu/handle/11299/185362

Angelaki, A., Singh Nain, S., Singh, V. & Sihag, P. 2021 Estimation of models for cumulative infiltration of soil using machine learning methods. *ISH Journal of Hydraulic Engineering* **27** (2), 162–169. https://doi.org/10.1080/09715010.2018.1531274.

Bhandari, S., Jobe, A., Thakur, B., Kalra, A. & Ahmad, S. 2018 Flood damage reduction in urban areas with use of low impact development designs. *World Environmental and Water Resources Congress* **1** (i), 52–61. https://doi.org/10.1061/9780784481431.006.

Davis, A. P., Shokouhian, M., Sharma, H. & Minami, C. 2006 Water quality improvement through bioretention media: nitrogen and phosphorus removal. *Water Environment Research* **78** (3), 284–293. https://doi.org/10.2175/106143005x94376.

Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J. L., Mikkelsen, P. S., Rivard, G., Uhl, M., Dagenais, D. & Viklander, M. 2015 SUDS, LID, BMPs, WSUD and more – the evolution and application of terminology surrounding urban drainage. *Urban Water Journal* **12** (7), 525–542. https://doi.org/10.1080/1573062X.2014.916314.

Ghasemi, F., Mehridehnavi, A., Pérez-garrido, A. & Pérez-sánchez, H. 2018 Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discovery Today* **23** (10), 1784–1790. https://doi.org/10.1016/j.drudis.2018.06.016.

Kumar, M. & Sihag, P. 2019 Assessment of infiltration rate of soil using empirical and machine learning-based models. *Irrigation and Drainage* **68** (3), 588–601. https://doi.org/10.1002/ird.2332.

Li, J., Li, Y. & Li, Y. 2016 SWMM-based evaluation of the effect of rain gardens on urbanized areas. *Environmental Earth Sciences* **75** (1), 1–14. https://doi.org/10.1007/s12665-015-4807-7.

Malaviya, P., Sharma, R. & Sharma, P. K. 2019 Rain gardens as stormwater management tool. *Sustainable Green Technologies for Environmental Management*, 141–166. https://doi.org/10.1007/978-981-13-2772-8_7.

Mohammed, W., Welker, A. L. & Press, J. 2019 Effect of geotechnical parameters on the percolation performance of an established rain garden in Pennsylvania. *Geo-Congress* **2014**, 733–742. https://doi.org/10.1061/9780784482124.074.

Muerdter, C., Özkök, E., Li, L. & Davis, A. P. 2016 Vegetation and media characteristics of an effective bioretention cell. *Journal of Sustainable Water in the Built Environment* **2** (1), 04015008. https://doi.org/10.1061/jswbay.0000804.

Muerdter, C. P., Wong, C. K. & Lefevre, G. H. 2018 Emerging investigator series: the role of vegetation in bioretention for stormwater treatment in the built environment: pollutant removal, hydrologic function, and ancillary benefits. *Environmental Science: Water Research and Technology* **4** (5), 592–612. https://doi.org/10.1039/c7ew00511c.

Natekin, A. & Knoll, A. 2013 Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* **7**, 21. https://doi.org/10.3389/fnbot.2013.00021

Nguyen, T. T., Ngo, H. H., Guo, W., Wang, X. C., Ren, N., Li, G., Ding, J. & Liang, H. 2019 Implementation of a specific urban water management – sponge city. *Science of the Total Environment* **652**, 147–162. https://doi.org/10.1016/j.scitotenv.2018.10.168.

NIUA 2016 *Urban Flooding, Urban Climate Change Fact Sheet.* Available from: http://www.parliament.uk/documents/post/postpn289.pdf (accessed 10 July 2015).

Osheen & Singh, K. K. 2019 Rain garden – a solution to urban flooding: a review. In: *Lecture Notes in Civil Engineering*, Vol. 30. Springer, Singapore. https://doi.org/10.1007/978-981-13-6717-5_4

Osheen & Singh, K. K. 2020 The influence of slope profile on rain gardens' hydrological performance. In: *World Environmental and Water Resources Congress 2020*, Gallagher, 2011, pp. 145–153. https://doi.org/10.1061/9780784482988.015

Panagopoulos, T. 2019 Special issue: landscape urbanism and Green infrastructure. *Land* **8** (7), 112. https://doi.org/10.3390/land8070112.

Patle, G. T., Sikar, T. T., Rawat, K. S. & Singh, S. K. 2019 Estimation of infiltration rate from soil properties using regression model for cultivated land. *Geology, Ecology, and Landscapes* **3** (1), 1–13. https://doi.org/10.1080/24749508.2018.1481633.

Philip, J. R. 1957 The theory of infiltration: 1. The infiltration equation and its solution. *Soil Sci* **83** (5), 345–358.

Rafiq, S., Salim, R. & Nielsen, I. 2016 Urbanization, openness, emissions, and energy intensity: a study of increasingly urbanized emerging economies. *Energy Economics* **56**, 20–28. https://doi.org/10.1016/j.eneco.2016.02.007.

Shafique, M. 2016 A review of the bioretention system for sustainable storm water management in urban areas. *Materials and Geoenvironment* **63** (4), 227–236. https://doi.org/10.1515/rmzmag-2016-0020.

Shuster, W. D., Darner, R. A., Schifman, L. A. & Herrmann, D. L. 2017 Factors contributing to the hydrologic effectiveness of a rain garden network (Cincinnati OH USA). *Infrastructures* **2** (3), 1–14. https://doi.org/10.3390/infrastructures2030011.

Sihag, P., Singh, B., Sepah Vand, A. & Mehdipour, V. 2020 Modeling the infiltration process with soft computing techniques. *ISH Journal of Hydraulic Engineering* **26** (2), 138–152. https://doi.org/10.1080/09715010.2018.1464408.

Singh, B., Sihag, P. & Deswal, S. 2019 Modelling of the impact of water quality on the infiltration rate of the soil. *Applied Water Science* **9** (1), 1–9. https://doi.org/10.1007/s13201-019-0892-1.

Skorobogatov, A., He, J., Chu, A., Valeo, C. & van Duin, B. 2020 The impact of media, plants and their interactions on bioretention performance: a review. *Science of the Total Environment* **715**, 136918. https://doi.org/10.1016/j.scitotenv.2020.136918.

Subramanya, K. 2020 *Engineering Hydrology*, 5th edn. Tata Mcgraw Hill Education Limited, New Delhi, India.

Tirpak, R. A., Afrooz, A. N., Winston, R. J., Valenca, R., Schiff, K. & Mohanty, S. K. 2021 Conventional and amended bioretention soil media for targeted pollutant treatment: a critical review to guide the state of the practice. *Water Research* **189**. https://doi.org/10.1016/j.watres.2020.116648

Weerasundara, L., Nupearachchi, C. N., Kumarathilaka, P., Seshadri, B., Bolan, N. & Vithanage, M. 2016 *Bio-retention Systems for Storm Water Treatment and Management in Urban Systems* (A. A. Ansari, S. S. Gill, R. Gill, G. R. Lanza & L. Newman, eds). Springer, New York, NY. https://doi.org/10.1007/978-3-319-41811-7.

Worland, S. C., Farmer, W. H. & Kiang, J. E. 2018 Improving predictions of hydrological low-flow indices in ungaged basins using machine learning. *Environmental Modelling and Software* **101**, 169–182. https://doi.org/10.1016/j.envsoft.2017.12.021.

Yang, F., Fu, D., Liu, S., Zevenbergen, C. & Singh, R. P. 2020 Hydrologic and pollutant removal performance of media layers in bioretention. *Water (Switzerland)* **12** (3). https://doi.org/10.3390/w12030921

Yuan, J., Dunnett, N. & Stovin, V. 2017 The influence of vegetation on rain garden hydrological performance. *Urban Water Journal* **14** (10), 1083–1089. https://doi.org/10.1080/1573062X.2017.1363251.

Zhang, J., Zhu, Y., Zhang, X., Ye, M. & Yang, J. 2018 Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology* **561**, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065.

Zhang, L., Oyake, Y., Morimoto, Y., Niwa, H. & Shibata, S. 2019 Rainwater storage/infiltration function of rain gardens for management of urban storm runoff in Japan. *Landscape and Ecological Engineering* **15** (4), 421–435. https://doi.org/10.1007/s11355-019-00391-w.

First received 12 July 2021; accepted in revised form 23 September 2021. Available online 7 October 2021