SCIENTIFIC PAPER
Published: December 15, 2025
https://doi.org/10.13167/2025.31.16

Field-based assessment and multilinear modelling of infiltration dynamics across heterogeneous zones

Sunith John David[1,2]✉, Abdu Rahiman[2] and Subha Vishnudas[1]

[1] *Department of Civil Engineering, School of Engineering, CUSAT, Kochi, India*

[2] *Department of Civil Engineering, Muthoot Institute of Technology and Science, Kochi, India*

Corresponding author:
Sunith John David

Received:
August 14, 2025

Revised:
September 26, 2025

Accepted:
October 2, 2025

Published:
December 15, 2025

Citation:
David, S.J.; Rahiman, A.; Vishnudas, S.
Field-based assessment and multilinear modelling of infiltration dynamics across heterogeneous zones.
Advances in Civil and Architectural Engineering, 2025, 16 (31), pp. 266-282.
https://doi.org/10.13167/2025.31.16

ADVANCES IN CIVIL AND ARCHITECTURAL ENGINEERING (ISSN 2975-3848)

Faculty of Civil Engineering and Architecture Osijek
Josip Juraj Strossmayer University of Osijek
Vladimira Preloga 3
31000 Osijek
CROATIA

Abstract:
Accurate estimation of infiltration is critical for hydrological modelling, particularly in regions with heterogeneous physiographic conditions. Conventional field measurements, while reliable, are labour-intensive and spatially constrained, underscoring the need for robust predictive models. This study presents a multiple linear regression (MLR) framework for predicting infiltration rates across diverse terrains in Kerala, India, using both primary field measurements and secondary soil property data. Key infiltration influencing parameters percentage silt, clay, sand, bulk density, initial moisture content, and time were incorporated into the model, with logarithmic transformation applied to linearize the relationship. Model coefficients were derived using the least squares method in IBM SPSS, with statistical significance, multicollinearity, and overall adequacy assessed through variance inflation factor (VIF), coefficient of determination (R^2), adjusted R^2, and standard error of estimate. Calibration was achieved by introducing an infiltration coefficient (K), determined from the ratio of field to predicted infiltration rates, with values ranging from 5,7 to 8,9 across locations. Validation using one-way ANOVA and R^2 analysis confirmed high predictive accuracy and model robustness. LOOCV sensitivity analysis identified time (T) as the most influential predictor, with soil texture parameters showing comparable effects. The developed MLR model provides a scalable, validated tool for infiltration estimation across varied physiographic conditions.

Keywords:
infiltration modelling; sensitivity analysis; Leave-One-Out Cross-Validation (LOOCV)

1 Introduction

Infiltration describes the process through which water moves from the land surface into the soil, which is a key step in controlling water availability, groundwater recharge, and surface runoff in various environments. In civil engineering, understanding how infiltration occurs is especially important in designing effective storm water systems, managing flood risk, and ensuring reliable agricultural irrigation. By quantifying how much water enters the soil during and after rainfall events, engineers can make informed decisions that influence infrastructure resilience and environmental sustainability [1; 2]. The conventional double ring infiltrometer remains the standard instrument for measuring field infiltration rates, as specified in Indian guidelines [3]. However, this method is often labour-intensive, time-consuming, and requires significant human involvement to obtain reliable results. To address these practical limitations, numerous infiltration models have been developed to predict the physical behaviour of infiltration under various soil and environmental conditions. These models offer efficient alternatives for estimating infiltration rates, and thus streamline the assessment process for hydrological and engineering applications [4-6].

Regression techniques are essential for linking observed infiltration behaviour with quantifiable soil and environmental variables in hydrological studies. Their principal applications include calibrating model parameters to align predictions with field measurements, evaluating and selecting the most suitable infiltration models based on statistical metrics such as the coefficient of determination and root mean square error, and constructing predictive tools by integrating influential factors such as soil texture, antecedent moisture, and rainfall characteristics through linear and non-linear analyses [4; 7]. In addition, the incorporation of advanced regression approaches, including machine learning methods, has expanded the capacity to interpret large datasets and discern the influence of site-specific conditions. Thus these methods have enhanced the predictive reliability and practical utility of infiltration models in engineering design and water resources management [8-10].

1.1 Regression models

In the context of infiltration modelling and hydrological studies, various regression models have been adopted to analyse, predict, and calibrate infiltration processes based on observed data and measurable site parameters. Regression approaches ranging from simple linear forms to advanced machine learning models allow hydrologists and engineers to effectively model, calibrate, and predict infiltration by leveraging both field data and multifactor environmental measurements. The choice of which regression model to use depends on the complexity of the system being modelled and the available data.

Linear regression models the relationship between the infiltration rate as the dependent variable and one or more independent variables such as the initial moisture content of the soil and, rainfall intensity.

$$f(t) = a + bt \tag{1}$$

where $f(t)$ is the infiltration rate at time t, and a and b are arbitrary constants.

Multiple Linear Regression (MLR) models are used to consider multiple predictors simultaneously such as soil texture, bulk density, slope, rainfall characteristics, and so on MLR models are used for complex field conditions with multiple factors affecting infiltration [4; 7; 11].

$$f(t) = a + bx_1 + cx_2 + \cdots \tag{2}$$

where $f(t)$ is the infiltration rate at time t, a, b, c are arbitrary constants, and x_1, x_2, x_3 are variables that have some impact on infiltration.

Nonlinear regression models where used when the relationship between infiltration and its influencing factors deviates from linearity. Infiltration phenomena are inherently complex and often exhibit nonlinear behaviour, which is better captured by empirical models such as the Horton, Kostiakov, and Philip models [12-14]. These models rely on nonlinear formulations to accurately represent temporal variations in infiltration rates under varying soil and hydrological conditions.

Machine Learning-Based Regression

With the advancement of data-driven approaches, machine learning (ML) techniques have emerged as powerful tools for modelling infiltration processes. By leveraging large datasets and learning complex patterns, ML-based regression models such as random forest regression, support-vector machine (SVM), artificial neural network (ANN) models offer improved predictive accuracy compared to conventional methods.

1.2 Literature review and research objectives

Numerous empirical and analytical models have been developed to simulate infiltration behaviour, each with varying applicability depending on the soil type, initial moisture content, and climatic conditions. Vand et al. [4] conducted a comparative analysis of infiltration models including the Kostiakov, modified Kostiakov, and Philip models, as well as a novel empirical model developed at the NIT Kurukshetra campus. In addition to evaluating model performance using statistical indices, the study proposed a multiple linear regression-based empirical equation to estimate infiltration rates using influencing parameters such as soil texture, bulk density, and moisture content. This hybrid approach highlights the potential of combining empirical data with regression modelling to enhance infiltration predictions. In a broader evaluation,

Igbadun et al. [15] assessed the predictive capability of ten different infiltration models under sandy clay loam soil conditions in Samaru, Zaria. Their findings provided critical insights into model performance in semi-arid soils, and demonstrated that no single model can universally capture infiltration behaviour across all scenarios.

Gregory et al. [16] examined the application of methods such as the Horton and Green-Ampt models in relation to soil physical properties. Their study emphasised parameter sensitivity and model calibration as key to achieve reliable predictions across diverse soil types. Wei et al. [17] focused on the effect of moisture content and found that measurements of infiltration within a 5-15 % soil moisture range yielded the most consistent results when using a double ring infiltrometer. This suggests that initial moisture conditions significantly influence field measurements and model accuracy.

Ogbe et al. [18] evaluated four commonly used models, including the Kostiakov, modified Kostiakov, Philip, and Horton models, and found that the Horton model provided the best statistical fit for cumulative infiltration. Their findings affirmed that the choice of which model to use must be aligned with field data and soil characteristics for effective prediction. Su et al. [8] introduced a novel theoretical approach to infiltration modelling using the principle of least action and a variational principle. This method produced an approximate analytical solution capable of simulating soil water dynamics with high precision, especially with higher-order Taylor series expansions. In the context of land management, Atta-Darkwa et al. [6] examined the influence of tillage systems on infiltration using the Kostiakov, Philip, and Horton models in tropical climates. Their results revealed that the Kostiakov model performed best in predicting cumulative infiltration under various soil disturbance conditions, which illustrates the impact of agricultural practices on infiltration behaviour.

Marqasi et al. [7] explored the use of combined simulation techniques to enhance infiltration modelling accuracy. While their hybrid methods generally improved prediction, one model (M3SE) underperformed, which indicates that not all integrated approaches guarantee better results and that the suitability of any given method remains context-specific. Despite substantial progress in infiltration modelling, several gaps in the relevant literature have been identified. Most studies have confirmed that infiltration models must be site-specific, which limits their transferability across regions with differing soil types and hydrological conditions. Although some studies have combined empirical modelling with regression, Vand et al. [4], noted that no methods have been developed to systematically integrate of multiple influencing parameters such as slope, vegetation, porosity, and compaction into data-driven frameworks. Most existing studies have been based on short-term or point-specific field measurements.

Thus there is a need for models that incorporate spatial and temporal variability using more granular datasets.
Kieu et al. [19] demonstrated the usefulness of multivariate statistical methods (including components analysis (CA) and principle component analysis PCA) along with water quality indices (including the water pollution index WPI, and groundwater quality index GWQI) for identifying pollution sources and seasonal variability. Their findings revealed that wastewater discharge, landfilling, and seawater intrusion were major contributors to statistical approaches to hydrological problems.
Utama et al. [20] studied infiltration in the Patuha geothermal area by comparing SVM and random forest (RF) methods using Landsat 8 and Sentinel 2 imagery for land use classification. Their analysis highlighted RF as the superior model in terms of accuracy and stability for supporting sustainable water management and ecosystem conservation, especially in complex urban settings. Such applications have demonstrated the growing role of machine learning in advancing environmental monitoring aligned with sustainable development goals (SDGs). Kawther et al. [21] explored the use of water treatment sludge ash (WTSA) to improve the performance of expansive clayey soils for landfill liners. The incorporation of WTSA into bentonite-treated soils reduced hydraulic conductivity, swelling, and desiccation cracking. Their approach also meets standard liner requirements and provided enhanced durability in tropical conditions that can affect soil properties.
Although theoretical and empirical models are well-documented, fewer studies have combined physical theory, statistical regression, and machine learning approaches to capture the complexity of infiltration processes comprehensively. Based on the gaps identified and the need for further study, the following objectives were derived:

- A multiple linear regression (MLR) model was developed designed to estimate soil infiltration rates using field-measured data and relevant hydrological parameters.
- The performance of the proposed MLR model was also validated by comparing predicted infiltration values with observed field measurements of infiltration through appropriate statistical metrics.
- Additionally, a sensitivity analysis was also performed to evaluate the relative influence of various independent variables on infiltration using leave one out cross-validation.

2 Materials and methodology

2.1 Study area

The present investigation was carried out in Kerala, a coastal state located in the southern part of India, between 10,1632° N latitude and 76,6413°E longitude. The state exhibits significant topographic and climatic diversity, which strongly influences its hydrological behaviour, particularly infiltration, runoff, and groundwater recharge. Bounded by the Western Ghats to the east and the Arabian Sea to the west, Kerala is characterised by a narrow strip of land with pronounced altitudinal gradients and intense monsoonal rainfall. The spatial framework for the study was developed using Survey of India (SoI) administrative boundary shape files processed in ArcGIS 10,8 to generate the base map Figure 1. This approach ensured that the chosen locations were both representative and reproducible for infiltration modelling.

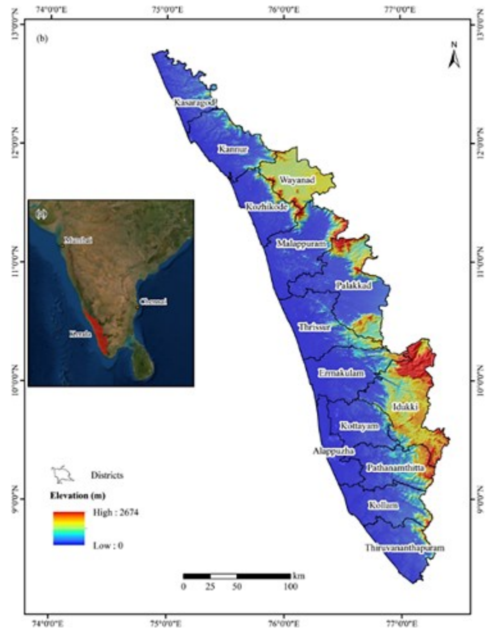


Figure 1. Location map of the study area in Kerala, India. The base map was prepared using Survey of India (SoI) administrative boundary shape files integrated in ArcGIS 10,8

Physiographically, the state is divided into three distinct zones based on elevation:

- The lowland zone (elevation < 7,6 m above mean sea level) constitutes 10,24 % of the total area and encompasses coastal plains, backwaters, and estuarine systems.
- The midland zone (7,6-76,0 m MSL) accounts for approximately 41,76 % and is marked by rolling hills and fertile valleys.
- The highland zone (elevation > 76,0 m MSL) makes up around 48,00 % of the state and includes the steep slopes of the Western Ghats, which serve as a major watershed region.

Figure 1 shows the variations in the physiographical conditions of the study area. These physiographic zones exhibit distinct hydrological regimes influenced by variations in rainfall intensity, land use, soil type, and drainage characteristics. The highland region, with steep gradients and rocky substrata, typically exhibits lower infiltration and higher runoff, whereas the midland and lowland areas show greater potential for water retention and infiltration, depending on soil permeability and vegetation cover.

To capture this hydrological variability, six representative locations were selected across the three physiographic zones of Kerala. These sites were chosen based on altitude, terrain type, and spatial distribution, and were used for field-based infiltration measurements in developing the regression model. Some details of the selected locations are summarized in Table 1.

Table 1. Study location and terrain

Study Location	Location	Terrain
Location 1	Vadavukode	Midland
Location 2	Thripunithura	Midland
Location 3	Inchathotty	Highland
Location 4	Thatekadu	Highland
Location 5	Chitoor	Lowland
Location 6	Kadamakudy	Lowland

To address potential concerns regarding spatial dependence, the study sites were deliberately selected from distinct physiographic zones including highland, midland, and lowland areas. They also focused on geographically well-separated locations to minimise the likelihood of spatial clustering effects. Although formal spatial autocorrelation tests were not performed due to the limited sample size, the independence of the observations was indirectly supported by the stratified site selection strategy.

2.2 Data collection

Field infiltration rates are known to be significantly affected by the existing soil moisture content at the time of testing. To account for this, soil moisture levels were systematically measured at all testing locations immediately before the infiltration experiments were conducted. The gravimetric oven-drying method was used to determine the initial soil moisture, following standardised protocols. Across all sampled sites, the soil moisture content was consistently found to be below 6 %, which indicate relatively dry field conditions prior to infiltration testing [4]. Comprehensive characterization of other relevant soil properties such as the percentage of clay, sit, sand, bulk density, and moisture content was performed as per IS2720:1985, and the results are compiled in Table 2. These parameters provide a critical context for interpreting infiltration behaviour and reflect the influence of inherent soil attributes on hydraulic performance in the upper soil profile.

Table 2. Secondary data

Study Location	Clay (%)	Silt (%)	Sand (%)	Bulk density (g/cc)	Moisture content (%)
Location 1	11	49	40	1,42	3,6
Location 2	13	52	35	1,55	4,2
Location 3	10	53	37	1,25	3,5
Location 4	14	48	38	1,72	2,8
Location 5	15	52	33	1,38	4,5
Location 6	12	55	33	1,68	5,5

To measure field infiltration rates, the double ring infiltrometer method was utilised at each site while adhering strictly to the procedures outlined in IS 12707:1989. The tests employed a diameter double ring infiltrometer with diameter of 30cm to ensure uniformity across sites [3]. The data from these tests, including the infiltration rate measurements, are presented in Table 3.

All laboratory and field testing methods associated with secondary data collection conformed to the guidelines in IS 2720:1985 in the interest of methodological consistency and reliability of the results. Soil samples for both moisture and other geotechnical analyses were consistently collected from a depth interval of 15 to 30cm below the ground surface. This depth selection followed the established standards for geotechnical and infiltration studies by targeting the soil horizon most relevant for near-surface hydrologic assessments.

Table 3. Field Infiltration value (cm/hr) for different locations in the study area

Time (min)/Location	1	2	3	4	5	6
5	44,83	65,87	50,55	60,45	36,38	45,21
10	34,09	57,98	48,91	55,82	31,76	38,38
15	34,09	44,73	45,76	52,38	28,22	35,76
20	25,57	40,67	38,81	48,71	25,91	30,51
25	23,01	38,34	36,28	42,68	20,76	27,66
30	23,44	34,64	32,65	36,87	18,45	25,87

35	34,09	35,67	31,36	33,21	16,87	22,65
40	20,45	33,45	29,01	34,76	15,88	20,38
45	20,45	30,56	28,76	33,87	15,71	16,76
50	23,01	29,67	27,88	26,19	14,72	16,45
55	19,60	29,67	25,82	26,54	14,44	15,76
60	22,50	29,67	24,37	24,18	13,28	15,24
65	17,04	29,01	23,87	23,87	13,19	12,76
70	22,67	27,46	22,76	25,41	12,45	12,45
75	17,04	27,21	20,97	24,39	11,96	9,65
80	17,04	24,67	19,83	23,73	11,62	8,46
85	22,16	23,85	18,57	21,65	11,45	8,46
90	21,73	23,85	17,76	20,65	10,94	8,46
95	13,64	23,85	17,21	18,42	10,84	8,46
100	15,34	23,85	16,49	18,52	10,80	8,46
105	15,34		15,76	15,38	10,72	
110	15,34		13,69	15,38	10,72	
115	15,34		12,45	15,38	10,72	
120		---	12,21		10,72	---
125	---		12,21	---		
130			12,21		---	
135			12,21			

3 Results and discussion

3.1 Comparative validation of observed infiltration across varying terrains

Randomly selected field infiltration measurements were carried out across multiple locations and categorised according to distinct terrain classifications. A comparative analysis of the observed infiltration rates revealed a notable consistency among sites within the same terrain category, which highlight the significant influence of physiographic characteristics on infiltration behaviour [19]. To statistically assess the degree of agreement among the infiltration values across similar terrains, one-way ANOVA was employed to test for significant differences, whereas Pearson correlation analysis was used to evaluate the strength of associations between paired terrain-specific observations [19]. An alternative approach to understanding infiltration behaviour involves the development of multiple infiltration models using both empirical and statistical techniques. These models can then be systematically evaluated by comparing their predicted infiltration values against observed field data. Such comparative analysis enables the identification of an optimal model that represents the infiltration process most accurately under varying field conditions [22; 23]. The predictive performance of each model can be quantitatively assessed by applying goodness of fit indicators such as the coefficient of determination (R^2), standard error (SE), and decision factor (DF) [19; 20]. This approach not only aids in model selection but also enhances the reliability of infiltration estimates for hydrological planning, soil-water management, and runoff prediction [21; 24; 25].

3.2 Multi linear regression approach

Direct field measurement of infiltration rates is often labour-intensive, time-consuming, and logistically challenging, especially in terrains with heterogeneous physiographic conditions. Such variability in soil texture, structure, and moisture regimes necessitates the development of robust predictive models to complement or replace extensive experimental field work. In this study, multiple linear regression (MLR) framework was employed to derive empirical equations capable of predicting infiltration rates across diverse landscape conditions. The model incorporated key infiltration-influencing parameters, namely the percentages of silt (*Si*), clay (*C*), and sand (*Sa*), as well as bulk density (*B*), initial moisture content (*W*) and time (*T*) which have been widely recognised in the literature as primary determinants of infiltration behaviour. Secondary datasets compiled from multiple geospatially distinct field locations shown in Table 2 served as the basis to calibrate and validate the model [4]. The linear functional relationship can be represented as given in Equation (3):

$$f(t) = aT^u + C^v + Si^w + Sa^x + B^y + W^z \tag{3}$$

where *a* is a proportionality constant and *u*, *v*, *w*, *x*, *y*, *z* are coefficients of the equation. Applying a logarithmic transformation to both sides of the equation converts it into a linear form as in Equation (4):

$$\log f(t) = loga + u\ logT + v\ logC + w \log Si + x \log Sa + y \log B + z\ logW \tag{4}$$

Model development and statistical analysis were conducted using the IBM SPSS software platform. The MLR model coefficients were evaluated for statistical significance, multicollinearity was assessed using variance inflation factors (VIF), and the overall adequacy of the model was tested using the coefficient of determination (R^2), adjusted R^2, and the standard error of the estimate. All predictor variables in the MLR model exhibited VIF values below 5. This threshold is widely recognised in regression diagnostics as an indicator of acceptable collinearity, as values exceeding 5 often suggest moderate correlation among predictors, whereas values greater than 10 are typically considered problematic. The fact that all predictors in this study remained well below this conservative cutoff confirms that multicollinearity was not a concern. Consequently, the stability of the regression coefficients is ensured which allows the model to reliably capture the independent contributions of each soil and hydrological parameter without distortion arising from inter-variable dependencies [25; 26]. To develop the multiple linear regression model, log *f*(*t*) was adopted as the dependent variable, while the selected primary and secondary parameters served as independent variables [21; 22]. The model parameters were estimated using the least squares method within the IBM SPSS statistical analysis platform. MLR results output by the SPSS platform are shown in Table 4.

Table 4. SPSS output interface for multiple linear regression results

Model	Unstandardized coefficients		Standardized coefficients Beta	t	Sig.	Correlations			Collinearity statistics	
	b	std. error				zero order	partial	part	tolerance	VIF
constant	0,822	9,674	---	0,085	0,946	---	---	---	---	---
clay	0,232	2,051	0,199	0,113	0,928	0,107	0,112	0,092	0,214	4,669
silt	−0,232	3,559	−0,094	−0,065	0,958	−0,294	−0,065	−0,053	0,323	3,093
bulk density	−0,733	3,199	−0,284	−0,229	0,857	−0,232	−0,223	−0,187	0,432	2,314
moisture content	−0,496	0,880	−0,492	−0,563	0,673	−0,411	−0,491	−0,459	0,871	1,149
time interval	−0,281	1,687	−0,172	−0,167	0,895	−0,113	−0,165	−0,136	0,628	1,593

The standardised coefficients obtained from the regression analysis were used to formulate an empirical equation to estimate the infiltration rate as shown in Equation (5). This derived relationship offers a practical and scalable approach for infiltration estimation, with potential applicability across a range of hydrological modelling applications.

$$f(t) = \frac{0{,}822 \cdot C^{0{,}232}}{Si^{0{,}232} \cdot B^{0{,}733} \cdot W^{0{,}496} \cdot T^{0{,}281}} \tag{5}$$

3.3 Comparison of regression model with field data

To account for the estimation error in infiltration rates derived from the regression equation, a calibration constant, referred to as the infiltration coefficient (*K*), was incorporated into the formulation of the model. The coefficient was determined by computing the ratio of the field infiltration rate (*field (f)*) to the model-predicted rate (*model (f)*) for each measurement location based on the Equation (6):

$$K = \frac{field\ (f)}{model\ (f)} \tag{6}$$

The mean value of these ratios, after screening for outliers, was adopted as the representative *K* for model adjustment. The calibrated infiltration rate was then obtained as per Equation (7):

$$f(t) = K \cdot \frac{0{,}822 \cdot C^{0{,}232}}{Si^{0{,}232} \cdot B^{0{,}733} \cdot W^{0{,}496} \cdot T^{0{,}281}} \tag{7}$$

Although the silt fraction exhibited a negative coefficient in the regression model, it was retained because of both statistical and hydrological considerations. Statistical diagnostics including VIF and tolerance values, confirmed that the parameter was not collinear with sand or clay fractions, which clarifies its independent contribution to the model. From a hydrological perspective, higher silt content is known to reduce infiltration by clogging soil pores and limiting hydraulic conductivity, which support the observed negative relationship. A similar approach can be applied to other governing factors such as bulk density, moisture content and time interval.

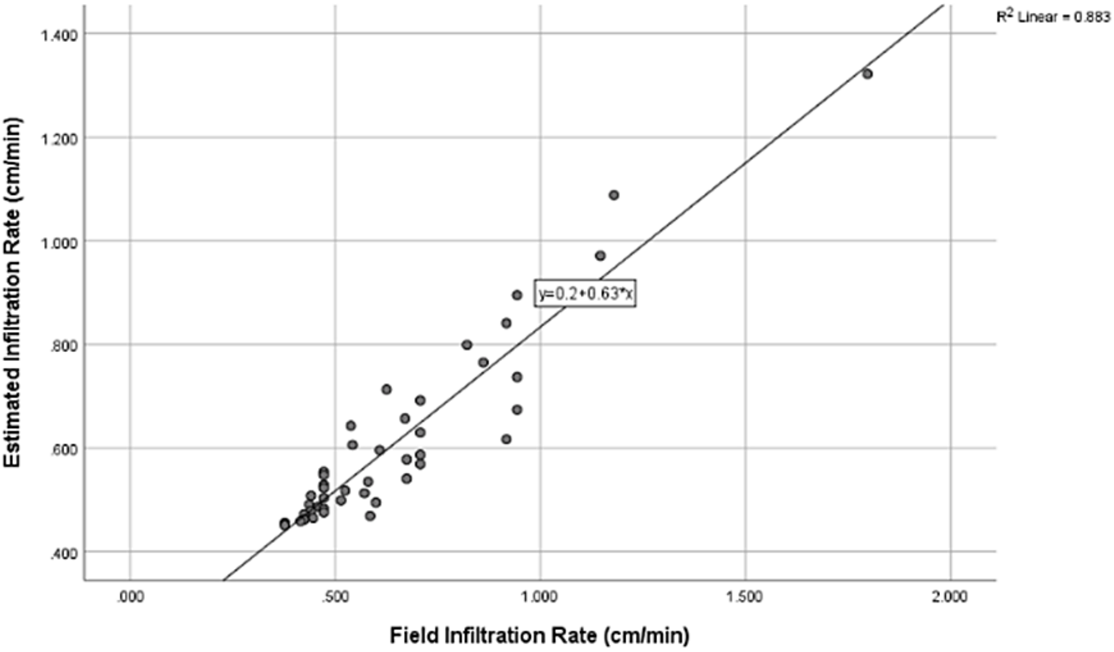
For the present study area, the infiltration coefficient (*K*) was found to vary between 5,7 and 8,9. The *K* values corresponding to individual locations are listed in Table 5.

Table 5. Infiltration constant (*K*) for various locations

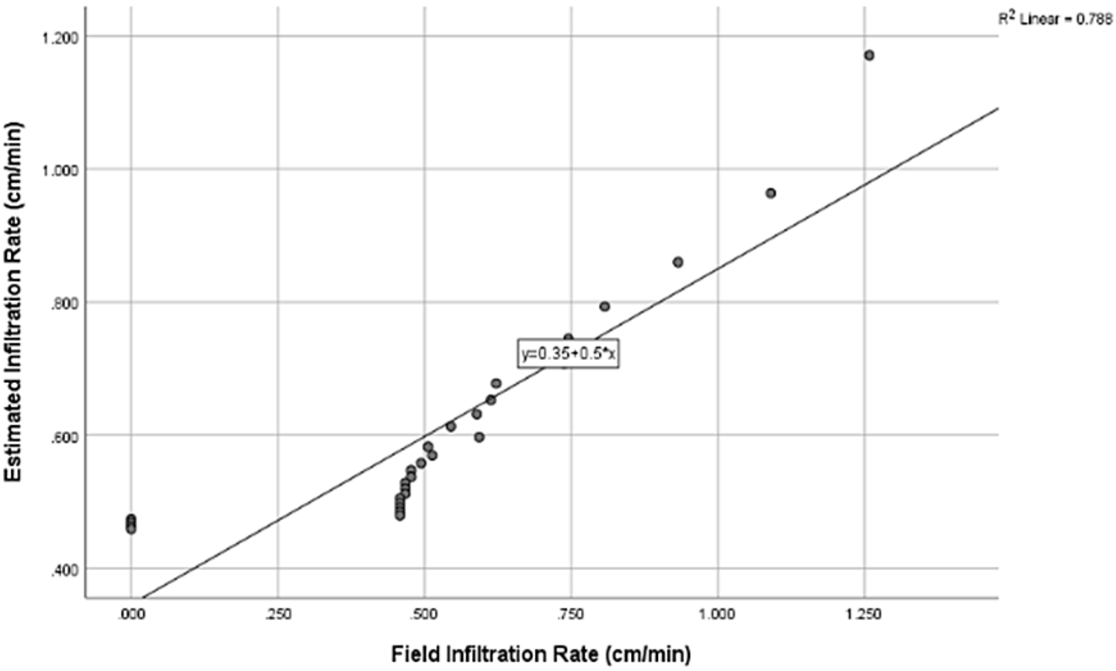
Study location	Infiltration coefficient *K*
Location 1	8,9
Location 2	6,9
Location 3	8,5
Location 4	7,2
Location 5	5,7
Location 6	7,4

The variation in the infiltration coefficient (*K*) across locations reflects site-specific soil and terrain attributes. For example, the highest *K* value observed at Location 1 corresponds to coarse-textured soils with higher sand content, lower bulk density, and better drainage conditions that favour greater infiltration. In contrast, locations with higher silt and clay fractions and relatively compacted soil profiles exhibited lower K values, which highlights the role of soil texture and structure in governing infiltration dynamics.

The corrected performance of the model was subsequently validated against independent field datasets using statistical indicators such as one-way ANOVA and the coefficient of determination (R^2). The ANOVA test was used to compare the mean infiltration rates predicted by the calibrated model with those measured in the field under the null hypothesis that no significant difference existed between the two datasets. A p-value greater than 0,05 indicated statistical agreement between predicted and observed means, confirming the adequacy of the calibration. Additionally, the coefficient of determination (R^2) was calculated to quantify the proportion of variance in the observed infiltration rates as explained by the model. The results for various locations are shown in Figures 2, 3 and 4.

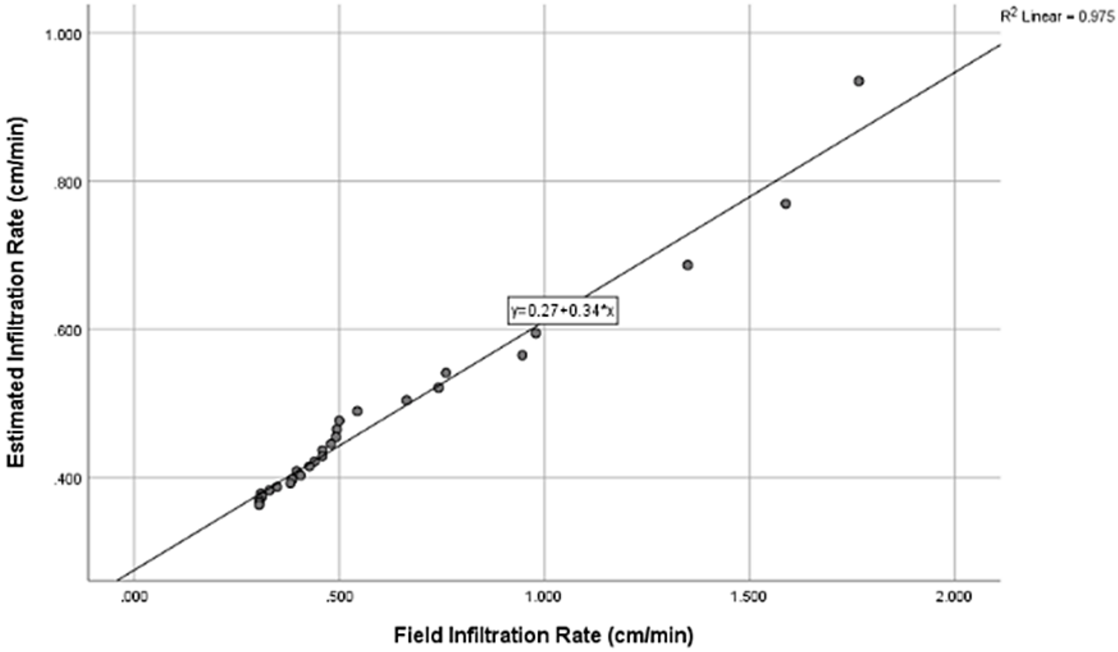


a)

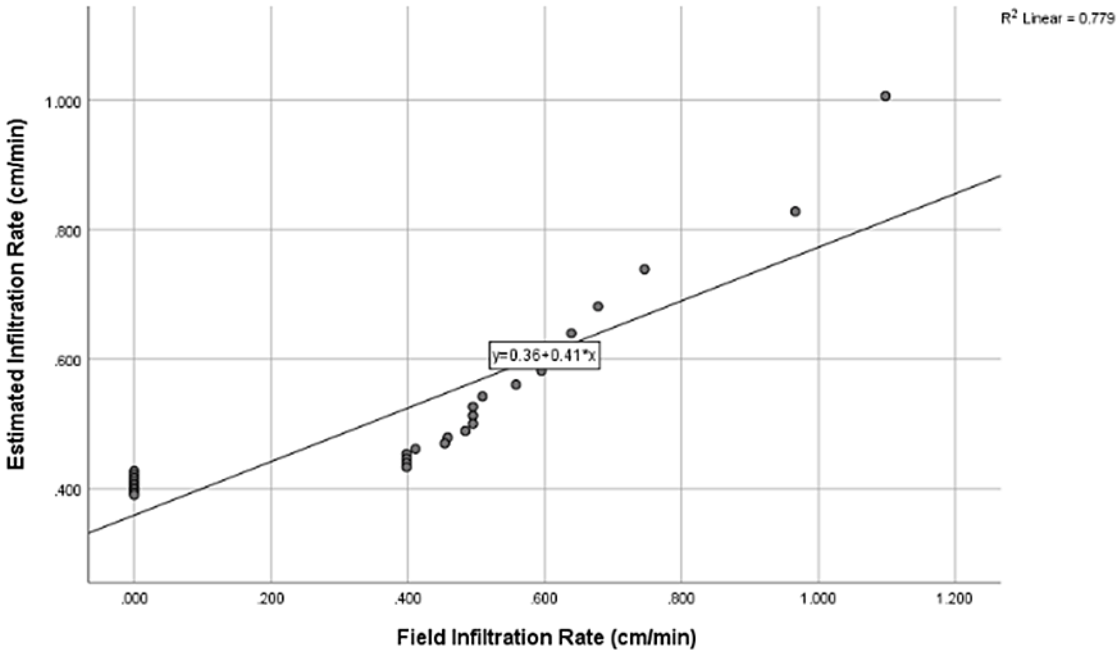


b)

Figure 2. Coefficient of determination for: a) location 1, and b) location 2

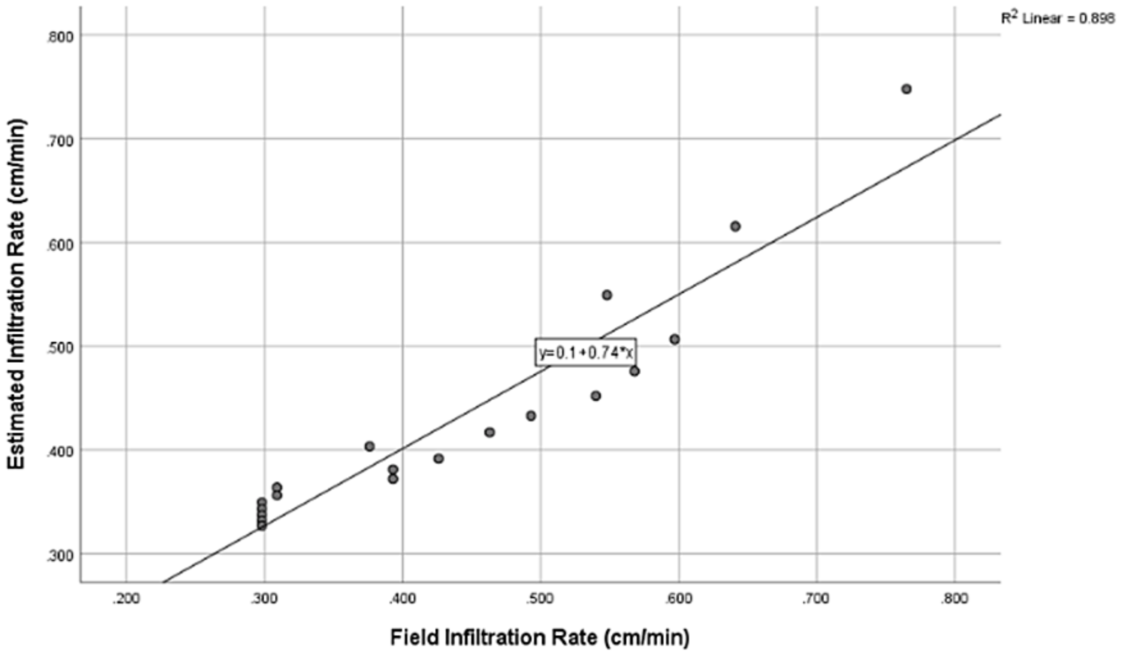


a)

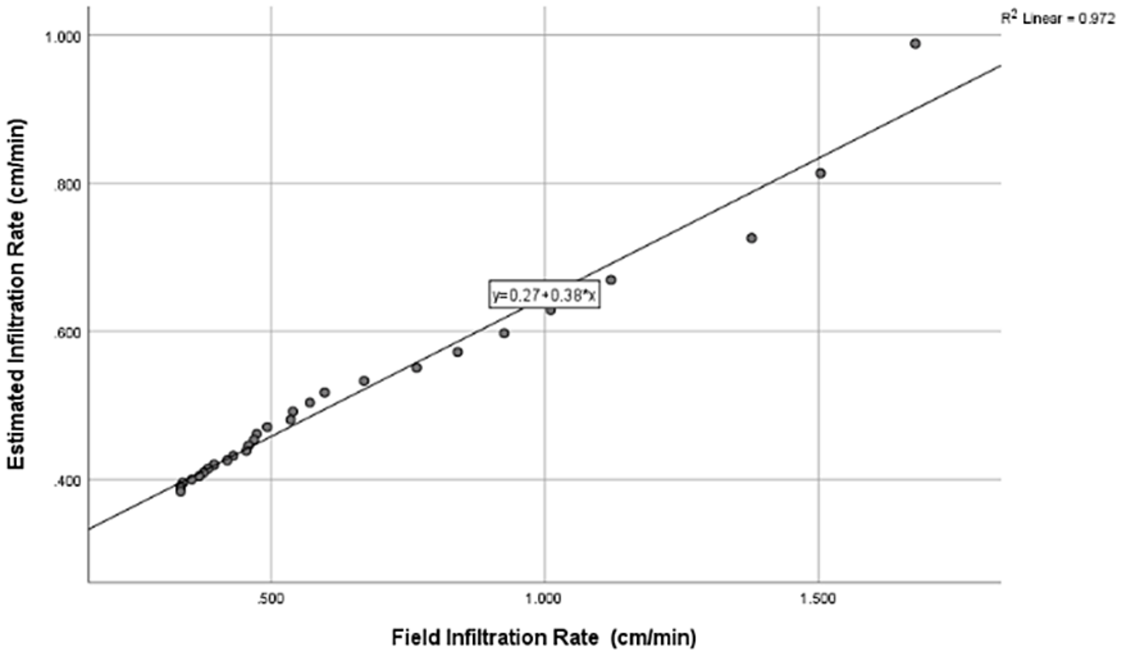


b)

Figure 3. Coefficient of determination for: a) location 3, and b) location 4



a)



b)

Figure 4. Coefficient of determination for; a) location 5, and b) location 6

Across the different study locations, the coefficient of determination (R^2) between the field-measured infiltration rates and the values estimated by the calibrated multiple linear regression (MLR) model ranged from 0,779 to 0,975. These values indicate that the model was capable of explaining between 77,9 % and 97,5 % of the variability in the observed infiltration rates [23; 24]. In the context of environmental and hydrological modelling, R^2 values above 0,75 are generally regarded as indicative of strong predictive performance, particularly when applied to field-scale datasets that inherently contain measurement variability and natural heterogeneity [7]. This level of agreement demonstrates both the predictive reliability and the robustness of the developed infiltration model across diverse soil textures and physiographic settings. Based

on these results, the model can be considered suitable for practical applications in infiltration assessment.
In addition to R² values, the performance of the model was evaluated in terms of root mean square error (RMSE) and mean absolute error (MAE) to provide a more comprehensive assessment. The RMSE values across different locations ranged between 0,82 and 1,47 cm/hr, which indicate that the magnitude of prediction errors remained within acceptable limits. Similarly, the MAE values varied from 0,61 to 1,12 cm/hr, reflecting the average deviation between observed and predicted infiltration rates. Together with R² (0,779-0,975), these indicators confirm that the developed MLR model achieved both strong explanatory power and reliable predictive accuracy across diverse physiographic conditions. The measures of the performance of the model derived from the SPSS platform are provided in Table 6.

Table 6. Model performance metrics

Location	Coefficient of determination R^2	Root Mean Square Error RMSE	Mean Absolute Error MAE
1	0,883	1,10	0,89
2	0,788	1,51	1,20
3	0,975	0,82	0,61
4	0,779	1,47	1,12
5	0,898	1,32	1,05
6	0,972	0,81	0,60

3.4 Model response to dependent parameters

Leave-one-out cross-validation (LOOCV) is a statistical validation technique used to assess the predictive performance and generalisability of a model. In LOOCV, a dataset containing n observations is partitioned such that, one observation is excluded from training and used solely for testing in each iteration. This process was repeated n times, leaving different observations each time. The overall prediction error was computed based on the coefficient of correlation (CC) and RMSE of the statistical parameters. Various dependent parameters considered for the LOOCV included the percentage of silt (*Si*), clay (*C*), sand (*Sa*), bulk density (*B*), initial moisture content (*W*) and time (*T*) [27; 28; 29]. Multiple linear regression models were compared with field data for various locations to identify the parameter that influenced regression the most [4; 7]. The data set for location 1 is provided in Table 7.

Table 7. LOOCV approach on dependent variables

Excluded parameter	Root Mean Square Error RMSE	Coefficient of correlation CC
Nil (All parameters considered)	0,095	0,91
Percentage of clay *C*	0,090	0,87
Percentage of sand *Sa*	0,080	0,84
Percentage of silica *Si*	0,070	0,89
Time *T*	0,180	0,64
Bulk density *B*	0,060	0,87
Moisture content *W*	0,040	0,89

A similar LOOCV based approach was applied across various locations within the study area to identify the parameters that exerted the greatest influence on infiltration. The results revealed that during LOOCV analysis in which time was excluded, the statistical evaluation yielded the highest RMSE and the lowest CC, which indicates that time (*T*) is the most influential predictor within the MLR framework. Furthermore, the analysis suggested that the

soil texture parameters of silt (*Si*), clay (*C*), and sand (*Sa*) exhibited comparable levels of influence on the infiltration estimates generated by the MLR equations.
Thus the LOOCV sensitivity analysis highlighted time (T) as the most influential predictor of infiltration, with its exclusion resulting in the highest RMSE and lowest correlation coefficients across locations. These findings have important implications for practical hydrological applications. Because infiltration is strongly time-dependent, the duration and intensity of rainfall events play a decisive role in determining infiltration capacity, surface runoff generation, and groundwater recharge. For example, in irrigation scheduling, accounting for the temporal dynamics of infiltration can help optimise watering intervals, reduce deep percolation losses, and improve water-use efficiency. Similarly, in rainfall–runoff modelling and storm water management, the timing of infiltration directly influences peak discharge estimation, flood detention design, and effectiveness of recharge structures. Thus, the identification of time as a dominant variable validates the robustness of the regression framework. Moreover, it also underscores its utility in guiding sustainable water resource management, irrigation planning, and flood risk reduction strategies.

3.5 Scalability novelty and limitations

This study introduced a calibrated MLR framework that integrates both primary field infiltration measurements and secondary soil property datasets across heterogeneous physiographic zones. Unlike conventional infiltration models that are limited to site-specific or single-parameter inputs, the proposed approach combines multiple influential parameters such as soil texture components, bulk density, moisture content, and time into a unified predictive equation. The inclusion of a location-specific infiltration coefficient (*K*) enhanced the adaptability and accuracy of the model across diverse types of terrain. Scalability is ensured through the model's reliance on readily measurable soil and hydrological variables to support, its application in various geographic and climatic contexts without the need for extensive, repeated field testing. Furthermore, the validation using ANOVA, R^2 analysis, and LOOCV established its robustness for broader hydrological and water resources management applications.
Future studies incorporating denser spatial datasets could further strengthen this framework by explicitly quantifying spatial autocorrelation using geostatistical approaches. Future research using larger spatial datasets could benefit from incorporating geostatistical techniques to account for spatial dependence explicitly.
Measurement errors in bulk density and soil texture, arising from sampling and laboratory methods, may introduce uncertainties in infiltration predictions. In the present study, such variability was minimised by adhering to the IS 2720 (1985) standards during soil testing. The LOOCV analysis further indicated that although soil texture parameters and bulk density influenced the model, their impact was less critical than that of time, which emerged as the most dominant factor. Nonetheless, future research could reduce this uncertainty by incorporating repeated laboratory measurements, improved sampling protocols, or advanced soil characterization techniques.
Although the findings of this study demonstrate the robustness and practical utility of the MLR-based infiltration model, it is important to acknowledge the growing role of machine learning (ML) techniques such as RF, SVM, and ANN models in hydrological modelling. Future research may explore hybrid approaches that integrate the interpretability of MLR with the predictive power of ML algorithms, particularly when larger datasets become available. Such integration has the potential to further enhance infiltration modelling accuracy and provide a balance between empirical transparency and computational complexity.

4 Conclusions

This study developed and validated a MLR model for estimating infiltration rates across heterogeneous physiographic zones in Kerala, India. By integrating the key soil parameters of the percentages of silt (*Si*), clay (*C*), sand (*Sa*), bulk density (*B*), initial moisture content (*W*),

and time (T) and applying a logarithmic transformation, the model effectively captured the relationship between infiltration and its influencing factors. Calibration through an infiltration coefficient (K), ranging from 5,7 to 8,9 substantially improved predictive accuracy.
Validation using one-way ANOVA and coefficient of determination (R^2) confirmed strong model agreement with field measurements (R^2 = 0,779-0,975), highlighting its robustness under diverse soil and terrain conditions. Sensitivity analysis via LOOCV identified time (T) as the most influential predictor, whereas soil texture parameters exhibited a comparable influence on infiltration predictions.
The developed MLR framework offers a practical, scalable, and statistically validated tool for estimating infiltration. Thus, this method reduces the need for labour-intensive field measurements. Its adaptability across physiographically varied terrain makes it suitable for applications in hydrological modelling, watershed management, irrigation planning, and flood risk assessment. Although the MLR framework developed in this study was calibrated for Kerala, its structure based on fundamental soil and hydrological parameters offers potential applicability to other regions with comparable physiographic conditions. However, due to regional differences in rainfall regimes, vegetation, and land management practices, site-specific recalibration using representative field infiltration data is recommended to ensure reliable performance outside the area considered in the present work. This study highlights the robustness and practical utility of the MLR-based infiltration model while recognizing the growing relevance of machine learning (ML) methods in hydrological modelling. Future studies should focus on hybrid MLR–ML approaches to combine interpretability with predictive accuracy, particularly when larger datasets are available. The proposed MLR framework can also serve as a foundation for more advanced modelling. Future extensions may involve incorporating non-linear functions to better represent temporal dynamics, or using the identified influential parameters (time, bulk density, and soil texture) as input features for machine learning algorithms such as RF, SVM, or ANN models. Such adaptations would enhance the predictive capability of the proposed method under more complex field conditions while retaining the model’s practical applicability.

References

[1] Fu, Q. et al. A new infiltration model for simulating soil water movement in canal irrigation under laboratory conditions. *Agricultural Water Management*, 2019, 213, pp. 433-444. https://doi.org/10.1016/j.agwat.2018.10.021

[2] Zhu, P. et al. Variation in soil infiltration properties under different land use/cover in the black soil region of Northeast China. *International Soil and Water Conservation Research*, 2023, 12 (2), pp. 379-387. https://doi.org/10.1016/j.iswcr.2023.07.007

[3] Shamohammadi, S.; Kadkhodahosseini, M.; Ostad-Ali-Askari, K. New procedure to estimate soil field capacity based on double ring infiltration data. *Ain Shams Engineering Journal*, 2023, 14 (11), 102511. https://doi.org/10.1016/j.asej.2023.102511

[4] Vand, A. S.; Sihag, P.; Singh, S.; Zand, M. Comparative Evaluation of Infiltration Models. *KSCE Journal of Civil Engineering*, 2018, 22 (10), pp. 4173-4184. https://doi.org/10.1007/s12205-018-1347-1

[5] Mao, L. et al. A new method to estimate soil water infiltration based on a modified Green–Ampt model. *Soil and Tillage Research*, 2016, 161, pp. 31-37. http://dx.doi.org/10.1016/j.still.2016.03.003

[6] Atta-Darkwa, T. et al. Performance evaluation of infiltration models under different tillage operations in a tropical climate. *Scientific African*, 2022, 17, e01318. https://doi.org/10.1016/j.sciaf.2022.e01318

[7] Sihag, P. et al. Modelling of infiltration using artificial intelligence techniques in semi-arid Iran. *Hydrological Sciences Journal*, 2019, 64 (13), pp. 1647-1658. https://doi.org/10.1080/02626667.2019.1659965

[8] Su, L.; Wang, J.; Qin, X.; Wang, Q. Approximate solution of a one-dimensional soil water infiltration equation based on the Brooks-Corey model. *Geoderma*, 2017, 297, pp. 28-37. https://doi.org/10.1016/j.geoderma.2017.02.026

[9] Fu, Q. et al. A new infiltration model for simulating soil water movement in canal irrigation under laboratory conditions. *Agricultural Water Management*, 2019, 213, pp. 433-444. https://doi.org/10.1016/j.agwat.2018.10.021

[10] Mao, L. et al. A new method to estimate soil water infiltration based on a modified Green–Ampt model. *Soil and Tillage Research*, 2016, 161, pp. 31-37. http://dx.doi.org/10.1016/j.still.2016.03.003

[11] Yang, M.; Zhang, Y.; Pan, X. Improving the Horton infiltration equation by considering soil moisture variation. *Journal of Hydrology*, 2020, 586, 124864. https://doi.org/10.1016/j.jhydrol.2020.124864

[12] Ma, Y. et al. Modeling water infiltration in a large layered soil column with a modified Green–Ampt model and HYDRUS 1D. *Journal of Computers and Electronics in Agriculture*, 2019, 71 (1), pp. S40-S47. http://dx.doi.org/10.1016/j.compag.2009.07.006

[13] Cui, G.; Zhu J. Prediction of unsaturated flow and water backfill during infiltration in layered soils. *Journal of Hydrology*, 2018, 557, pp. 509-521. https://doi.org/10.1016/j.jhydrol.2017.12.050

[14] Zhu, P. et al. Variation in soil infiltration properties under different land use/cover in the black soil region of Northeast China. *International Soil and Water Conservation Research*, 2024, 12 (2) pp. 379–387. https://doi.org/10.1016/j.iswcr.2023.07.007

[15] Igbadun, H.; Othman, M.; Ajayi, A. Performance of Selected Water Infiltration Models in Sandy Clay Loam Soil in Samaru Zaria. *Global Journal of Researches in Engineering: J General Engineering*, 2016, 16 (4), pp. 8-14.

[16] Gregory, J. H.; Dukes, M. D.; Miller, G. L.; Jones, P. H. Analysis of Double-Ring Infiltration Techniques and Development of a Simple Automatic Water Delivery System. *Applied Turfgrass Science*, 2005, 2 (1), pp. 1-7. https://doi.org/10.1094/ATS-2005-0531-01-MG

[17] Wei, L. et al. Experimental Investigation of Relationship between Infiltration Rate and Soil Moisture under Rainfall Conditions. *Water*, 2022, 14 (9), 1347. https://doi.org/10.3390/w14091347

[18] Ogbe, V. B.; Jayeoba, O. J.; Ode, S. O. Comparison of Four Soil Infiltration Models on A Sandy Soil in Lafia, Southern Guinea Savanna Zone of Nigeria. *Production Agriculture and Technology*, 2011, 7 (2), pp. 116-126.

[19] Kieu, L. D.; Nguyen, P. Q. Groundwater Quality Assessment in the Middle-Upper Pleistocene Aquifer. *Civil Engineering Journal*, 2024, 10 (7), pp. 2357-2369. http://dx.doi.org/10.28991/CEJ-2024-010-07-018

[20] Utama, W. et al. Towards Improving Sustainable Water Management in Geothermal Fields: SVM and RF Land Use Monitoring. *Journal of Human, Earth, and Future*, 2024, 5 (2), pp. 216-242.

[21] Al-Soudany, K. Y. H.; Fattah, M. Y.; Rahil, F. H. Enhancement of Expansive Soil Properties by Water Treatment Sludge Ash in Landfill Liners. *Civil Engineering Journal*, 2024, 10 (11), pp. 3508-3530. http://dx.doi.org/10.28991/CEJ-2024-010-11-04

[22] David, S. J.; Rahiman, A.; Subha, V. A statistical approach towards performance of infiltration models with field infiltration value. *Physics and Chemistry of the Earth*, 2025, 140, 103972. https://doi.org/10.1016/j.pce.2025.103972

[23] David, S. J.; Rahiman, A.; Subha, V. A Novel Statistical Approach for Infiltration Model Analysis. *Advanced Modelling and Innovations in Water Resources Engineering*, 2021, 176, pp. 577-583. https://doi.org/10.1007/978-981-16-4629-4_39

[24] Winner, L. *Applied Statistical Methods*. USA: Department of Statistics, University of Florida, 2009.

[25] Ratner, B. A Closer Look The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, 2009, 17, pp.139-142. https://doi.org/10.1057/jt.2009.5

[26] Sawyer, S. F. Analysis of Variance: The Fundamental Concepts. *Journal of Manual & Manipulative Therapy*, 2009, 17 (2), pp. 27-38. https://doi.org/10.1179/jmt.2009.17.2.27E

[27] Marqasi, M. et al. Simulation model of water infiltration in soil using combination technique. *Ain Shams Engineering Journal*, 2023, 14 (11), 102550. https://doi.org/10.1016/j.asej.2023.102550

[28] O'brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 2007, 41 (5), pp.673-690. https://doi.org/10.1007/s11135-006-9018-6

[29] Kunter, M. *Applied Linear Statistical Models*. 5th Edition, USA: McGraw-Hill/Irwin, 2005.