J. Indian Water Resour. Soc.,
Vol. 39, No. 1, Jan., 2019

ESTIMATION OF CUMULATIVE INFILTRATION OF SOIL USING FIELD DATASET

Parveen Sihag

ABSTRACT

This paper examine the potential of machine learning approaches i.e. Support Vector Machine (SVM) and Gaussian Process (GP) and Adaptive Neuro-Fuzzy Inference System (ANFIS) regression and compared with Kostiakov and Soil Conservation Service (SCS) models of the cumulative infiltration of soil. Data set consisting of 340 samples were used and obtained from the field experiments. Out of 340 samples arbitrarily selected 238 samples were used for training whereas remaining 102 were used for testing the models. Input data set consists of time, percentage of sand, percentage of silt, percentage of clay, bulk density and moisture content where as cumulative infiltration was considered as output. Two kernel functions i.e. Pearson VII and radial based kernel function were used with both SVM and GP regression. A comparison of results suggest that GP approach works better than SVM and ANFIS approaches for estimation of cumulative infiltration of soil. Results of sensitivity analysis conclude that time and moisture content were the most important parameter in estimation of cumulative infiltration for this dataset.

Keywords: *cumulative infiltration, Support vector regression, Gaussian process regression, Adaptive neuro-fuzzy inference system.*

INTRODUCTION

Infiltration is significant for both agriculture and irrigation system. The knowledge of infiltration features of a soil is the primary information necessary for designing and scheduling of effective irrigation system. In irrigation, most capable functions depend on the infiltration process of the soil. Infiltration rate, relative to the rate of water at which it moves into the soil from the surface sources like rainfall, irrigation etc. (Hillel, 1998). Over the decades, the significance of the infiltration process resulted in the development of numerous simple equations/models for estimation of cumulative infiltration and infiltration rate. These infiltration models range from to physically based (Green & Ampt, 1911; Philip, 1957) to empirical based (Kostiakov, 1932; Horton, 1941; Holtan, 1961; Sihag et al., 2017a (Noval model)). A standard review with a comprehensive evaluation of various infiltration models is presented by Philip (1969), and Swartzendruber and Hillel (1973). A brief review of different empirical models, Richards' equation-based models can be found in Ravi and Williams (1998). Out of numerous developed infiltration models, only some have been used effectively to field data in most studies. The main criteria in selecting the models based on the simplicity of the model parameters. Several researchers used soft-computing for the estimation of infiltration process (Sy, 2006; Singh et al., 2017; Sihag et al., 2017b; Sihag, 2018; Sihag et al. 2018a).

In the last decades, Support vector regression, Gaussian Process regression and adaptive neuro inference fuzzy system have been used as dominant tools in solving water resources problems (Parsaie and Haghiabi, 2014; Parsaie and Haghiabi, 2015; Azamathulla et al., 2016; Tiwari et al., 2017; Sihag et al., 2018b; Tiwari et al., 2018). The advantages of using SVM, GP and ANFIS are that these techniques require few user-defined parameters. Keeping in view of the improved performance by SVM, GP and ANFIS approaches in water engineering problems; this study compares its performances with empirical models (Kostiakov model and SCS model) of cumulative infiltration of soil.

1. *Civil Engineering Dept., NIT, Kurukshetra*
(Email: parveen12sihag@gmail.com)
Manuscript No. 1480

An Overview of Support vector Machines (SVM)

This method was introduced by Vapnik (1995) and derived from statistical learning theory. Main principle of SVM is optimal separation of classes, from the separable classes SVM selects the one which have least generalisation error from infinite number of linear classifier or set upper limit to error which is obtained from structural risk minimisation. Thus maximum margin between two classes could be obtained from the selected hyper plane and sum of distances of the hyper plane from the closest point of two classes will set maximum margin between two classes. For further more details study of SVM readers are referred to (Vapnik (1995) and Smola (1996).

An Overview of Gaussian Process Regression (GP)

Rasmussen and Williams (2006) assumption on which GP regression model works are that adjoining observation should express information about each other, it is a method of specifying a prior directly over function space. Mean and covariance of Gaussian distribution is vector and matrix

where as Gaussian process is over function.GP regression model are able to recognize the predictive distribution analogous to test input.

A GP is a collection of random variables, any finite number of which has a joint multivariate Gaussian distribution. Let x X y represent the domains of inputs and outputs, respectively, out of which n pairs (x_i, y_i) are drawn independently and identically distributed. For regression, let $y \subseteq \Re$; then, a GP on χ is defined by a mean function $\mu: \chi \to \Re$ and a covariance function $k: \chi \times \chi \to \Re$. For further information about GP regression and different covariance functions readers are referred to Kuss (2006).

An Overview of Adaptive Neuro fuzzy inference system (ANFIS)

It uses reasoning of fuzzy logic and algorithms of neural network to generate output. Figure 1 shows the structural design of first order Sugeno fuzzy model of ANFIS having 2 inputs (a and b), 4 rules and 1 output (c).

First-order model of Sugeno fuzzy type (Takagi and Sugeno, 1985) have four fuzzy rules (if-then), given as:

Rule 1: if a is X_1 and b is Y_1, then $f_{11} = m_{11}a + n_{11}b + q_{11}$, (3)

Rule 2: if a is X_1 and b is Y_2, then $f_{12} = m_{12}a + n_{12}b + q_{12}$, (4)

Rule 3: if a is X_2 and b is Y_1, then $f_{11} = m_{21}a + n_{21}b + q_{21}$, (5)

Rule 4: if a is X_2 and b is Y_2, then $f_{22} = m_{22}a + n_{22}b + qc_{22}$, (6)

Where X_1, X_2, Y_1 and Y_2 are fuzzy sets of input a and b, f_{ij} (i, j = 1,2) are the outputs within the fuzzy specified region by the fuzzy rule, for input a and b, m_{ij} , n_{ij} and q_{ij} (i, j = 1,2) are the design parameters that are evaluated during the training process.

Figure 1 contain five layers, each layer executes different function explained below:

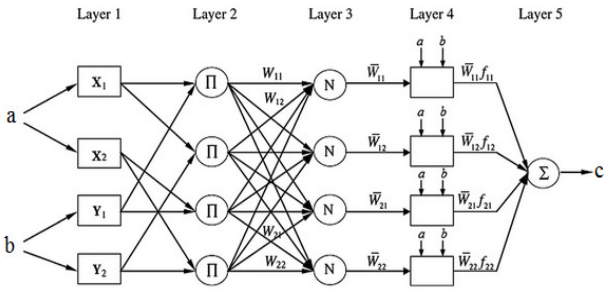

Fig. 1: ANFIS Structure

Layer 1(Input nodes): Every node is adaptive nodes and produce membership grade of input and output given by this layer are:

$$O^1_{Xi} = \mu_{Xi}(a), \quad i = 1,2, \tag{7}$$

$$O^1_{Yi} = \mu_{Yj}(b), \quad j = 1,2, \tag{8}$$

where a and b are crisp inputs, and X_i and Y_j are fuzzy set.

Layer 2 (Rule nodes): All nodes are fixed nodes and labeled as Π, which plays a role of a simple multiplier and output is given as below:

$$O^2_{ij} = Wij = \mu_{Xi}(a)\, \mu_{Yj}(b), \quad i, j \ldots\ldots\ldots\ldots = 1,2, \tag{9}$$

Layer 3(Average nodes): Every node are again fixed node and labeled as N and plays a normalization role in the network, output is give below as:

$$O^3_{ij} = \bar{W}_{ij} = \frac{W_{ij}}{W_{11}+W_{12}+W_{21}+W_{22}}, \quad i, j \ldots\ldots\ldots = 1,2, \tag{10}$$

Layer 4 (Consequent nodes): Every node is adaptive nodes and output is product of normalized firing strength and first order polynomial and is given as below.

$$O^4_{ij} = \bar{W}_{ij} f_{ij} = \bar{W}_{ij}(m_{ij}a + n_{ij}b + q_{ij}), \quad i,j \ldots\ldots = 1,2. \tag{11}$$

Layer 5 (Output nodes): The only node output in the layer is the summation output of the system.

$$O^5_1 = \sum_1^2 \sum_1^2 \overline{W_{ij}}\, f_{ij} = \sum_1^2 \sum_1^2 \overline{W_{ij}}\, (m_{ij}a + n_{ij}b + q_{ij})$$

$$= \sum_1^2 \sum_1^2 [(W_{ij}a)m_{ij} + (W_{ij}b)n_{ij} + (W_{ij})q_{ij}] \tag{12}$$

Choice of membership function

Membership function may be of many shapes for example trapezoidal, triangular, generalized bell shaped, Gaussian functions and many more. In present study, performances of triangular, trapezoidal, generalized bell-shaped and Gaussian function have been compared. All four Membership functions (MFs) are defined below:

Triangular(trimf)**:** $\mu_X(a)$ = (a-x)/(y-x), x≤ a ≤y (13)

= (z-a)/(z-y), y≤ a ≤ z

= 0

Trapozoidal(trapmf)**:** $\mu_X(a)$ = (a-x)/(y-x), x≤ a ≤y (14)

= 1, y≤ a ≤ z

= (w-a)/(w-z), z≤ a ≤ w

where x is lower limit, an upper limit w, a lower support limit y, and an upper support limit z, where x < y < z < w.

Gaussian (gaussmf)**:** $\mu_X(a) = e^{\frac{-(a-m)^2}{2\sigma^2}}$, (15)

Where a is a function of a vector and depends on the parameters σ is the Standard Deviation and m is the mean used in gaussmf.

Bell shaped (gbellmf)**:** $\mu_{Xi}(a) = \frac{1}{1+\left(\frac{a-C_j}{A_j}\right)^{2B_j}}$ i =1,2,… (16)

The generalized bell function depends on three parameters A, B, and C, where the parameter B is usually positive. The parameter C locates the center of the curve. Enter the parameter vector prams, the second argument for gbellmf, as the vector whose entries are A, B, and C respectively.

Empirical models

Kostiakov and SCS models are the empirical models. The least square techniques were used to drive regression coefficients of the Kostiakov and SCS models with the help of training data set.

Kostiakov model:

$$F(t) = at^b \quad (17)$$

$$F(t) = 1.57t^{0.6909} \quad (18)$$

SCS model: Soil Conservation Service (Jury et al. 1991), is expressed as follows:

$$F(t) = at^b + 0.6985 \quad (19)$$

$$F(t) = 1.1706t^{0.7737} + 0.6985 \quad (20)$$

Where F*(t)* is the cumulative infiltration (LT) as a function of time, *a* and *b* are the equation's parameters and *t* is time (T)

Model Performance Evaluation Criteria

To analyze the capability of various modeling methods in estimating the cumulative infiltration of soil, correlation coefficient (R), mean squared error (MSE), root mean squared error (RMSE) and Nash-Sutcliffe model efficiency (NSE) values were calculated using the training and the testing dataset.

$$R = \frac{n\sum HF - (\sum H)(\sum F)}{\sqrt{n(\sum H^2) - (\sum H)^2}\sqrt{n(\sum F^2) - (\sum F)^2}} \quad (21)$$

$$MSE = \frac{1}{n}(\sum_{i=1}^{n}(H-F)^2 \quad (22)$$

$$RMSE = \sqrt{\frac{1}{n}(\sum_{i=1}^{n}(H-F)^2} \quad (23)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(H-F)^2}{\sum_{i=1}^{n}(H-\bar{H})^2} \quad (24)$$

Where:

$H = observed\ values$

$F = predicted\ values$

$\bar{H} = mean\ of\ observed\ values$

$n = number\ of\ observations$

STUDY AREA

Kurukshetra district falls in the north-east part of the Haryana State, India and is bounded by North latitudes $29^0$53'00" and $30^0$15'02" and East longitudes $76^0$26'27" and $77^0$07'57". Thanesar Tehsil of Kurukshetra district is selected as study area. The total area of Kurukshetra district is 1530 Km^2. The district covers 3.46% area of the State. The topographical map of the Kurukshetra district can be referred to the Toposheets of 53B and 53C of survey of India. The river Markhanda provides the major drainage in the area. Location map of the study area is shown in Figure 2. Study area (Thanesar) is a part of Ghaggar basin. Total 20 different locations were selected for experimentation in the study area. The coordinates of all the locations are listed in Table 1. The texture of the soil are listed and shown in Table 2 and Figure 3, respectively.

Table 1: The coordinate system of all locations

Site No.	Location	Latitude	Longitude
1	Dayalpur	29.939648N	76.814545E
2	Samshipur	29.925980N	76.803795E
3	Kirmach (SKS)	29.911368N	76.794275E
4	Alampur	29.938222N	76.824080E
5	Sanheri Khalsa	29.918557N	76.826591E
6	Mirzapur	29.950163N	76.781358E
7	Khanpur Roran	29.939504N	76.757209E
8	Barna	29.924569N	76.733358E
9	Pindarsi	29.919078N	76.702227E
10	Kamoda	29.936836N	76.736818E
11	Lohar Majra	29.958742N	76.727137E
12	Jyotisar	29.960166N	76.760195E
13	Narkatari	29.962200N	76.797872E
14	Kurukshetra University	29.95.5052N	76.815767E
15	Thim Park	29.967055N	76.832005E
16	Darra Khera	29.981300N	76.822550E
17	Bhiwani Khera	29.994305N	76.826474E
18	Bahadur Pura	30.008150N	76.834262E
19	Hansala	30.011900N	76..811639E
20	Durala	30.025939N	76.809048E

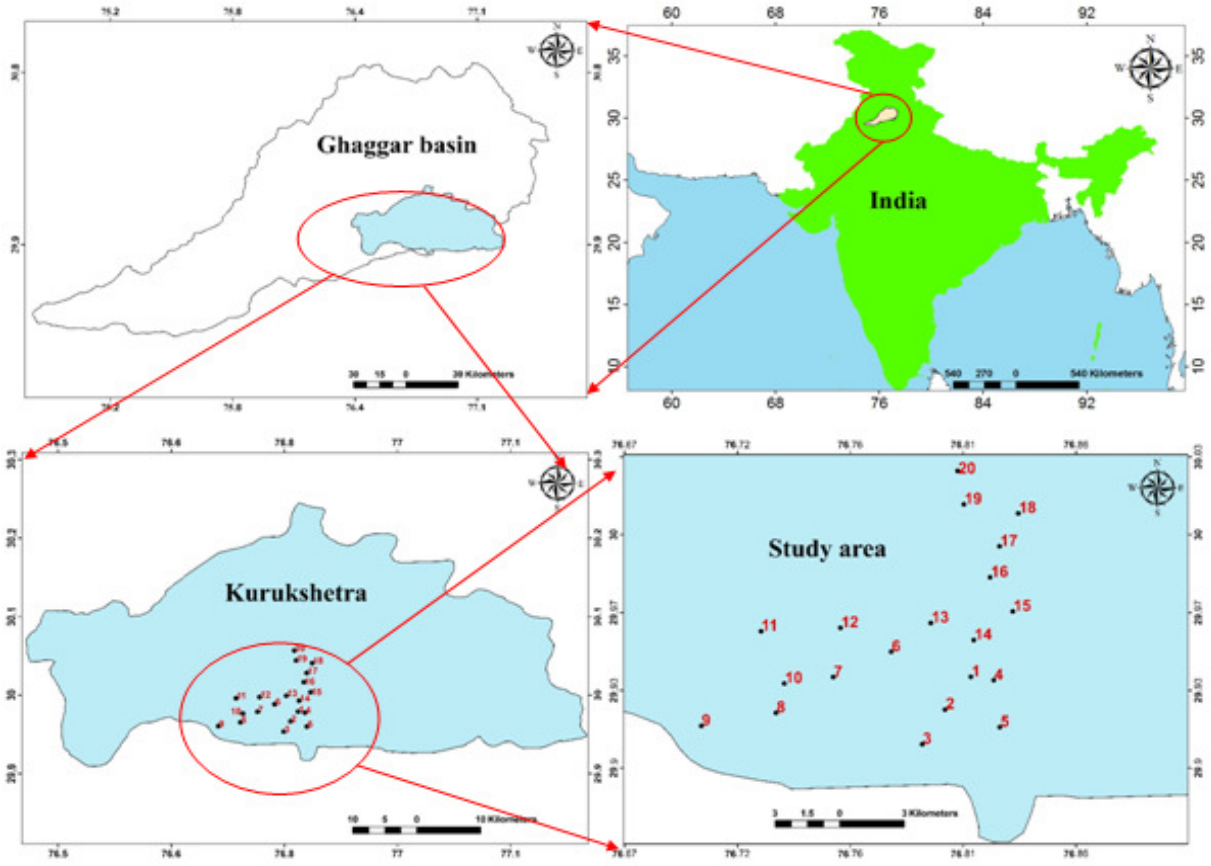

Fig. 2: Study area

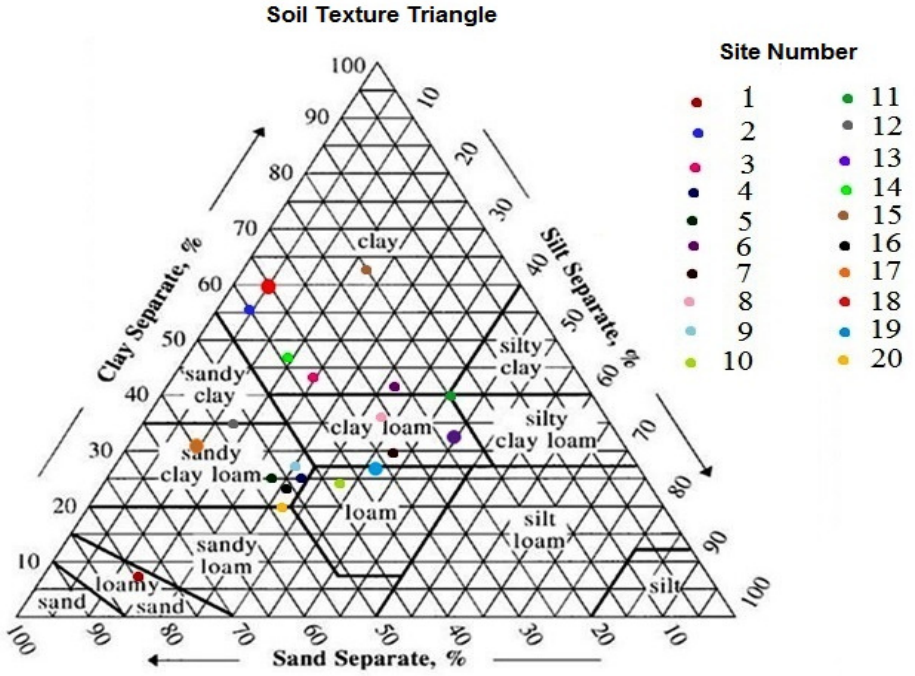

Fig. 3. Texture of the soil for the study area

Table 2: Texture of the soil

Site No.	Location	Texture	Sand (%)	Clay (%)	Silt (%)
1	Dayalpur	Loamy Sand	78.73	7.4445	13.8255
2	Samshipur	Clay	39.84	55.3472	4.8128
3	Kirmach (SKS)	Clay	37.14	43.3734	19.4866
4	Alampur	Sandy clay Loam	47.5	25.2	27.3
5	Sanheri Khalsa	Sandy clay Loam	52.11	24.9028	22.9872
6	Mirzapur	Clay	26.63	41.8209	31.5491
7	Khanpur Roran	Clay loam	32.94	29.5064	37.5536
8	Barna	Clay loam	31.52	35.7133	32.7667
9	Pindarsi	Sandy clay Loam	47.6	27.248	25.152
10	Kamoda	Loam	42.85	24.003	33.147
11	Lohar Majra	Clay loam	24.6	39.962	35.438
12	Jyotisar	Sandy clay Loam	52.71	34.5217	12.7683
13	Narkatari	Clay loam	22.93	32.3694	44.7006
14	KUK	Clay	52.74	19.85	27.41
15	Thim Park	clay	36.7	26.586	36.714
16	Dara kheda	Sandy clay Loam	35.31	59.5148	5.1752
17	bhiwani kheds	Sandy clay Loam	59.58	30.7192	9.7008
18	bhaderpura	Clay	50.78	23.6256	25.5994
19	Singhpura	Loam	19.74	62.6028	17.6572
20	Durala	Sandy Loam	39.13	46.2612	14.6088

Observation of cumulative infiltration

The cumulative infiltration of soil was experimentally observed in the field with the help of mini disk infiltrometer (Decagon Devices Inc., 2014). Mini disk infiltrometer contains two parts (water reservoir and bubble), which are coupled via a Mariotte tube to supply a steady water pressure head of 0.05 to 0.7 kPa. The base of the mini disk infiltrometer contains a porous sintered steel disk having diameter of 4.5 cm and thickness of 3 mm. The water filled tube is located on the soil surface (Figure 4) resulting in water infiltrating into the soil. During the observation, the volume of the water in the reservoir part was recorded at standard intervals.

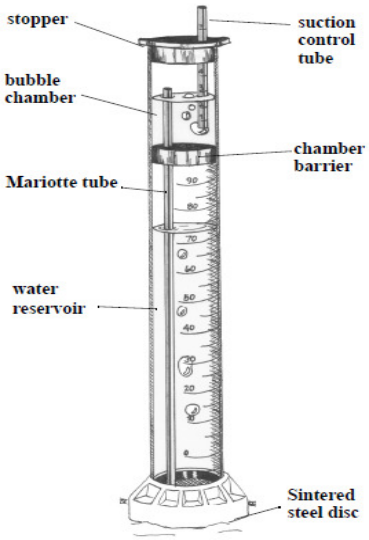

Fig. 4: Mini disk Infiltrometer (Infiltrometer User's Manual, 2014)

Data Set

The whole dataset containing 340 observations from field infiltration experiments was divided into two separate parts of training and testing, respectively. Training data involves 70% of the total data chosen randomly from whole data set, while testing data involves remaining 30% of the total data. The features of training and testing data sets are represented in Table 3 in which time, sand, clay, silt, bulk density and moisture content are input parameters and cumulative infiltration of soil is target.

Table 3: Features of the data set

Parameter	Unit	Training data				Testing data			
		Lower	Higher.	mean	Std. deviation	Lower	Higher.	mean	Std. deviation
Time (t)	min.	1.00	17.00	9.08	4.98	1.00	17.00	8.80	4.75
Sand(S)	(%)	19.74	78.73	41.32	13.81	19.74	78.73	40.88	13.84
Clay(C)	(%)	7.44	62.60	33.58	13.35	7.44	62.60	34.87	14.38
Silt (Si)	(%)	4.81	44.70	25.09	10.99	4.81	44.70	24.25	10.96
bulk density (ρ)	gm/cc	1.39	1.90	1.67	0.13	1.39	1.90	1.66	0.13
moisture content (MC)	(%)	1.49	14.19	7.72	3.14	1.49	14.19	7.72	3.07
Cumulative Infiltration (F(t))	mm	0.63	25.90	6.95	4.86	0.94	23.89	6.82	4.55

Details of GP and SVM

The GP and SVM-based regression approaches design involves the idea of kernel function. A number of kernels are discussed in the literature, but studies suggest a better performance by radial basis kernels for different civil engineering problems (Sihag et al., 2017c; Kumar et al, 2018; Sihag et al., 2018b) proposed a universal Pearson VII function-based kernel and suggested it to be an alternative to the linear, polynomial and radial basis function kernels. The present study uses the radial basis kernel ($e^{-\gamma\left|\mathbf{x}_i-\mathbf{x}_j\right|^2}$) and the Pearson VII function kernel $\left(1\Big/\left[1+\left(2\sqrt{\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2}\sqrt{2^{(1/\omega)}-1}\Big/\sigma\right)^2\right]^{\omega}\right)$, where γ, σ and ω are kernel-specific parameters. The parameter σ controls the Pearson width, whereas ω is the tailing factor of the peak when the Pearson VII function is used for curve-fitting purposes.

In the present study, a manual method (carrying out a large number of trials by using different combinations of user-defined parameters with both modeling approaches) was used to select user-defined parameters (i.e., C, γ, σ, ω, ε and Gaussian noise). For SVM, multiple trials were also carried out to find a suitable value of the error-insensitive zone with a fixed value of C and kernel-specific parameters. Optimal values of various user-defined parameters are chosen in such a way so as to minimize the root mean square error and maximize the correlation coefficient. The same kernel-specific parameters were used for both GP regression and SVM. Table 4 provides the optimal values of the user-defined parameters for SVM, GP and ANFIS. To assess the performance of SVM, GP and ANFIS modeling approaches and empirical relation (Kostiakov model and SCS model), R, MSE, RMSE and NSE values were used.

Table 4: User-defined parameters using SVM, GP and ANFIS

	RBF kernel	PUK kernel
Support vector machines	C = 3, γ = 2.5	C =3, ω = 0.1, σ = 0.1
Gaussian process regression	Gaussian noise = 0.01, γ = 2.5	Gaussian noise = 0.01, ω = 0.1, σ = 0.1
ANFIS	MFs= triangular, trapezoidal, generalized bell and Gaussian shape Number of MFs =2,2,2,2,2,2 and Epoch= 10	

RESULT AND DISCUSSION

Figure 5 and 6 provide plots between actual and estimated cumulative infiltration of soil by SVM, GP and ANFIS regression approaches using training and testing data set respectively. Results from SVM, GP and ANFIS regression approaches are in accordance with the actual values. Results of test data set from Table 5 indicate that the performance of RBF kernel function based GP is best among GP, SVM and ANFIS approaches. Correlation coefficient value of 0.9999, MSE= 0.0041 (RMSE = 0.0641mm and NSE=

0.9998 was achieved by RBF kernel function based GP suggests a better performance in comparison to RBF kernel based SVM. In comparison to both GP and SVM regression approaches, GP approach performance was quite better (Table 6). The Correlation coefficient value of 0.9994, MSE= 0,0241mm, RMSE= 0.1552mm and NSE= 0.9988 was achieved by triangular Membership function based ANFIS approach. Table 6 suggest that triangular Membership function based ANFIS model works better than other membership function based ANFIS models.

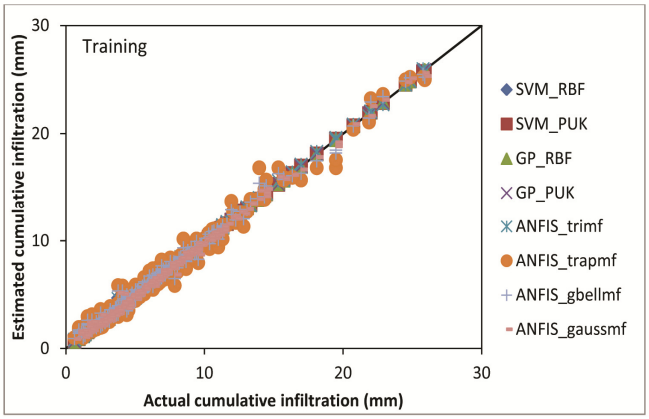

Fig. 5: Actual vs. estimated values of cumulative infiltration of soil using GP, SVM and ANFIS with training data

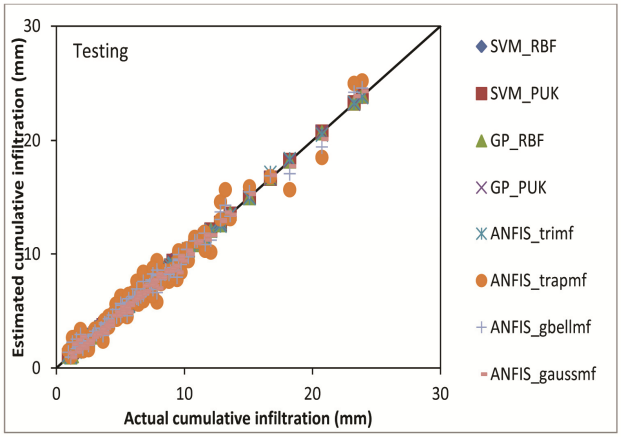

Fig. 6: Actual vs. estimated values of cumulative infiltration of soil using GP, SVM and ANFIS with testing data

Figure 7 and 8 provide plots between actual and estimated cumulative infiltration of soil by Kostiakov and SCS models using training and testing data set respectively. These figures indicate that values obtained from Kostiakov and SCS models are not closer to line of perfect agreement. Table 5 indicates that soft computing based models has higher prediction capability than conventional models for this data set.

Table 5: Performance of SVM, GP, ANFIS, Kostiakov model and SCS model.

Approaches	Training data set				Testing data set			
	R	MSE	RMSE	NSE	R	MSE	RMSE	NSE
SVM_RBF	0.9998	0.0118	0.1084	0.9995	0.9998	0.0067	0.0819	0.9997
SVM_PUK	1.0000	0.0022	0.0465	0.9999	0.9999	0.0059	0.0767	0.9997
GP_RBF	1.0000	0.0017	0.0409	0.9999	0.9999	0.0041	0.0641	0.9998
GP_PUK	1.0000	0.0002	0.0150	1.0000	0.9999	0.0052	0.0719	0.9997
ANFIS_trimf	0.9995	0.0245	0.1565	0.9990	0.9994	0.0241	0.1553	0.9988
ANFIS_trapmf	0.9897	0.4830	0.6950	0.9795	0.9815	0.7634	0.8737	0.9627
ANFIS_gbellmf	0.9960	0.1857	0.4310	0.9921	0.9939	0.2497	0.4997	0.9878
ANFIS_gaussmf	0.9982	0.0843	0.2903	0.9964	0.9977	0.0944	0.3072	0.9954
Kostiakov Model	0.5844	15.5020	3.9373	0.3415	0.5757	13.7021	3.7016	0.3301
SCS	0.5846	15.4955	3.9364	0.3417	0.5738	13.7341	3.7060	0.3286

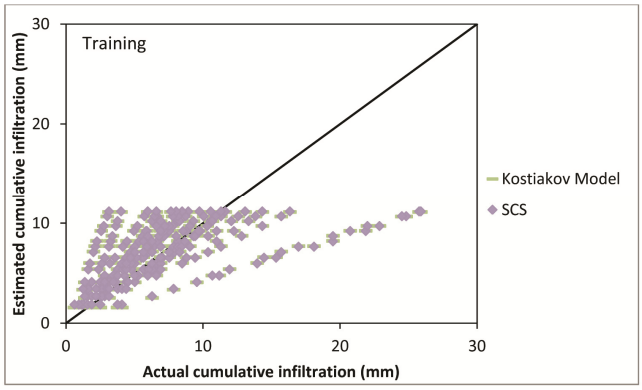

Fig. 7: Actual vs. estimated values of cumulative infiltration of soil using Kostiakov and SCS model with training data

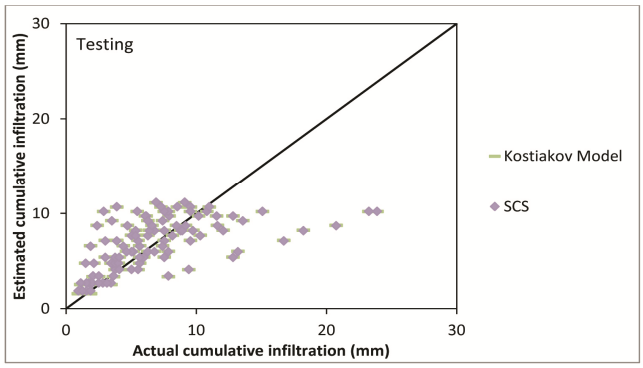

Fig. 8: Actual vs. estimated values of cumulative infiltration of soil using Kostiakov and SCS model with testing data

Keeping in view of the improved performance of RBF kernel function based GP modeling approaches, a graph between test data set number and cumulative infiltration of soil is plotted (Figure 9). It can be inferred from this figure that estimated values using RBF kernel function based GP found to follow the same patterns of actual cumulative infiltration of soil.

Sensitivity Analysis

Sensitivity analysis was carried out to determine the most significant input parameter in cumulative infiltration of soil. GP_RBF as the ablest method of this research was selected to run sensitivity analysis. Several set of training data was created by removing one input parameter at a time and results were reported in terms of coefficient of correlation and root mean square error (RMSE) with training data set. Results from Table 6 suggest that time and moisture content of the soil have significance role in estimation of cumulative infiltration of soil in comparison to other input parameter.

Table 6: Sensitivity analysis using RBF kernel based GP regression

Input combination	Input parameter Removed	GP regression	
		Coefficient of correlation	Root mean square error (mm)
t, S, C, Si, ρ, MC		0.9999	0.0641
S, C, Si, ρ, MC	T	0.7624	2.9289
t, C, Si, ρ, MC	S	0.9999	0.0656
t, S, Si, ρ, MC	C	0.9999	0.0645
t, S, C, ρ, MC	Si	0.9999	0.0623
t, S, C, Si, MC	P	0.9998	0.0827
t, S, C, Si, ρ	MC	0.8948	2.0595

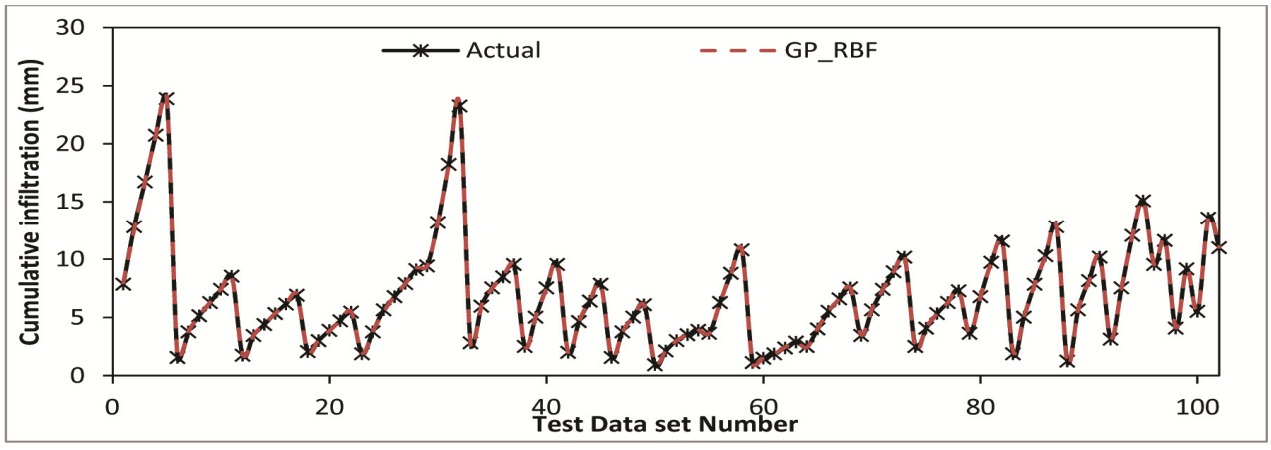

Fig. 9: Variation in estimated values of cumulative infiltration using GP_RBF regression approach in comparison to actual values of cumulative infiltration.

CONCLUSION

This paper investigates the potential of SVM, GP and ANFIS based regression approach in estimation of cumulative infiltration of soil. From the comparison of performance evaluation parameters it has been found that RBF based GP regression approach works better than in comparison to SVM, ANFIS and conventional models (Kostiakov and SCS model) for this data set. One of important conclusion was that GP regression works better than all other approaches. Triangular membership based ANFIS model is superior to other membership based ANFIS models in estimation of cumulative infiltration of soil. Results conclude that soft computing based models have suitable capability than conventional models in estimation of cumulative infiltration of soil. Results of sensitivity analysis recommend that time and moisture content were the most important parameter in estimation of cumulative infiltration for this dataset.

REFERENCES

1. Azamathulla, H.M., Haghiabi, A.H. and Parsaie, A. (2016). Prediction of side weir discharge coefficient by support vector machine technique. *Water Science and Technology: Water Supply*, *16*(4), pp.1002-1016.

2. Devices, D. (2014). Mini disk infiltrometer user's manual, Version 9. *Decagon Devices, Pullman, WA*.

3. Green, W.H. and Ampt, G.A. (1911). Studies on Soil Phyics. *The Journal of Agricultural Science*, *4*(1), pp.1-24.

4. Hillel, D. (1998). *Environmental soil physics: Fundamentals, applications, and environmental considerations*. Elsevier.

5. Holtan, H.N. (1961). Concept for infiltration estimates in watershed engineering. USDA-ARS, pp 41–51.

6. Horton, R.E. (1941). An Approach Toward a Physical Interpretation of Infiltration-Capacity 1. *Soil Science Society of America Journal*, *5*(C), pp.399-417.

7. Jury W.A., Gardner W.R., Gardner W.H. (1991). Soil physics. 5th ed. New York (NY): John Wiley & Sons, pp.328.

8. Kostiakov, A.N. (1932). On the dynamics of the coefficient of water percolation in soils and the necessity of studying it from the dynamic point of view for the purposes of amelioration. *Trans. Sixth Comm. Int. Soc. Soil Sci.*, *1*, pp.7-21.

9. Kumar, M., Tiwari, N.K. and Ranjan, S. (2018). Prediction of oxygen mass transfer of plunging hollow jets using regression models. *ISH Journal of Hydraulic Engineering*, pp.1-8

10. Kuss, M., 2006. *Gaussian process models for robust regression, classification, and reinforcement learning* (Doctoral dissertation, Technische Universität).

11. Parsaie, A. and Haghiabi, A. (2014). Predicting the side weir discharge coefficient using the optimized neural network by genetic algorithm. *Scientific Journal of Pure and Applied Sciences*, *3*(3), pp.103-112.

12. Parsaie, A. and Haghiabi, A. (2015). The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. *Water Resources Management*, *29*(4), pp.973-985.

13. Philip, J.R. (1957). The theory of infiltration: 4. Sorptivity and algebraic infiltration equations. *Soil science*, *84*(3), pp.257-264.

14. Philip, J.R. (1969). Theory of infiltration. In *Advances in hydroscience* (Vol. 5, pp. 215-296). Elsevier.

15. Rasmussen, C.E. and Williams, C.K. (2006). Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, *38*, pp.715-719.

16. Ravi, V., Williams, J.R. and Ouyang, Y. (1998). Estimation of infiltration rate in the vadose zone: compilation of simple mathematical models. Vol. I. United States Environmental Protection Agency, EPA/600/R-97/128a, 26pp

17. Sihag, P. (2018). Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Modeling Earth Systems and Environment*, pp.1-10.

18. Sihag, P., Jain, P. and Kumar, M. (2018b). Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. *Modeling Earth Systems and Environment*, pp.1-8.

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

20. Sihag, P., Tiwari, N.K. and Ranjan, S. (2017b). Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH Journal of Hydraulic Engineering*, pp.1-11.

21. Sihag, P., Tiwari, N.K. and Ranjan, S. (2017c). Modelling of infiltration of sandy soil using gaussian process regression. *Modeling Earth Systems and Environment*, *3*(3), pp.1091-1100.

22. Sihag, P., Tiwari, N.K. and Ranjan, S. (2018a). Support vector regression-based modeling of cumulative infiltration of sandy soil. *ISH Journal of Hydraulic Engineering*, pp.1-7.

23. Singh, B., Sihag, P. and Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment*, *3*(3), pp.999-1004.

24. Smola, A.J. (1996). *Regression estimation with support vector learning machines* (Doctoral dissertation, Master's thesis, Technische Universität München).

25. Swartzendruber, D. and Hillel, D. (1973). The physics of infiltration. In *Physical aspects of soil water and salts in ecosystems* (pp. 3-15). Springer, Berlin, Heidelberg.

26. Sy, N.L. (2006). Modelling the infiltration process with a multi-layer perceptron artificial neural network. *Hydrological sciences journal*, *51*(1), pp.3-20.

27. Tagaki, T. and Sugeno, M. (1985). Fuzzy identification of systems and its application to modelling and control. *IEEE Trans. Syst. Man and Cybernetics*, *15*(1), pp.116-132.

28. Tiwari, N.K., Sihag, P. and Ranjan, S. (2017). Modeling of Infiltration of Soil using Adaptive Neuro-fuzzy Inference System (ANFIS). *Journal of Engineering & Technology Education*, 11 (1), pp. 13-21.

29. Tiwari, N.K., Sihag, P., Kumar, S. and Ranjan, S. (2018). Prediction of trapping efficiency of vortex tube ejector. *ISH Journal of Hydraulic Engineering*, pp.1-9.

30. Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. New York: Springer-Verlag.