RESEARCHERID
THOMSON REUTERS

[I-CONCEPTS-18] ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

MODELING OF INFILTRATION RATE USING DATA MINING MODELS

Parveen Sihag*[1] & Somvir Singh[2]
*[1]Department of Civil Engineering, NIT Kurukshetra, Haryana,India
[2]Department of Mechanical Engg., CMR College of Engg& Technology

ABSTRACT

Infiltration rate of soil was investigated using predictive models of artificial neural networks (ANN) and their performances were analysed with four traditional models: Kostiakov model, Philips model, Multiple Non-linear regression (MNLR) and multiple linear regression (MLR) methods. These models are judged using laboratory data. Data set consists of 392 observations of sand samples with varying initial conditions and different mixture of rice husk ash, fly ash and sand. Out of 392 data sets, 272 data were selected randomly for training and remaining 120 data were used for testing the models. Input data set consists of time, sand, rice husk ash, fly ash, suction head, bulk density and moisture content whereas the infiltration rate was considered as output. The results of the models were compared using suitable performance criterion. The judgement of the results suggests that ANN approach works wells than four conventional models. The results of the sensitivity analysis suggest that time and moisture content are the most significant parameters in estimating the infiltration rate of the soil.

Keyword: *Infiltration rate, artificial neural networks, Kostiakov model, Philips model, Multiple Non-linear regressions and multiple linear regressions.*

I. INTRODUCTION

Infiltration is the movement of water into the soil from the surface of soil. It divides water into two major hydrological components, ground water flow and surface runoff. It is the main factor in watershed modeling for the estimation of surface runoff. Exact estimation of infiltration rate is necessary for reliable estimation of surface runoff (Diamond & Shanley, 2003). The infiltration rate is important for agriculture and irrigation engineers. At catchment level, infiltration characteristics are one of the main factors in determining the flooding condition (Bhave and Sreeja 2013). Water holding capacity varies with the soil texture and soil physical properties. Sand contains comparatively larger pore size than clay and thus has higher infiltration rate and very low water holding capacity. The actual rate at which water percolates into the soil at any time is identified as the infiltration rate (Haghighi et al., 2011). The significance of the infiltration process imposed the researchers to generate several models (e.g. Green & Ampt, 1911; Richards, 1931; Philip's 1957; Holtan, 1961; Singh & Yu, 1990; Kostiakov, 1932; Horton, 1940, Modified kostiakov model, SCS model). These infiltration models can be categorized as Physical models, Semi-empirical models and empirical models.

In last few years, soft computing approaches such as artificial neural networks (ANN), Support vector machines (SVM), fuzzy logic, M5P, forest regression, and many other techniques have been widely used in water resources, hydrology applications and in other fields (Pal et al. 2011; Pal et al. 2012; Pal et al. 2014; Ghorbani et al. 2016, Angelaki et al. 2018, Nain et al. 2017 & 2018). This paper uses on artificial neural networks (ANN) based models. ANN is an adaptable system that by knowledging relationships from the input and output datasets is capable to estimate the datasets not observed before, but similar characteristics connected with the input data sets (Haykin 1999). The paper examines the performance of ANN models with respect to four conventional models: Kostiakov model, Philips model, Multiple Non-linear regression (MNLR) and multiple linear regression (MLR) in predicting the infiltration rate of the soil.

II. METHODOLOGY AND DATA SET

Experiments were conducted in a hydraulic laboratory located at National Institute of Technology, Kurukshetra, India. The soil used for experimentation, was selected as sand. Rice husk ash and fly ash were mixed with the sand in different proportions. All the measurements had been taken on predetermined initial conditions of moisture content and bulk density. The properties of the material are recorded in Table 1. The moisture content of the samples was measured in an electric oven by keeping them for 24 hours at 100°C. The soil samples, after proper mixing were carefully compacted in a proctor having a volume of 1000 cm^3.

Table 1: Properties of the material used for experimentation

Properties	Sand	Rice Husk Ash	Fly Ash
Specific gravity	2.48	1.89	2.07
D_{50} (mm)	0.438	0.190	0.180
C_u	3.1290	3.200	2.7333
Colour	White	Black	Gray

III. MEASUREMENT OF INFILTRATION IN THE LABORATORY

The cumulative infiltration was observed in the laboratory using a mini disk infiltrometer (Decagon Devices Inc. 2006). The mini disk infiltrometer consists of two chambers (water reservoir and bubble chamber), which are connected via a Mariotte tube to provide a constant negative water pressure head of 0.5 to7 cm (equivalent to 0.05 to 0.7 kPa). The bottom of the mini disk infiltrometer contains a porous sintered steel disk having 4.5 cm diameter and thickness of 3 mm. The water filled tube is placed upon the soil surface resulting in water infiltrating into the soil, with the volume of water and rate of infiltration depending on the sorptivity and hydraulic conductivity of the soil. A suction head of -0.5 cm to -6 cm (equivalent to -0.05-0.6 kPa) was chosen in this study. During the measurement, the volume of the water in the reservoir chamber was recorded at regular intervals.

Kostiakov model

Kostiakov (1932) developed an empirical model for the estimation of infiltration rate as follow:

$$f(t) = at^{-b} \tag{1}$$

Philip's model

Philip (1957) developed an infinite-series solution to resolve the non-linear partial differential Richards equation (Richards 1931), which describes transient fluid flow in a porous medium. The Philip's model is expressed as (Philip, 1957):

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

Where t is infiltration time (T), *f(t)* is the infiltration rate at time t(LT^{-1}), *A* is the rate factor (LT^{-1}) as a function of soil properties and water contents, *S* is the Sorptivity that is function of soil ($LT^{-0.5}$) and a and b are dimensionless empirical constants.

Multiple Non-linear regression (MNLR)

Multiple non-linear regression (MNLR) is applied on more than one predictors parameters. The common structure of the MNLR model is:

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

Multiple linear regression (MLR)

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

$$Z = c_0 + {x_1}^{c_1} + {x_2}^{c_2} + {x_3}^{c_3} {x_4}^{c_4} + \cdots \dots\dots\dots\dots\dots + {x_n}^{c_n} \tag{4}$$

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

Artificial neural networks (ANN)

The artificial neural network (ANN) is a machine learning method widely used for numerical prediction of hydrology problems (Kia et al. 2012; Aggarwal et al. 2012). It is inspired by the functioning of the nervous 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 between these layers represents the link between the nodes. Input layer having nodes equal to the number of input parameters, distributes the data presented to the network and does not help in processing. This layer follows one or more hidden layers which help in the processing of data. The output layer is the final processing unit. 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 (z_j) to the unit

$$P_j=\sum_i X_{ij} \times y_i \quad (5)$$

Where, X_{ij} is the weight of interconnection from unit i to j, y_i is the input value at input layer, P_j is output obtained by activation function to produce an output for unit j. The detailed discussion about ANN is provided Haykin (1999). In present analysis an ANN based one and two hidden layers are used.

Parameter		Train data	Test data	Parameter		Train data	Test data	Parameter	
	Unit	Min.	Max.	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.
t	Sec.	6.75	6056.8	341.26	511.93	22.93	4731.2	372.16	504.33
S	%	50	90	61.65	14.50	50	90	61.58	14.84
R_{ha}	%	5	45	21.01	13.23	5	45	21.58	13.63
F_a	%	5	45	17.34	11.48	5	45	16.83	11.41
S_h	cm	0.5	6	1.215	0.989	0.5	6	1.19	1.01
B_d	gm/cc	0.84	1.73	1.18	0.23	0.84	1.73	1.17	0.24
M_c	%	2	20	10.29	5.12	2	20	10.04	4.84
f (t)	cm	10.25	1341.36	433.67	256.28	11.54	1039.27	402.96	244.84

Performance criteria

Two stastical performance criterion are used in the study to evaluate the goodness of fit of the models. They are: correlation of coefficient (c.c) and root mean square error (RMSE). The correlation coefficient, can take a range of values from -1 to +1. A value of zero shows that there is no link and a value greater than zero shows a positive link between the observed and predicted data. +1 means perfect correlation and negative value means the relationship is negative i.e. when one goes up the other goes down. The RMSE provides a balance estimation of the goodness of fit of the model as it is more responsive to the large relative errors caused by low values. The ideal model will have a RMSE value of zero.

The combined use of c.c and RMSE provides a sufficient evaluation of every models performance and allows a comparison of the exactness of the five modeling strategies applied in this study.

Data Set

Data set used consist of 392 observations which are resulted from the laboratory experiments Where 272 observations were selected for training, whereas remaining 120 were used for testing the models. Input data set consists of time, sand, rice husk ash, fly ash, suction head, bulk density and moisture content whereas infiltration rate was considered as output.

IV. RESULT AND DISCUSSION

The Kostiakov model, MNLR and MLR are the empirical models where as Philip's model is physical model. The Kostiakov, Philip's, MNLR and MLR models were implemented using least square technique to drive regression coefficients using the training data set.

Kostiakov model

$$f(t) = 3810t^{-0.46} \quad (6)$$

Philip's model

$$f(t)=\frac{1}{2}5145.87t^{-0.5}+190.79 \quad (7)$$

MNLR

$$f(t)=43.82t^{-0.23}S^{0.77}F_a^{-0.09}R_{ha}^{0.277}S_h^{-0.13}B_d^{1.0024}M_c^{-0.22} \quad (8)$$

MLR

$$f(t)=514.55-0.19t+10.43S-1.95F_a-3.80R_{ha}-11.19S_h-347.87B_d-11.69M_c \quad (9)$$

The results of these models for the training and testing data set are summarized in Table 2. It shows the performance of these models, higher value of coefficient of correlation and lower value of RMSE confirms that MNLR works well in comparison to Kostiakov model, Philips, and MLR models in predicting the infiltration rate of soils. Further, single factor ANNOVA results that *F*-values (0.013268) was less than *f*- critical (3.880827) and *P*- values (0.90839) was greater than 0.05 suggest that difference in predicted values by MNLR and actual values is insignificant.

Table 2 Performance measures for different models using training and testing data set

Sr. No.	Models	Performance evaluation parameters			
		Training		Testing	
		c.c	RMSE	c.c	RMSE
1	Kostikov model	0.70955	208.8765	0.632445	197.8236
2	Philips model	0.701284	182.3603	0.623105	198.073
3	MNLR model	0.903849	110.624	**0.929017**	**105.7279**
4	MLR model	0.792577	155.9844	0.839467	135.6185

Table 2 higher value of c.c and lower value of RMSE suggests that MNLR is the best among other models. Fig. 1 and 2 display the agreement diagram for MNLR model for the training and testing data set respectively. Fig. 3 shows the actual and predicted infiltration rate of the MNLR model for

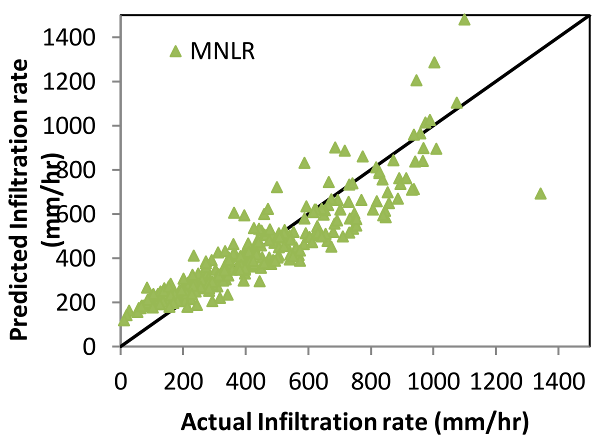

Fig. 1 Agreement Plot of actual and predicted infiltration rate of MNLR model using training data set.

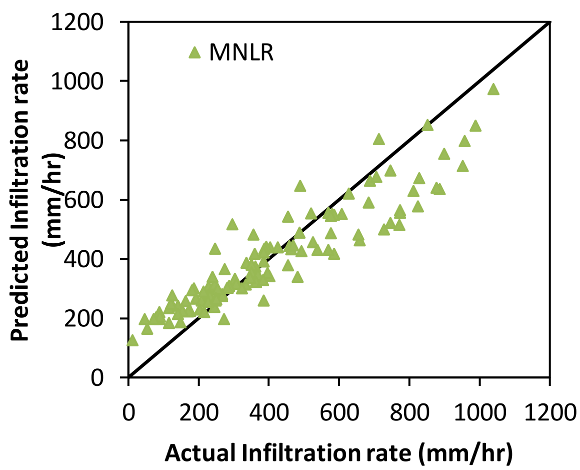

Fig. 2 Agreement Plot of actual and predicted infiltration rate of MNLR model using testing data set.

ISSN 2348 – 8034
Impact Factor- 5.070

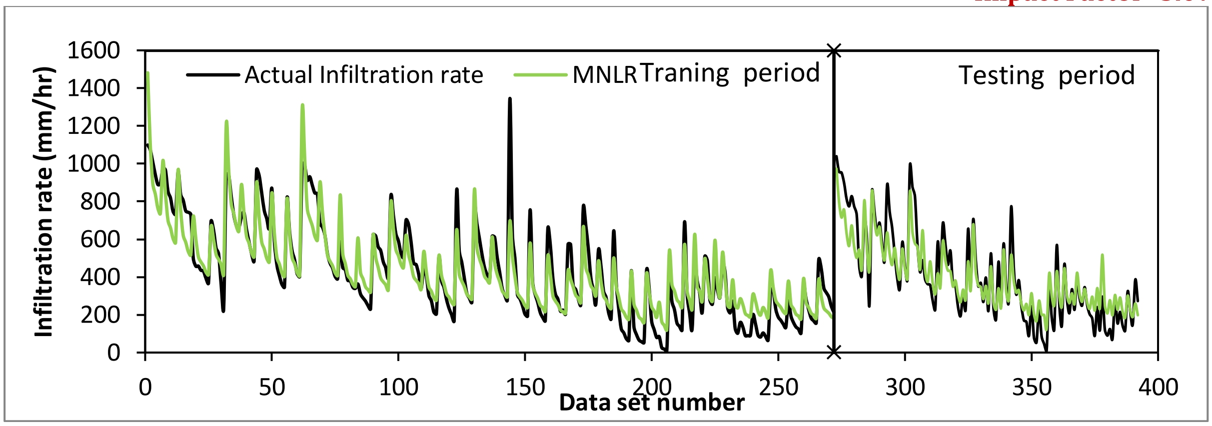

Fig. 3 Comparison of actual and predicted infiltration rate –MNLR model, training and testing data set

V. ANN MODELS

The ANN modeling was implemented using WEKA 3.9 software. The momentum and learning rate were kept as 0.2 and 0.1respectively. In this study one and two hidden layers were used to find best model at 1500 iterations. The number of neurons was changed 1 to 15 for single hidden layer and 1 to 5 for double hidden layers. The effect of change of hidden layer and number of neurons on c.c and RMSE for each model is combined in Table 3. From Table 4 ANN model 19 (7-14-4-1) performs better than other ANN models for the given data set.

Table 3 Performance measures of different ANN models using training and testing data sets

Sr. No.	ANN models	Performance evaluation parameters			
		Training		Testing	
		c.c	RMSE	c.c	RMSE
1	7-1-1	0.9164	109.1298	0.9280	99.4885
2	7-2-1	0.9343	95.6714	0.9524	81.4069
3	7-3-1	0.9300	100.2400	0.9432	89.0983
4	7-4-1	0.9089	107.9100	0.9371	85.2718
5	7-5-1	0.9099	107.1582	0.9342	87.1074
6	7-6-1	0.9093	108.2368	0.9357	86.3157
7	7-7-1	0.9466	96.3249	0.9616	83.4762
8	7-8-1	0.9378	98.8414	0.9552	84.4408
9	7-9-1	0.9471	92.2300	0.9622	79.0356
10	7-10-1	0.9411	101.9974	0.9463	96.5605
11	7-11-1	0.9416	93.2429	0.9538	82.2183
12	7-12-1	0.9523	90.1993	0.9667	78.7458
13	7-13-1	0.9351	101.8820	0.9520	89.1946
14	7-14-1	0.9513	91.5306	0.9682	77.2850
15	7-15-1	0.9339	104.2083	0.9446	95.0051
16	7-14-1-1	0.9667	72.0803	0.9743	63.1813
17	7-14-2-1	0.9663	71.1472	0.9720	63.2983
18	7-14-3-1	0.9611	75.8299	0.9688	66.9298
19	7-14-4-1	0.9605	74.5255	**0.9748**	**59.9944**
20	7-14-5-1	0.9471	91.4761	0.9638	76.6922

ISSN 2348 – 8034
Impact Factor- 5.070

Table 3 Suggests that ANN model 19 is the best among other ANN model with coefficient of correlation value is 0.9748 and RMSE value is 59.9944 for testing data set. Single factor ANNOVA results that *F*-values (0.65303) was less than *f*- critical (3.8808) and *P*- values (0.4198) was greater than 0.05 suggest that difference in predicted values by ANN model 19 and actual values is insignificant. Fig. 4 and 5 displays the agreement diagram for ANN model 19 for the training and testing data set respectively. Fig. 6 shows the actual and predicted infiltration rate of the ANN model. In particular, there is better agreement in the predicted higher values compared with their actual values.

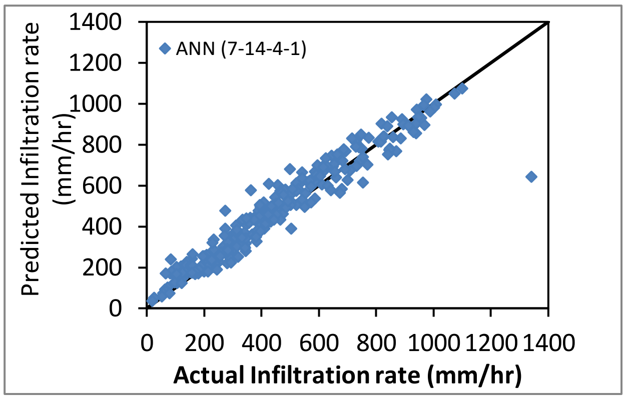

Fig. 4 Agreement Plot of actual and predicted infiltration rate of ANN model (7-14-4-1) using training data set.

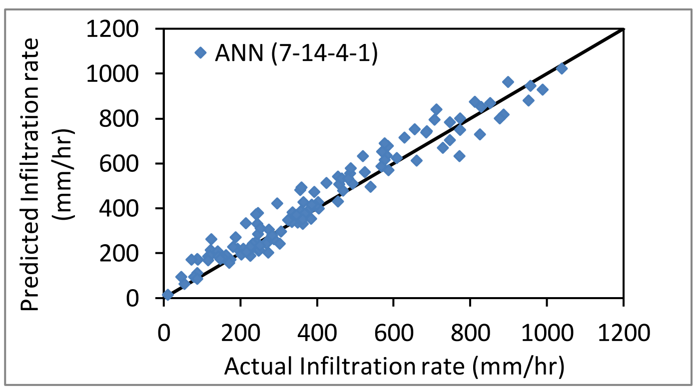

Fig. 5 Agreement Plot of actual and predicted infiltration rate of ANN model (7-14-4-1) using testing data set.

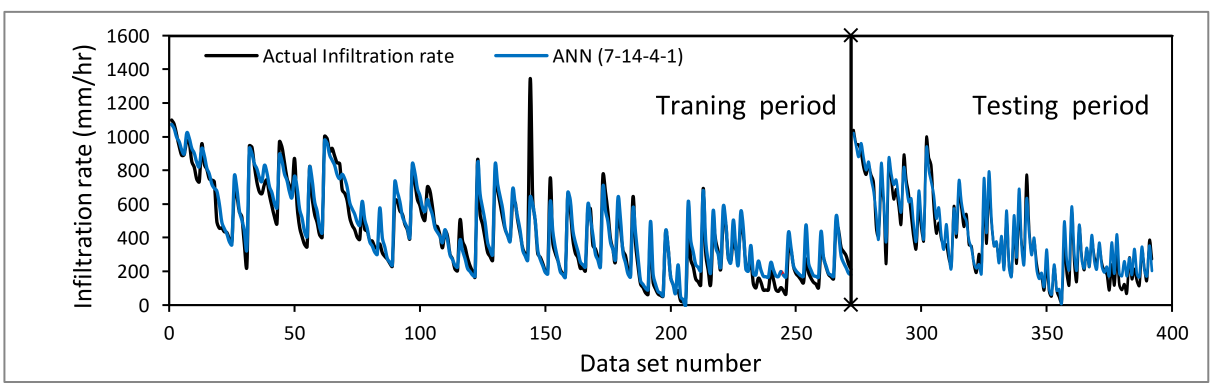

Fig. 6 Comparison of actual and predicted infiltration rate –ANN model (7-14-4-1), training and testing data set

Fig. 7 shows the scatter plot of actual and predicted values of Kostiakov, Philips, MNLR, MLR and ANN models using testing data set. It is inferred from figure that ANN has closer agreement to the perfect line. Fig. 8 shows the difference between predicted and actual infiltration rate values against test data set number for MNLR and ANN model. This Figure indicates that residuals of local picks are relatively insignificant. Single factor ANNOVA results that *F*-values (1.1152) was less than *f*- critical (3.8808) and *P*- values (0.2920) was greater than 0.05 suggest that difference in predicted values of MNLR and ANN model is insignificant. In which MNLR have higher picks of residual than ANN, So ANN model display relatively low residual.

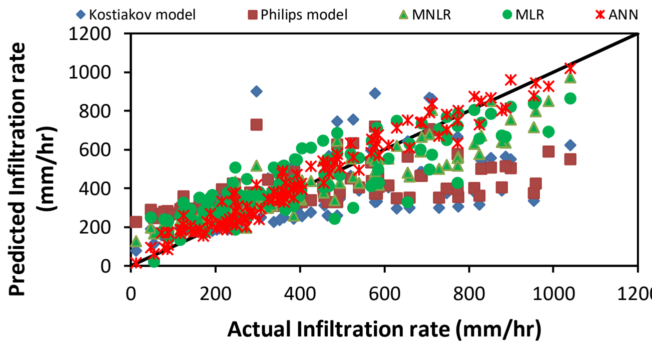

Fig. 7 Agreement Plot of actual and predicted infiltration rate of different models using testing data set.

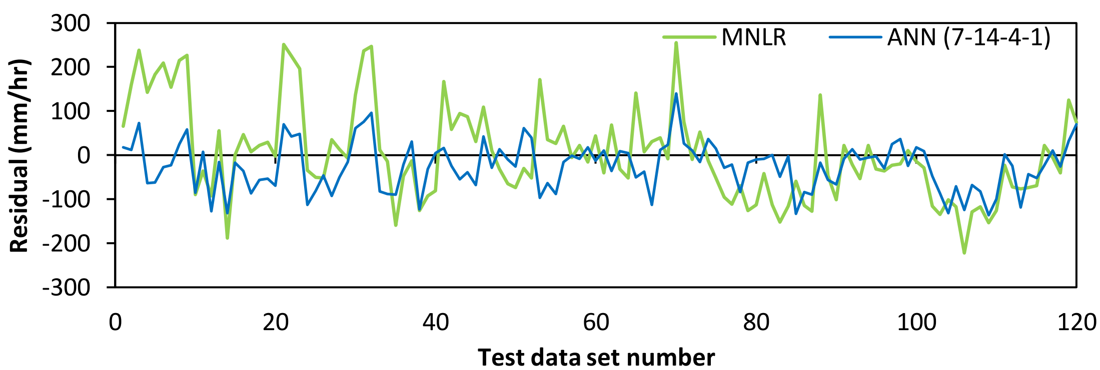

Fig. 7 Residuals from predicted values and actual values for test data set number using MNLR and ANN models.

VI. SENSITIVITY ANALYSIS

Sensitivity analysis was carried out to determine the most significant input parameter in infiltration rate of soil. For this, ANN (7-14-4-1) performing also best with the data set, was used. A different set of training data was created by removing one input parameter at a time and results were reported in terms of the coefficient of correlation and root mean square error (RMSE) with a test data set. Results from Table 4 suggest that time and moisture content of the soil has a significant role in predicting the infiltration rate of soil in comparison to other input parameter.

Table 4 Sensitivity analysis using ANN

Input combination	Input parameter removed	ANN(7-14-4-1)	
		Coefficient of correlation	Root mean square error (mm/hr)
t,S, R_{ha}, F_a, S_h, B_d, M_c		0.9748	59.9944
S, R_{ha}, F_a, S_h, B_d, M_c	t	0.8839	114.0512
t, R_{ha}, F_a, S_h, B_d, M_c	S	0.9739	58.3339
t, S, F_a, S_h, B_d, M_c	R_{ha}	0.9676	70.2919
t, S, R_{ha}, S_h, B_d, M_c	F_a	0.963	72.9673
t, S, R_{ha}, F_a, B_d, M_c	S_h	0.9567	78.8649

t, S, R_{ha}, F_a, S_h, M_c	B_d	0.967	65.1447
t, S, R_{ha}, F_a, S_h, B_d	M_c	0.9352	90.9974

VII. CONCLUSION

This paper investigates the potential of Kostiakov model, Philip's model, MNLR, MLR and Artificial neural network approaches in predicting the infiltration rate of soil. From the comparison of performance evaluation parameters, it has been found that ANN (7-14-4-1) approach works well in comparison to Kostiakov model, Philip's model, MNLR and MLR for this data set. One of the important conclusions was that MNLR works well than other conventional models: Kostiakov model, Philips model and MLR. Single factor ANNOVA results suggest that difference in predicted values of MNLR and ANN model is insignificant. Results of sensitivity analysis conclude that time and moisture content were most important parameter in measurement of the infiltration rate of soil.

REFERENCES

1. *Aggarwal, S.K., Goel, A. and Singh, V.P., 2012. Stage and discharge forecasting by SVM and ANN techniques. Water resources management, 26(13), pp.3705-3724.*
2. *Bhave, S. and Sreeja, P., 2013. Influence of initial soil condition on infiltration characteristics determined using a disk infiltrometer. ISH Journal of Hydraulic Engineering, 19 (3), pp. 291-296.*
3. *Devices, D., 2006. Mini disk infiltrometer user's manual. Decagon Devices: Pullman, WA) Available at http://www. decagon. com/products/hydrology/hydraulic-conductivity/mini-disk-portable-tension-infiltrometer/[Verified 12 May 2014].*
4. *Diamond, J. and Shanley, T., 2003. Infiltration rate assessment of some major soils. Irish Geography, 36(1), pp.32-46.*
5. *Ghorbani, M.A., Khatibi, R., Goel, A., FazeliFard, M.H. and Azani, A., 2016. Modeling river discharge time series using support vector machine and artificial neural networks. Environmental Earth Sciences, 75(8), pp.1-13.*
6. *Green WH, Ampt G. 1911. Studies on soil physics, 1. The flow of air and water through soils. The Journal of Agricultural Science 4: pp. 1–24.*
7. *Haghighi, F., Saghafian, B. and Kheirkhah, M., 2011. Evaluation of Soil Hydraulic Parameters in Soils and Land Use Change. INTECH Open Access, Publisher.*
8. *Angelaki, A., Singh Nain, S., Singh, V., &Sihag, P. (2018). Estimation of models for cumulative infiltration of soil using machine learning methods. ISH Journal of Hydraulic Engineering, 1-8.*
9. *Nain, S. S., Garg, D., & Kumar, S. (2017). Prediction of the Performance Characteristics of WEDM on Udimet-L605 Using Different Modelling Techniques. Materials Today: Proceedings, 4(2), 546-556.*
10. *Nain, S. S., Garg, D., & Kumar, S. (2017). Modeling and optimization of process variables of wire-cut electric discharge machining of super alloy Udimet-L605. Engineering Science and Technology, an International Journal, 20(1), 247-264.*
11. *Somvir Singh Nain, Dixit Garg & Sanjeev Kumar (2018) Performance evaluation of the WEDM process of aeronautics super alloy, Materials and Manufacturing Processes, 33:16, 1793-1808, DOI: 10.1080/10426914.2018.1476761*
12. *Nain, S. S., Garg, D., & Kumar, S. (2018). Evaluation and analysis of cutting speed, wire wear ratio, and dimensional deviation of wire electric discharge machining of super alloy Udimet-L605 using support vector machine and grey relational analysis. Advances in Manufacturing, 6(2), 225-246.*
13. *Nain, S. S., Garg, D., & Kumar, S. (2018). Investigation for obtaining the optimal solution for improving the performance of WEDM of super alloy Udimet-L605 using particle swarm optimization. Engineering science and technology, an international journal, 21(2), 261-273.*
14. *Nain, S. S., Sihag, P., &Luthra, S. (2018). Performance Evaluation of Fuzzy-Logic and BP-ANN Methods for WEDM of Aeronautics Super Alloy. MethodsX.*
15. *Haykin, S., 1999. Neural Networks: A comparative Foundation. Macmillan Publishing, New York.*
16. *Holtan, H.N., 1961. Concept for infiltration estimates in watershed engineering. URS41-41.US. US Dept of Agric. Service, Washington, D.C.*

ISSN 2348 – 8034
Impact Factor- 5.070

17. *Horton, R.E., 1941. An approach toward a physical interpretation of infiltration-capacity. Soil science society of America journal, 5(C), pp.399-417.*
18. *Kia, M.B., Pirasteh, S., Pradhan, B., Mahmud, A.R., Sulaiman, W.N.A. and Moradi, A., 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environmental Earth Sciences, 67(1), pp.251-264.*
19. *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, pp.17-21.*
20. *Pal, M., Singh, N.K. and Tiwari, N.K., 2011. Support vector regression based modeling of pier scour using field data. Engineering Applications of Artificial Intelligence, 24(5), pp.911-916.*
21. *Pal, M., Singh, N.K. and Tiwari, N.K., 2012. M5 model tree for pier scour prediction using field dataset. KSCE Journal of Civil Engineering, 16(6), pp.1079-1084.*
22. *Pal, M., Singh, N.K. and Tiwari, N.K., 2014. Kernel methods for pier scour modeling using field data. Journal of Hydroinformatics, 16(4), pp.784-796.*
23. *Philips, J.R., 1957. The theory of infiltration: the infiltration equation and its solution. Soil Sci, 83(5), pp.345-357.*
24. *Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Journal of Applied Physics, 1(5), pp.318-333.*