www.czasopisma.pan.pl PAN www.journals.pan.pl

JOURNAL OF WATER AND LAND DEVELOPMENT
e-ISSN 2083-4535

Polish Academy of Sciences (PAN), Committee on Agronomic Sciences
Section of Land Reclamation and Environmental Engineering in Agriculture
Institute of Technology and Life Sciences (ITP)

JOURNAL OF WATER AND LAND DEVELOPMENT
2020, No. 47 (X–XII): 77–88
https://doi.org/10.24425/jwld.2020.135034

Available (PDF): http://www.itp.edu.pl/wydawnictwo/journal; http://journals.pan.pl/jwld

Received 22.01.2020
Reviewed 28.03.2020
Accepted 04.06.2020

Determination of soil infiltration rate equation based on soil properties using multiple linear regression

Donny HARISUSENO ✉, Evi N. CAHYA

University of Brawijaya, Faculty of Engineering, Water Resources Engineering Department, MT. Haryono Street No. 167, 65145, Malang, Indonesia

For citation: Harisuseno D., Cahya E.N. 2020. Determination of soil infiltration rate equation based on soil properties using multiple linear regression. Journal of Water and Land Development. No. 47 (X–XII) p. 77–88. DOI: 10.24425/jwld.2020.135034.

Abstract

Infiltration process plays important role in water balance concept particularly in runoff analysis, groundwater recharged, and water conservation. Hence, increasing knowledge concerning infiltration process becomes essential for water manager to gain an effective solution to water resources problems. This study employed multiple linear regression for estimating infiltration rate where the soil properties used as the predictor variable and measured infiltration rate as the response variable. Field measurement was conducted at sixteen points to obtain infiltration rate using double ring infiltrometer and soil properties namely soil porosity, silt, clay, sand content, degree of saturation, and water content. The result showed that measured infiltration rate had an average initial infiltration rate (f_0) of 6.92 mm·min^{-1} and final infiltration rate (f_c) of 1.49 mm·min^{-1}. Soil porosity and sand content showed a positive correlation with infiltration rate by 0.842, 0.639, respectively, while silt, clay, water content, and degree of saturation exhibited a negative correlation by –0.631, –0.743, –0.66 and –0.49, respectively. Three types of regression equations were established based on type of soil properties used as predictor variables. The model performance analysis was conducted for each equation and the result shows that the equation with five predictor variables $f_{\mathrm{MLR\_3}}$ = – 62.014 + 1.142 soil porosity – 0.205 clay, – 0.063 sand – 0.301, silt + 0.07 soil water content with R^2 (0.87) and Nash–Sutcliffe (0.998) gave the best result for estimating infiltration rate. The study found that soil porosity contributes mostly to the regression equation that indicates great influence in controlling soil infiltration behavior.

Key words: *infiltration rate, model performance, multiple linear regression, soil property*

INTRODUCTION

Increasing knowledge on infiltration processes and its influencing variables become very essential to have a better understanding concerning hydrological processes including rainfall-runoff relationship, groundwater, and water conservation in a basin. As a part of processes in the hydrologic cycle, infiltration seems to play important role in water balance concept particularly in transformation of rainfall into runoff. From the perspective of environmental view, infiltration becomes a fundamental process concerning water conservation where most of recharge of groundwater comes from portion of rainfall that infiltrated into ground soil. Most of the hydrological models required a rigorous description on infiltration process since it plays as major loess that control rainfall-runoff relationship in a basin and the development of appropriate, reliable, and sustainable water resources management scheme [RASHIDI *et al.* 2014]. Many water resources practices employed infiltration concepts such as irrigation design, water conservation, rainfall-runoff relationship, and groundwater replenishment [MAO *et al.* 2016]. Thus, the efficiency and effectiveness design of water resources system depends on to what extend the accuracy of infiltration is estimated since the volume of infiltration becomes the main input in the design of water resources system analysis. Having a better knowledge concerning infiltration process strongly relates to adequate knowledge of soil properties as the main factor that controlling infiltration process. Many variables in basin such as topography, soil properties, climate, and landuse/landcover were recognized as influential factor that drives the variability of infiltration rate [NIE *et al.*

2017]. RASHIDI *et al.* [2014] confirmed that the property of soil including physical, chemical, and biological along with initial soil moisture played an essential role to control infiltration process. The prediction of soil infiltration rate generally deals with the adoption of an equation that involves some soil properties as predictor variable and the determination of a numerical constant as parameter. Studies concerning infiltration rate models have been conducted by some researchers and widely applied in many infiltration studies. Some of the models included in physical and empirical-based infiltration models [GREEN, AMPT 1911; HORTON 1940; KOSTIAKOV 1932; PHILIP 1956]. However, those models are unsatisfactory for practical field application due to logistical reasons in the non-linear least sum of squares technique to derive the model parameters adopted [NIE *et al.* 2017]. MUBARAK *et al.* [2010] adopted geostatistical techniques approach to predict soil infiltration rate, however, the method needs a large number of measuring points. Some researchers attempt to use a statistical approach by developing multiple regression analyses that describe the relationship between infiltration rate with soil properties as predictor variables. They conducted research to determine optimum soil infiltration rate model based on some physical properties of soil using multiple linear regression analysis where moisture content, bulk density, particle density, texture and organic carbon content as predictor variables. VAN DE GENACHTE *et al.* [1996] considered the soil properties consist of texture, organic carbon content, dry bulk density, initial moisture content, and root content and analysing their influence on behaviour of infiltration rate. DEWIDAR *et al.* [2019] performed a comparative analysis to examine the efficiency of artificial neural network, fuzzy logic, and multiple linear regression for predicting infiltration rate. RASHIDI *et al.* [2014] classified several multiple linear regression based on the number of soil properties used as a predictor variable to estimate the infiltration rate where silt content and clay content, bulk density, organic matter, and moisture content were used as the predictor variables. From the previous researches mentioned above, it could be known that still a few studies considered to include soil porosity and degree of saturation as a predictor variable and examine the influence of them on soil infiltration rate behaviour. The influence of soil porosity and degree of saturation on infiltration process was explained by HARISUSENO *et al.* [2019] who used the soil porosity and degree of saturation on runoff time of concentration which directly related to infiltration process. CZYŻYK and ŚWIERKOT [2017] attempted to discuss the soil porosity along with topographical slope on infiltration process. However, those researches did not discuss the soil porosity and degree of saturation as a predictor variable that influence infiltration rate in a quantitative relationship such as in an equation of multiple linear regression, they only focus on qualitative explanation of soil properties on infiltration process.

In the view above, it is necessary to conduct a study concerning how the relationship of soil porosity and degree of saturation along with other soil properties on the infiltration process using the multiple linear regression analysis. The present study attempts to examine the relationship between soil properties and infiltration rate using the multiple linear regression analysis and to determine the optimum infiltration rate equation by comparing the estimated infiltration rate with the measured one derived from double ring infiltrometer.

MATERIALS AND METHODS

STUDY AREA AND FIELD MEASUREMENT

This study was conducted at the University of Brawijaya campus located in Malang City, East Java Province, Indonesia. Geographically, the campus is situated between 7°57’6.79” to 7°57’15.74” N latitude and 112°36’43.23” to 112°36’54.51” E longitude. The campus has 22.04 ha of area dominated by bare and vegetated land approximately 21.00 ha and the remaining 1.04 ha is building area. The study was conducted in eight locations spread throughout the University of Brawijaya as shown in Figure 1, among others Faculty of Agriculture (B). Location of each infiltration measurement point was marked using the Global Positioning System device. Detail information of each location of measurement points such as altitude, longitude, and latitude was also recorded.

The field measurement was carried out from May 2019 to September 2019 comprise of infiltration measurement and soil sampling at each point of measurement. Soil sampling was conducted at each location around the measurement point of infiltration, while the measurement of the rate of infiltration is performed 2 times at different points at each location thus, there are sixteen measurement data at the study location. Refer to the general rule of statistical, the total number of data include in a small sample size that likely affects the accuracy of the predicted value. In the present study, the accuracy of prediction was designed at coefficient of determination (R^2) 0.75–0.80 and desired confidence interval width (ω) 0.5, accordingly the recommended sample size was ranging from 16–19 samples [HAIR *et al.* 2019]. It is thus the number of sample size used in this study remain in the range of acceptable size.

The land surface which owing grassland cover with an average height of 2–3 cm was selected as the points of measurement considering land cover similarity. The soil physic properties measured in this study consist of water content [JIANG *et al.* 2017], porosity [LIPIEC *et al.* 2006], degree of saturation [SUN *et al.* 2010], and soil texture [SINHA, SINGH 2016] considering that these soil properties influence on infiltration rate. The laboratory analysis of water content and porosity was performed on the disturbed sample while the soil grain size analysis was obtained from the sieve and hydrometer analysis.

Field measurement of infiltration rate was carried out using a Turf-tec Infiltrometer Model of the IN2-W (Photo 1). Turf-tec Infiltrometer Model IN2-W is a double-ring infiltrometer which consists of two concentric metal cylindrical ring, inner and outer ring that owing 6.03 cm diameter and 17.78 cm high (inner ring) and 10.79 cm diameter and 15.24 cm high (outer ring), respectively. The infiltration device was equipped with a count-down timer clock

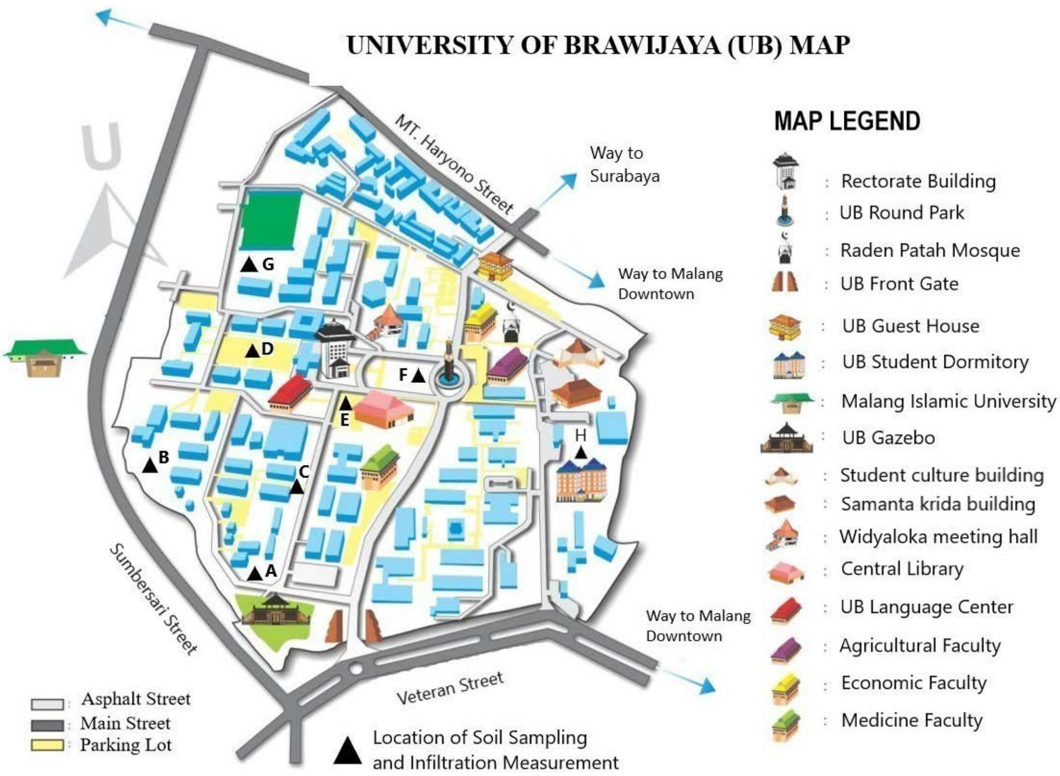

Fig. 1. Location of measurement point at University of Brawijaya (UB); source: own elaboration

Photo 1. Turf-tec Infiltrometer Model IN2-W (phot. *D. Harisuseno*)

with a beeper alarm and indicator of water level scale in inches and millimeters.

LABORATORY ANALYSIS

The soil laboratory analysis was conducted in the Laboratory of Soil and Groundwater, Water Resources Engineering Department, Faculty of Engineering, University of Brawijaya. Turf-tec double ring infiltrometer was used for measuring infiltration rate at eight points within the study area as shown in Figure 1. The inner and outer rings were set concentric and hammered into the soil uniformly utilizing a rammer up to 12.7 cm deep for the inner ring and 10.1 cm deep for the outer ring. A hand screw auger was utilized for collecting soil sample at the depth of 30 cm at a location surrounding the point of infiltration measurement. The soil samples obtained were then used for analyzing soil physical properties including soil texture, specific gravity, water content, porosity, and degree of saturation.

The sieve analysis was performed based on the standard of ASTM Test Designation D-421 and used to identify the distribution of the size of grains in terms of content of sand, clay, silt, and gravel in the soil sample. The procedure of hydrometer analysis was carried out referring to the standard of ASTM Designation D-422 which is a method employed for analyzing a fraction of soil particle size that is finer than No. 200 sieve size (0.075 mm). The result of sieve and hydrometer analyses were considered to estimate soil type and texture according to the classification system of United States Department of Agriculture (USDA).

The water content (w) analysis was conducted referring to the standard of ASTM Test Designation D-2216. The diameter of moisture cans used in the laboratory analysis has a size of 50.8 mm diameter and 22.2 mm high. The soil samples were drying for 24 h with temperature was kept between 105°C to 110°C [FRATTA *et al.* 2007].

The specific gravity (Gs) is a fundamental parameter for computing the soil weight-volume relationship particularly for determining the soil porosity and void ratio in soils and for evaluating the results of the hydrometer test. In this study, the specific gravity analysis was performed according to the standard of ASTM Test Designation

D-854 [FRATTA *et al.* 2007]. The soil porosity (n) is the proportion of pore space of soil contained in a volume of soil that can be occupied by water and air and computed using the equation explained in NIMMO [2004]. The degree of saturation (S) is the ratio between the air volume (V_w) to the volume of voids of the soil (V_v) and is determined using the equation mentioned in YE *et al.* [2018].

STATISTICAL ANALYSIS

To assure the quality of soil properties data as the predictor variable and the infiltration rate data as the response variable in the multiple linear regression, the normality and homogeneity tests were performed on each series of soil properties and infiltration rate data [AHMED *et al.* 2018]. The Shapiro–Wilk test was used for normality test, while the homogeneity test was conducted using the Levene's test. In addition, the scatter plot was developed to identify the normality of soil properties and infiltration rate data. The statistical program packages IBM SPSS Statistics ver. 25 was applied as a tool for the statistical tests. The decision of the statistical test result was determined by evaluating p-value and the significance level where p-value > significance level indicates acceptance of null hypothesis, otherwise was rejected.

The techniques of regression are used to derive information concerning the relationship between the response and predictor variables. Regression analysis requires a sample of the predictor variables and the response variable data which is then illustrated the relationship of both variables through scattering diagrams resulting from plotting the response variable and the predictor variables and subsequently summarized the relationship in a useful form of equation [FERRARO *et al.* 2011]. If there is more than one predictor variable that affects the response variable, then multiple regression analysis has been required to applied [PATLE *et al.* 2019]. The multiple regression analysis model investigates the simultaneous effects of several different predictor variables, or factors, on the response variable and describes the relationship between a single response variable and two or more predictor variables [ABDUL-WAHAB *et al.* 2005].

The present study used infiltration rate (f) as the response variable and soil texture, water content (w), porosity (n), and degree of saturation (S) as predictor variables. Microsoft Excel data analysis tool was employed to develop the multiple linear regression for summarizing the relationship between the response variable and the predictor variable.

BIVARIATE CORRELATION ANALYSIS

In this study, the multiple regression analysis was performed by taking into account only variables that show significant influence on the infiltration rate. Thus, simple bivariate correlation analyses were conducted to investigate the degree of relationship between measured infiltration rate as the response variable and each soil properties as the predictor variable. In addition, the correlation analysis was also carried out to identify whether there is a multicollinearity problem between any pairs of predictor variables or not. The problem with collinearity may appear when a predictor variable X_1 is adequately highly correlated with another predictor variable X_2, thus the independent effects of X_1 and X_2 on the response variable are difficult to disentangle [YORK 2012). The high multicollinearity problem is indicated by the pairwise correlation coefficient between the predictor variables that greater than 0.90 [PAWLICZ, NAPIERALA 2017]. If high multicollinearity problem exists, it is recommended to eliminate predictor variables having collinearity problem from the regression model to assure reliability and stability of the regression model [LIN 2008].

ROOT MEAN SQUARE ERROR (*RMSE*)

The root mean square error (*RMSE*) is affirmed as a method that has been frequently used to assess model performance through calculating a residual value between predicted value and measured value [MONTESINOS-LÓPEZ *et al.* 2018].

COEFFICIENT OF DETERMINATION (R^2)

The coefficient of determination is defined as quadratic value of the correlation coefficients according to Bravais–Pearson [RENAUD, VICTORIA-FESER 2010]. Coefficient of determination (R^2) is ability of predictor variable to explain the response variable. It also means variation in the response variable can be explained by the predictor variable X. Value of R^2 ranges from 0 to 1 where if $R^2 = 1$ variation of predictor variables can explain variation of response variable by 100%.

NASH–SUTCLIFFE EFFICIENCY (*NSE*)

Nash–Sutcliffe's value defined as a minus number of absolute squares between predicted data and normalized observational data by variants of measured data during a certain period under investigation [KNOBEN *et al.* 2019]. *NSE* value ranges between 1 (perfect) and $-\infty$. Efficiency lower than zero indicates that the average value of a set of data with measured time will be a better predictor of model. The criteria of efficiency of the Nash–Sutcliffe model was determined according to the value of *NSE* where $NSE \geq 0.75$ (good), $0.75 > NSE > 0.36$ (satisfying), and $NSE < 0.36$ (bad) [MCCUEN *et al.* 2006].

RESULTS AND DISCUSSION

SOIL PHYSICAL PROPERTIES

The results of the laboratory test of soil properties and textures at the sixteen points of measurement were shown in Table 1, whereas the descriptive statistics of soil properties were given in Table 2.

As shown in Table 1, the results of laboratory test found that the porosity (n) are ranging from 74–77%, 21–42% for water content (w), 7–28% for clay content, 20–50% for sand content, 36–53% for silt content, 17–41% for degree

Table 1. Soil physical properties at the point of measurement from laboratory analysis

Point	Porosity *n* (%)	Water content *w* (%)	Content (%)			Degree of saturation *S*	Texture
			clay (*CL*)	sand (*SN*)	silt (*SL*)		
A1	76.29	21.67	7.68	50.22	40.69	21.78	sandy loam
A2	76.05	23.77	11.45	44.32	40.44	20.33	sandy loam
B1	77.21	25.67	8.66	46.87	38.27	22.67	sandy loam
B2	77.11	26.28	10.34	43.75	40.81	28.79	sandy loam
C1	76.50	26.39	16.61	44.52	38.16	21.19	sandy loam
C2	73.58	20.66	7.11	49.77	35.78	16.78	sandy loam
D1	74.40	27.08	11.14	35.24	42.38	22.29	sandy loam
D2	75.39	27.73	16.67	38.54	44.65	24.35	sandy loam
E1	74.41	25.00	13.10	30.67	51.41	24.04	silt loam
E2	74.78	32.78	15.45	32.44	49.55	28.45	silt loam
F1	74.63	35.76	17.92	30.78	50.92	36.86	silt loam
F2	74.42	26.55	15.65	38.74	43.23	22.78	silt loam
G1	74.73	27.72	17.61	30.96	50.27	22.77	silt loam
G2	74.77	27.56	16.54	38.65	44.68	23.88	silt loam
H1	74.11	40.23	23.67	24.79	50.66	40.78	silt loam
H2	73.89	42.16	27.68	19.65	52.61	34.77	silt loam

Explanations: A = Faculty of Medicine, B = Faculty of Agriculture, C = Faculty of Humanities, D = Faculty of Mathematics and Natural Science, Department of Physics); E = Faculty of Mathematics and Natural Science, Department of Mathematics, F = Rectorate Building, G = Faculty of Mathematics and Natural Science, Department of Biology, H = UB Hall Stadium.
Source: own study.

Table 2. Statistics attribute of soil properties

Soil property	Min.	Max	Mean	Standard deviation (*SD*)	Skewness	Coefficient of variation *CV*
Porosity *n* (%)	73.58	77.21	75.14	1.14	0.66	0.02
Water content *w* (%)	20.66	42.16	28.56	6.13	1.16	0.21
Clay content *CL* (%)	7.11	27.68	14.83	5.59	0.72	0.38
Sand content *SN* (%)	19.65	50.22	37.49	8.89	0.34	0.24
Silt content *SL* (%)	35.78	52.61	44.66	5.52	0.06	0.12
Degree of saturation *S* (–)	16.78	40.78	25.91	6.49	1.14	0.25

Source: own study.

of saturation (*S*) 17–41%, respectively. The soil texture properties consist of sandy loam and silt loam. Regarding the statistical attribute of soil properties, Table 2 found that only porosity (*n*) has a relatively small range in the dispersion of data value, while other soil properties showed a relatively wide range of data value dispersion. The coefficient of variation (*CV*) was in the range of 0.02–0.38 which indicated that the soil properties data situated around their mean value that means that the soil properties data were feasible to be used in the regression analysis.

Table 3 summarizes the normality and homogeneity test for the soil properties. The Levene's test and the Shapiro–Wilk test were employed to examine the homogeneity and normality of soil properties data. Table 3 confirms that soil properties data fulfilled the assumption of homogeneity and normality data as indicated by *p*-values > 0.05 for the Levene's and Shapiro–Wilk test. Figure 2 presents the normality plot of soil properties at the sixteen measurement points. As shown in Figure 2, the scatter plot of the six soil properties showed a value of R^2 ranging 0.863–0.966 which indicated that each of the soil properties data fulfilled assumption of normality.

Table 3. Summary of the normality and homogeneity test for the soil properties

Soil property	Levene's test *p*-value	Shapiro–Wilk test *p*-value
Porosity *n*	0.075	0.095
Water content *w*	0.065	0.062
Clay content *CL*	0.744	0.306
Sand content *SN*	0.737	0.648
Silt content *SL*	0.833	0.151
Degree of saturation	0.095	0.072

Source: own study.

INFILTRATION RATE

The values of infiltration rate at the each measurement point were highly variable and presented as infiltration rate curve as shown in Figure 3. The infiltration rate curve has a high infiltration rate at the beginning of period of infiltration measurement then, decreases to reach a final infiltration rate at the remaining period of infiltration measurement. The infiltration rate at the location of point B (Faculty of Agriculture) clearly exhibited highest infiltration rate curve while location of point H (UB Hall Stadium) demonstrated lowest infiltration rate. Refer to the Table 1, the relatively high percentage of porosity and sand grain at the location of point B compared with other location of measurement point most likely lead to high of infiltration rate. In addition, low of degree of saturation, clay content, and water content contribute to high infiltration rate at point B as well. Conversely, the low infiltration rate deals with low percentage of porosity and sand content, high value of de-

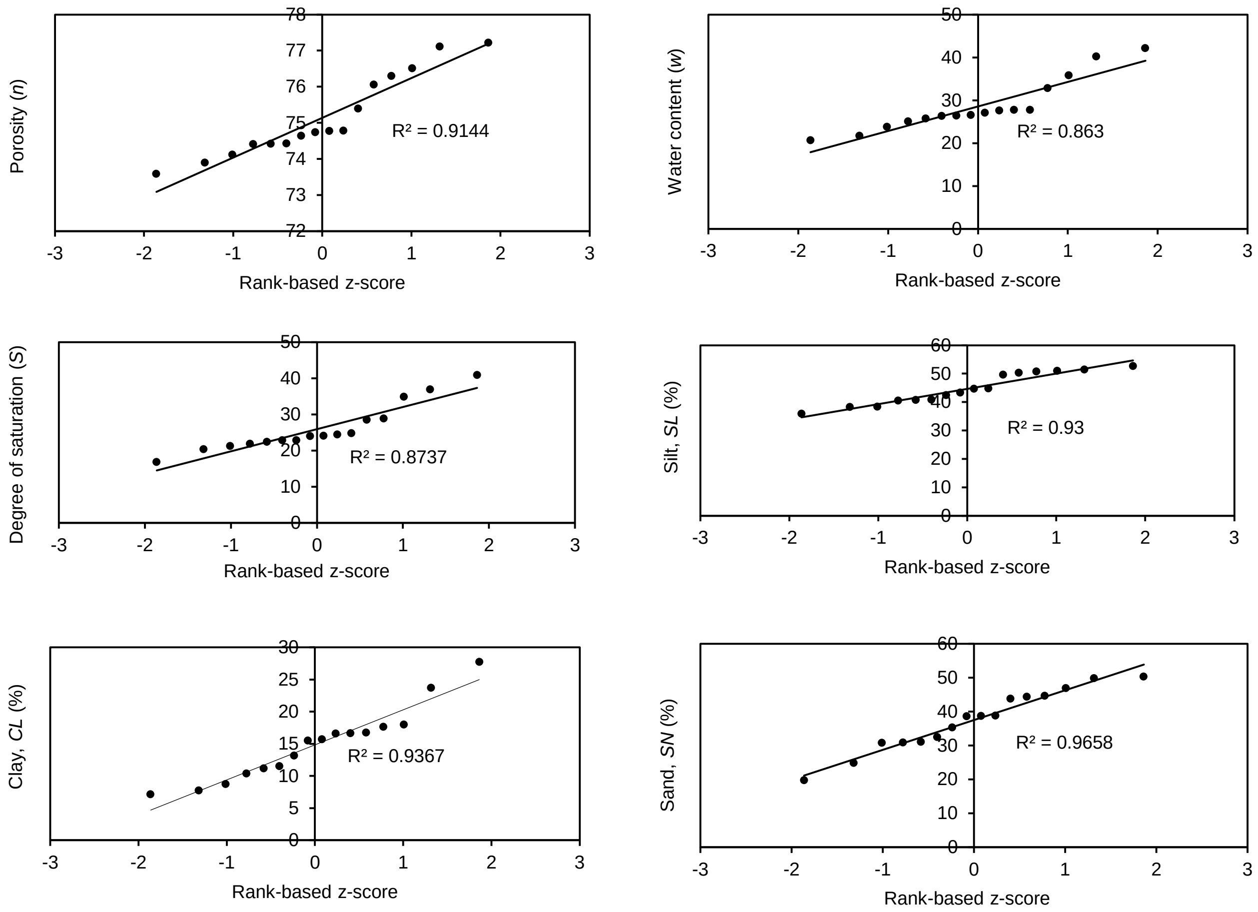

Fig. 2. Normality plot of soil properties; source: own study

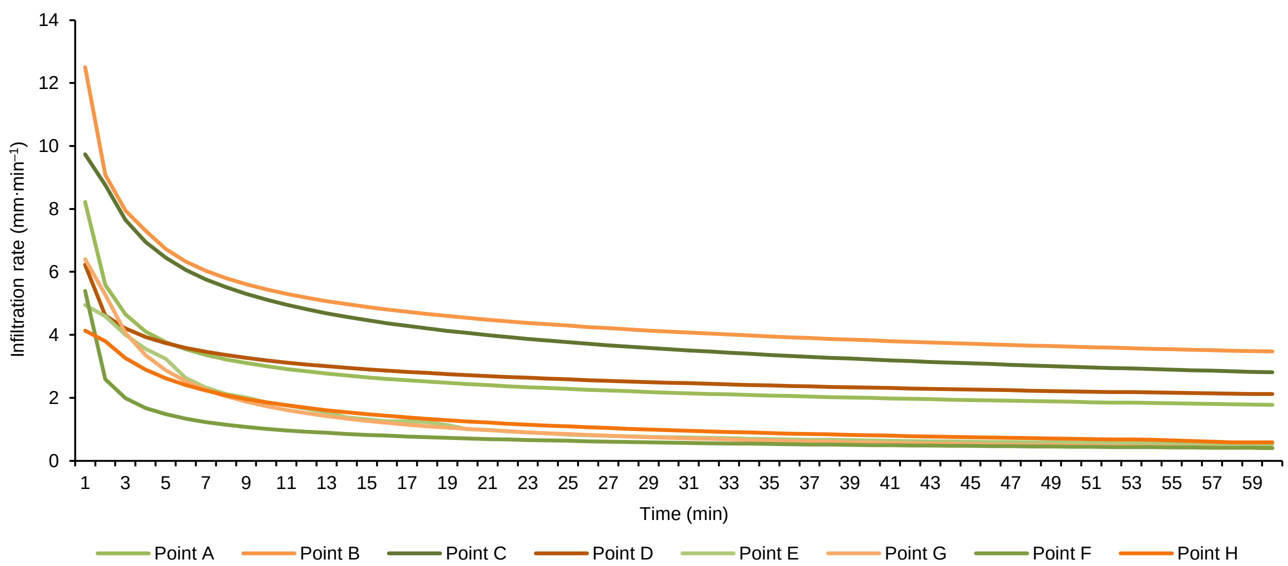

Fig. 3. The result of infiltration rate at the measurement points; A–H = as in Fig. 1 and Tab. 1; source: own study

gree of saturation, clay, and water content. From the infiltration rate curve obtained, it could be identified the value of the initial infiltration rate (f_0) and a final infiltration rate (f_c) at each measurement point. The initial infiltration rate (f_0) is the infiltration rate at the beginning of measurement which depends on the initial characteristic of soil, while the final infiltration rate (f_c) defined as the infiltration rate which owing to a constant value at a certain time indicated that soil has reached field capacity condition. Table 4 shows the value of the initial infiltration rate (f_0) and the final infiltration rate (f_c) at the entire measurement point. The measurement results showed that infiltration rate has an average initial infiltration rate (f_0) of 6.92 mm·min^{-1} and final infiltration rate (f_c) of 1.49 mm·min^{-1}.

Table 4. Initial infiltration rate (f_0) and final infiltration rate (f_c)

Point	Initial infiltration rate (f_0)	Final infiltration rate (f_c)
	(mm·min^{-1})	
A1	8.78	1.96
A2	7.67	1.59
B1	13.00	3.12
B2	12.15	3.82
C1	8.63	3.03
C2	9.35	2.59
D1	5.88	2.16
D2	6.12	2.08
E1	4.80	0.55
E2	4.88	0.50
F1	4.28	0.42
F2	6.51	0.40
G1	5.75	0.48
G2	6.11	0.34
H1	3.50	0.47
H2	3.35	0.70
Average	6.92	1.49

Explanations: points as in Fig. 1.
Source: own study.

Table 5 summarizes the bivariate correlation between the measured infiltration rate and each soil property. The results exhibit that only sand content and soil porosity show a positive correlation against measured infiltration rate, while the remaining soil properties indicate a negative correlation. All of the soil properties display a relatively high correlation coefficient (above 0.6), except 0.49 for the degree of saturation. The highest correlation coefficient of 0.842 was shown by soil porosity, whereas the lowest one was owned by the degree of saturation. From Table 5, it could be identified that there is a problem of multicollinearity between the predictor variables. It seems that the water content and degree of saturation experiencing a severe problem of collinearity each other (r = 0.909).

Considering the lowest correlation coefficient 0.49 of degree of saturation, therefore it is excluded from the list of predictor variables that influence the infiltration rate in the regression model. Consequently, it will not be considered in the composing of multiple linear regression equation.

MULTIPLE REGRESSION ANALYSIS (MLR)

In this study, the multiple regression analysis was carried out in three types of multiple linear regression considering the correlation coefficient of each soil property as shown in Table 5. The MLR-Type 1 consisted of the group of soil property that owing positive correlation coefficient (sand content *SN* and soil porosity *n*) as predictor variables, while the MLR-Type 2 considered the clay (*CL*), silt (*SL*), and water content (*w*) (group of soil property with negative correlation). The MLR-Type 3 included soil porosity (*n*), sand (*SN*), silt (*SL*), clay (*CL*), and water content as predictor variables. The MLR-Type 3 described a broad overview of the contribution of entire variables in the infiltration process simultaneously, whereas MLR-Type 1 and MLR-Type 2 are of special interest. The value of entire soil properties as the predictor variable in the MLR-Type 1, MLR-Type 2, and MLR-Type 3 was presented in Table 1.

ANALYSIS OF MLR–TYPE 1 (TWO PREDICTOR VARIABLES: SOIL POROSITY AND SAND CONTENT)

In the MLR-Type 1, sand and soil porosity were used as predictor variables in the composing of multiple regression analysis, while infiltration rate (f) as response variable. As displayed in Table 1, the mean value of soil porosity (n) is 75.14% with values ranging from 73.58% to 77.21%, whereas sand content (SN) has values ranging from 19.65% to 50.22% with a mean value of 37.49%. The MLR-Type 1 yields weights or regression coefficients β_0, β_1, β_2 were –77.712, 0.188, and 1.032 respectively. The equation of Type 1 ($f_{\mathrm{MLR\_1}}$) was shown as follows:

$$f_{\mathrm{MLR\_1}} = -77.712 + 0.188\ SN + 1.032\ n \qquad (1)$$

Where: SN = sand content (%), n = porosity (%).

As shown in Equation (1), the coefficient of sand and soil porosity shows a positive sign that implies an increasing trend in the infiltration rate. The analysis revealed that the soil porosity gives more contribution to the MLR-Type 1 (84.5%) compared with sand content (15.5%). The degree of contribution was indicated by the value of weights or regression coefficients (β) in Equation (1) where for soil porosity showed a value of 1.032 that was higher than sand content (0.188). This makes sense since the infiltration process was great influenced by the soil porosity that closely associates with the availability of pore space in the soil structure [HARISUSENO *et al.* 2019]. This result was consistent with LIPIEC *et al.* [2006] and SINHA and SINGH [2016], who stated that the soil porosity has an important role in process of water entering the soil. The result of MLR-Type 1 found that that the soil properties with positive correlation coefficient have a good performance in estimating infiltration rate.

Table 5. Summary of bivariate correlation between infiltration rate and soil properties

Variable	Infiltration rate (mm·min^{-1})	Porosity n (%)	Water content w (%)	Clay content CL (%)	Sand content SN (%)	Silt content SL (%)	Degree of saturation S
Infiltration rate	1						
Porosity n (%)	0.842	1					
Water content w (%)	–0.660	–0.399	1				
Clay content CL (%)	–0.743	–0.453	0.592	1			
Sand SN (%)	0.839	0.603	–0.464	–0.661	1		
Silt SL (%)	–0.831	–0.526	0.750	0.746	–0.727	1	
Degree of saturation S	–0.490	–0.214	0.909	0.707	–0.727	0.691	1

Source: own study.

ANALYSIS OF MLR-TYPE 2 (THREE PREDICTOR VARIABLES: WATER CONTENT, CLAY, AND SILT CONTENT)

The MLR-Type 2 was composed of three predictor variables namely water content (*w*), clay (*CL*), and silt content (*SL*). These variables had a negative correlation coefficient against measured infiltration rate (*f*), thus it gives the interest to know the performance of the model compared with the MLR-Type 1. The water content *(w)* varied from 20.66–42.16% with an average value of 28.56%, the clay content (*CL*) had a value within 7.11–27.68% with 14.83% on averagely, and the silt content (*SL*) had an average value of 44.66% with values vary between 35.78–52.61%.

Fitting the measured infiltration rate and the three predictor variables yielded the regression coefficients β_0, β_1, β_2 *and* β_3 were 22.43, 0.135, –0.253, –0.35, respectively. The resulted equation was displayed in the form:

$$f_{\mathrm{MLR\_2}} = 22.43 + 0.135\, w - 0.253\, CL - 0.35\, SL \qquad (2)$$

The negative sign of the coefficient of clay and silt variables indicates a declining trend in the infiltration rate. Further analysis was conducted to examine the significance of coefficients of regression through *p*-value where the analysis showed that the *p*-value is lower than 0.05 that the predictor variables have a significant influence in the regression at a 95% confidence level.

The results of MLR-Type 2 exhibit that the variation of infiltration rate was explained by 74.1% of the behaviour of predictor variables. The analysis revealed that the silt content *SL* gives high contribution to the regression model (47.4%) compared with clay content *CL* (34.3%) and soil water content *w* (18.3%) respectively. The degree of contribution of each predictor variable was shown by the value of the unstandardized coefficient of the regression model of silt content *SL* (0.35), clay content *CL* (0.253), and soil water content *w* (0.135).

ANALYSIS OF MLR-TYPE 3 (FIVE PREDICTOR VARIABLES: SOIL POROSITY, WATER CONTENT, SAND, CLAY, AND SILT CONTENT)

Five predictor variables were used to compose the MLR-Type 3 namely soil porosity (*n*), water content (*w*), sand (*SN*), clay (*CL*), and silt content (*SL*). These predictor variables were taken into consideration since they have a high correlation with the infiltration rate (*f*). The statistical regression analysis was carried out by fitting the measured infiltration rate as the response variable and the five predictor variables as the predictor variable. The result of regression coefficients β_0, β_1, β_2, β_3, β_4, and β_5 were –62.014, 1.142 (soil porosity), –0.205 (clay), –0.063 (sand), –0.301 (silt), and 0.07 (soil water content), respectively. The equation was presented as follows:

$$f_{\mathrm{MLR\_3}} = -62.014 + 1.142\, n - 0.205\, CL - 0.063\, SN + {} - 0.301\, SL + 0.07\, w \qquad (3)$$

As shown in Equation (3), the negative sign of the coefficient of clay, sand, and silt variables denote a declining trend in infiltration rate. Further, the result of *p*-value displayed a magnitude lower than 0.05 which indicates that the predictor variables influence significantly the regression at a 95% confidence level. The result of the determination coefficient (R^2) demonstrated that 87.4% of the infiltration rate event was greatly influenced by the characteristic of five predictor variables. From the multiple regression analysis, it could be known that the soil porosity (*n*) gives a high contribution to the regression model (64.1%) compared with silt *CL* (16.9%), clay *CL* (11.5%), water content *w* (3.9%), and sand *SN* (3.5%) respectively. The degree of contribution of each predictor variable was shown by the value of the unstandardized coefficient of the regression model of soil porosity *n* (1.14), silt content *SL* (0.30), clay content *CL* (0.21), soil water content *w* (0.07), and sand content *SN* (0.06). Using Equation (3), the estimated infiltration rate for all measurement points could be derived and then evaluated by comparing with the measured one.

Using Equations (1), (2), and (3), the estimated infiltration rate for all measurement points could be derived by inputting the value of sand content (*SN*) and soil porosity (*n*) into the three MLR equations. Subsequently, the results of estimated infiltration rate from each of the MLR equation were compared and evaluated with the measured infiltration rate which obtained from field measurement. Figure 4 exhibits plotting of the measured infiltration rate along with the estimated infiltration rate from the MLR-Type 1 ($f_{\mathrm{MLR\_1}}$), MLR-Type 2 ($f_{\mathrm{MLR\_2}}$), and MLR-Type 3 ($f_{\mathrm{MLR\_3}}$).

According to Figure 4, it was found that the estimated infiltration rate derived from the MLR-Type 1 has values ranging from 2.24 to 10.78 mm·min^{-1} with a mean value of 6.88 mm·min^{-1}. Further, Figure 4 demonstrated that both measured and estimated infiltration rates from the MLR-Type 1 show a similar pattern while the value of both infiltration rates shows relatively similar as well, except for measurement points A1, A2, B1, B2, C1, and C2. However, the results of MLR-Type 1 confirms that the sand and soil porosity are proved to be a good predictor for characterizing infiltration rate.

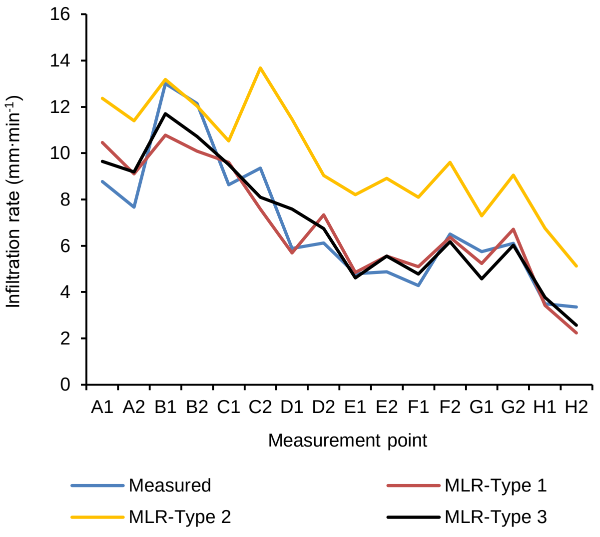

Fig. 4. Plotting measured vs. estimated infiltration rate (MLR-Type 1, 2, and 3); points as in Fig. 1 and Tab. 1, MLR = multiple regression analysis; source: own study

The estimated infiltration rate from MLR-Type 2 varies from 5.12 to 13.68 mm·min^{-1} with an average value of 9.80 mm·min^{-1} as indicated in Figure 4. The values of infiltration rate significantly different between the measured and estimated for all measurement points. Consequently, the MLR-Type 2 show less reliable compared with the MLR-Type 1 for describing soil infiltration rate characteristic. Despite those soil properties own a high negative correlation coefficient as displayed in Table 5, it seems that they fairly show low performance. This result revealed that the soil properties with negative of correlation coefficient show slightly no good in estimating infiltration rate when they are considered as the stand-alone predictor variable. Therefore, it is necessary to consider all soil properties that have positive and negative correlation coefficients as the predictor variables and investigate how their influence on estimating infiltration rate.

The results of the estimated infiltration rate from MLR-Type 3 (f_{MLR_3}) gives values ranging from 2.57 to 12.71 mm·min^{-1} with a mean value of 6.99 mm·min^{-1}. As shown in Figure 4, both measured and estimated infiltration rate displays consistent results which indicate high similarity and strong relationship between them. Thus, the result of MLR-Type 3 is confirmed as the most reliable and accurate infiltration equation comparing with the MLR-Type 1 and 2 for representing infiltration rate behaviour in the study area.

MODEL PERFORMANCE EVALUATION

In the present study, due to the minimum of measurement data (sixteen data) used to form the multilinear regression equation, consequently, the whole sixteen data were used in the model performance evaluation. The performance analysis was performed by comparing the total number of measured infiltration data and estimated data obtained from the three MLR equations. The five statistical parameters namely coefficient of correlation (R), coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), root mean square error ($RMSE$), and relative error (RE) were employed to assess model performance on estimating infiltration rate.

MULTIPLE LINEAR REGRESSION TYPE 1 (MLR-TYPE 1)

Figure 5 shows the scatter plot between the measured and estimated infiltration rate, while the consistency plot between cumulative measured and estimated infiltration rate is given in Figure 6. According to Figure 5, it could be known that the scatter points are on surrounding the straight-line trend which means that the estimated infiltration rate has a good agreement with the measured one. The value of R^2 = 81.3% indicates that the variation of the measured infiltration rate is highly influenced by the estimated one. Furthermore, as shown in Figure 6 both measured and estimated infiltration rates show a consistent pattern, where most of the cumulative points are on surrounding in the straight line with a slope of 1:1 (45°). In addition, the NSE of MLR-Type 1 shows a value of 0.997

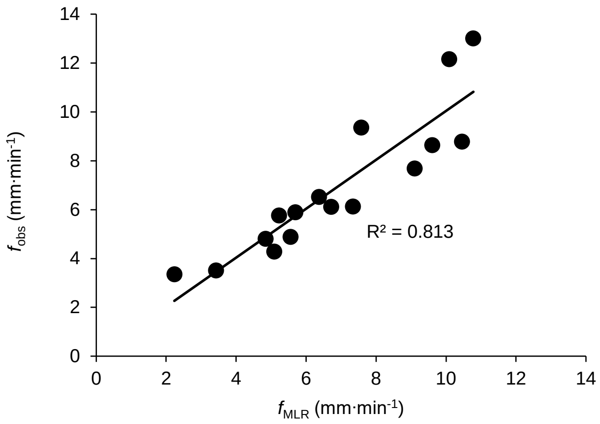

Fig. 5. Scatter plot of measured vs. estimated infiltration rate in the case of multiple linear regression type 1; source: own study

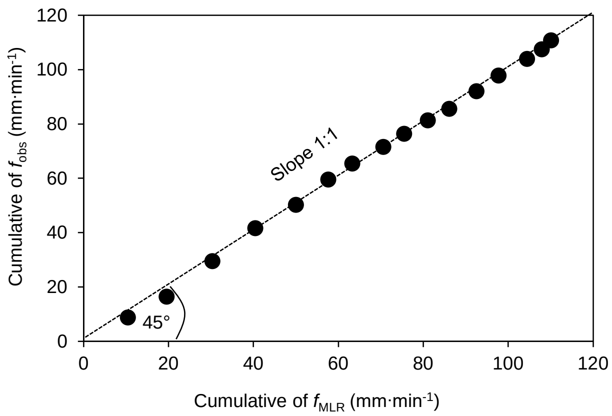

Fig. 6. Consistency plot of cumulative measured and estimated infiltration in the case of multiple linear regression type 1; source: own study

which is categorized as good in model performance. The $RMSE$ has a value ranging from 0.045–2.22 with a mean value of 0.97, RE shows a mean value of 13.45%.

Hence, the performance analysis of MLR-Type 1 demonstrated that the equation is a good predictor of describing infiltration rate characteristics.

MULTIPLE LINEAR REGRESSION TYPE 2 (MLR-TYPE 2)

Scatter and consistency plot between the measured and estimated infiltration rate are presented in Figures 7 and 8. The degree of relationship between the measured and estimated infiltration rate was shown with the value of determination coefficient (R^2) 72.2% as displayed in Figure 7. It means that approximately 72.2% of the variation of measured infiltration rate was explained by the change of the predictor variables. Based on the value of R^2, it could be known that there is a quite good relationship between the measured and estimated infiltration rate, however, the relationship is less good compared with the MLR-Type 1.

In addition, Figure 8 demonstrates that both measured and estimated infiltration rates display inconsistent patterns which are indicated by the change in the direction of the slope of lines where the direction of the line tends to move

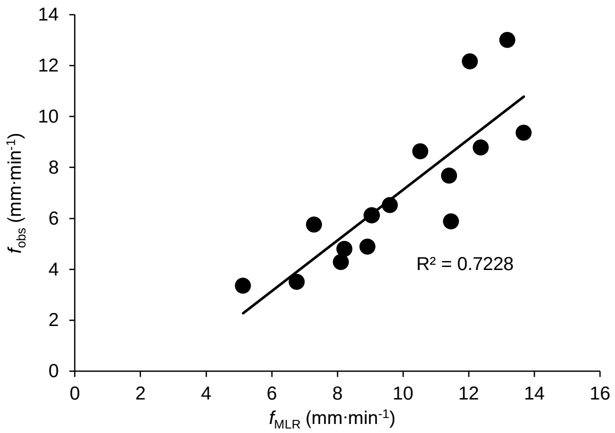

Fig. 7. Scatter plot of measured vs. estimated infiltration rate in the case of multiple linear regression type 2; source: own study

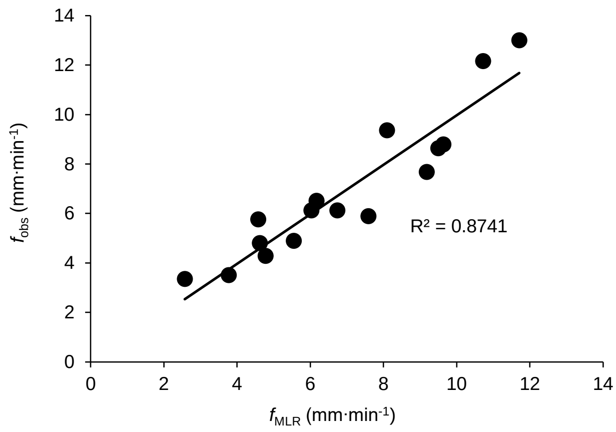

Fig. 9. Scatter plot of measured *vs* estimated infiltration rate in the case of multiple linear regression type 3; source: own study

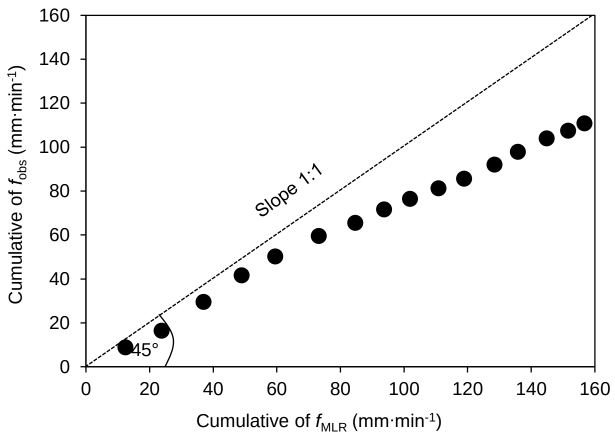

Fig. 8. Consistency plot of cumulative measured and estimated infiltration in the case of multiple linear regression type 2; source: own study

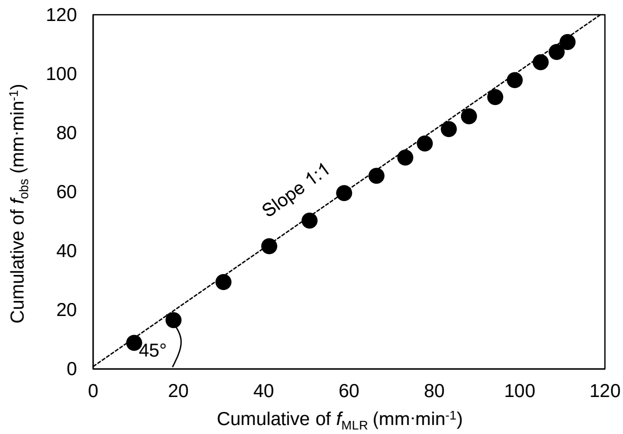

Fig. 10. Consistency plot of cumulative measured and estimated infiltration in the case of multiple linear regression type 3; source: own study

away from the straight line with a slope of 1:1 (45°). Accordingly, it could be known that the estimated infiltration rate resulted from the MLR-Type 2 has less good and reliability compared with the MLR-Type 1. Analysis of *NSE* value showed a magnitude of 0.98 while the range of *RMSE* was between 0.12 and 5.58 with a mean value of 2.89 and the *RE* exhibited a mean value of 50.87%. Hence, the MLR-Type 2 is a slightly good predictor of infiltration rate though it is still less good if compared with the MLR-Type 1.

MULTIPLE LINEAR REGRESSION TYPE 3 (MLR–TYPE 3)

The scatter plot between the measured and estimated infiltration rate for MLR-Type 3 is given in Figure 9 while Figure 10 presents the consistency plot between the measured and estimated one. As shown in Figure 9, the determination coefficient (R^2) indicates that approximately 87.4% of the variation of measured infiltration rate was explained by the change of the predictor variables. Thus, it could be confirmed that there is a good relationship between the measured and estimated infiltration rate. In addition, Figure 10 denotes a consistent pattern between the measured and estimated infiltration rate where most of the cumulative points are on surrounding in the straight line with a slope of 1:1 (45°). Accordingly, it could be known that the estimated infiltration rate resulted from the MLR-Type 3 shows a better quality and more reliable compared with the MLR-Type 1 and Type 2.

The performance of MLR-Type 3 was categorized in a good performance that indicated by the *NSE* value of 0.98 while the range of *RMSE* was between 0.08 and 1.71 with a mean value of 0.85 and *RE* displayed a mean value of 12.56%. Table 6 summarizes model performance analysis for all multiple linear regression types. From all the model performance analyses, it was revealed that an increase in the number of predictor variables leads to increasing the magnitude of R^2 and R that indicates an increase in the model reliability on estimating soil infiltration rate. According to Table 6, the MLR-Type 2 showed the lowest value of R^2, R, *NSE*, and the highest value of *RMSE* and *RE* whereas the MLR-Type 3 had the highest value of R^2, R, *NSE*, and lowest value of *RMSE* and *RE*. This confirms that the MLR-Type 3 is the best amongst all of the types of MLR for estimating soil infiltration rate in the study area.

Table 6. Summary of model performance analysis in the case of multiple linear regression (MLR)

Model type	Model performance parameter				
	R	R^2	*NSE*	*RMSE*	*RE* (%)
MLR-Type 1	0.90	0.81	0.997	1.19	13.46
MLR-Type 2	0.85	0.72	0.980	3.22	50.84
MLR-Type 3	0.93	0.87	0.998	0.97	12.56

Explanations: *R* = correlation coefficient, R^2 = determination coefficient, *NSE* = Nash–Sutclife efficiency, *RMSE* = root mean square error, *RE* = relative error.
Source: own study.

CONCLUSIONS

The present study employed multiple linear regression to estimate infiltration rate where the soil properties used as the predictor variable and measured infiltration rate as the response variable. The usage of soil properties with a high negative correlation coefficient as a stand-alone predictor variable shows no reliability in estimating infiltration rate whereas good performance was given by soil properties with positive correlation. The analysis of model performance revealed that MLR-Type 3 showed the highest value of determination coefficient (R^2), correlation coefficient (*R*), Nash–Sutclife efficiency (*NSE*) and lowest value of root mean square error (*RMSE*) and relative error (*RE*) which indicating high similarity between estimated infiltration rate and those from field measurement. These results highly confirmed that the multiple linear regression MLR-Type 3 with five soil predictors (soil porosity, silt, clay, sand content, and water content) show the best performance in estimating infiltration behaviour in the study area. In addition, the study found that the soil porosity contributed mostly to the multiple linear regression model followed by silt, clay, water content, and sand, respectively. Thus, it indicates clearly that soil porosity is the most essential variable which affects soil infiltration rate. The multiple linear regression model could be used as a reliable method to estimate infiltration rate since its practicability, simplicity, and easiness of being implemented. The study acknowledges that there is a weakness particularly in the model testing and validation step due to the limitation of sample data. Further research needs to involve many samples to obtain better results that represent behaviour of true infiltration rate spatially and temporally.

ACKNOWLEDGEMENTS

The authors are sincerely gratitude and give appreciation toward the Laboratory of Soil and Groundwater, Department of Water Resources Engineering, Faculty of Engineering, University of Brawijaya for kindly assisting field measurement and soil sampling analysis for our research.

REFERENCES

ABDUL-WAHAB S.A., BAKHEIT C.S., AL-ALAWI S.M. 2005. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environmental Modelling and Software. Vol. 20. No. 10 p. 1263–1271. DOI 10.1016/j.envsoft.2004.09.001.

AHMED K., SHAHID S., ISMAIL T., NAWAZ N., WANG X.J. 2018. Absolute homogeneity assessment of precipitation time series in an arid region of Pakistan. Atmosfera. Vol. 31. No. 3 p. 301–316. DOI 10.20937/ATM.2018.31.03.06.

CZYŻYK F., ŚWIERKOT Z. 2017. Recharging infiltration of precipitation water through the light soil, in the absence of surface runoff. Journal of Water and Land Development. Vol. 32 p. 25–30. DOI 10.1515/jwld-2017-0003.

DEWIDAR A.Z., AL-GHOBARI H., ALATAWAY A. 2019. Developing a fuzzy logic model for predicting soil infiltration rate based on soil texture properties. Water SA. Vol. 45. No. 3 p. 400–410. DOI 10.17159/wsa/2019.v45.i3.6737.

FERRARO M.B., COLUBI A., GONZÁLEZ-RODRÍGUEZ G., COPPI R. 2011. A determination coefficient for a linear regression model with imprecise response. Environmetrics. Vol. 22. No. 4 p. 516–529. DOI 10.1002/env.1056.

FRATTA D., AGUETTANT J., SMITH L. 2007. Introduction to soil mechanics laboratory testing. 1st ed. Boca Raton. CRC Press. ISBN 978-1-4200-4562-8 pp. 229.

GREEN W.H., AMPT G.A. 1911. Studies on soil physics. The Journal of Agricultural Science. Vol. 4. No. 1 p. 1–24.

HAIR J.F., BLACK W.C., BABIN B.J., ANDERSON R.E., BLACK W.C., ANDERSON R.E. 2019. Multivariate data analysis. 8th ed. Hampshire. Cengage Learning EMEA. ISBN 978-1-4737-5654-0 pp. 813.

HARISUSENO D., KHAERUDDIN D.N., HARIBOWO R. 2019. Time of concentration based infiltration under differ-ent soil density, water content, and slope during a steady rainfall. Journal of Water and Land Development. Vol. 41 p. 61–68. DOI 10.2478/jwld-2019-0028.

HORTON R.E. 1940. An approach toward a physical interpretation of infiltration-capacity. In: Soil Science Society of American p. 399–417.

JIANG Y., CHEN W., WANG G., SUN G., ZHANG F. 2017. Influence of initial dry density and water content on the soil–water characteristic curve and suction stress of a reconstituted loess soil. Bulletin of Engineering Geology and the Environment. Vol. 76. No. 3 p. 1085–1095. DOI 10.1007/s10064-016-0899-x.

KNOBEN W.J.M., FREER J.E., WOODS R.A. 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Science Vol. 23. No. 10 p. 4323–4331. DOI 10.5194/HESS-23-4323-2019.

KOSTIAKOV A. 1932. On the dynamics of the coefficient of water percolation in soils and on the necessity of studying it from a dynamic point of view for the purposes of amelioration. Transactions of the Sixth Congress of International Society of Soil Science Moscow Part A p. 15–21.

LIN F.J. 2008. Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. Quality and Quantity. Vol. 42. No. 3 p. 417–426. DOI 10.1007/s11135-006-9055-1.

LIPIEC J., KUŚ J., SŁOWIŃSKA-JURKIEWICZ A., NOSALEWICZ A. 2006. Soil porosity and water infiltration as influenced by tillage methods. Soil and Tillage Research. Vol. 89. No. 2 p. 210–220. DOI 10.1016/j.still.2005.07.012.

MAO L., LI Y., HAO W., ZHOU X., XU C., LEI T. 2016. A new method to estimate soil water infiltration based on a modified Green-Ampt model. Soil and Tillage Research. Vol. 161 p. 31–37. DOI 10.1016/j.still.2016.03.003.

MCCUEN R.H., KNIGHT Z., CUTTER A.G. 2006. Evaluation of the Nash–Sutcliffe efficiency index. Journal of Hydrologic Engineering. Vol. 11. No. 6 p. 597–602. DOI 10.1061/(ASCE)1084-0699(2006)11:6(597).

MONTESINOS-LÓPEZ A., MONTESINOS-LÓPEZ O.A., DE LOS CAMPOS G., CROSSA J., BURGUEÑO J., LUNA-VAZQUEZ F.J. 2018. Bayesian functional regression as an alternative statistical analysis of high-throughput phenotyping data of modern agriculture. Plant Methods. Vol. 14. No. 1 p. 1–17. DOI 10.1186/s13007-018-0314-7.

MUBARAK I., ANGULO-JARAMILLO R., MAILHOL J.C., RUELLE P., KHALEDIAN M., VAUCLIN M. 2010. Spatial analysis of soil surface hydraulic properties: Is infiltration method dependent? Agricultural Water Management. Vol. 97. No. 10 p. 1517–1526. DOI 10.1016/j.agwat.2010.05.005.

NIE W., MA X., FEI L. 2017. Evaluation of infiltration models and variability of soil infiltration properties at multiple scales. Irrigation and Drainage. Vol. 66. No. 4 p. 589–599. DOI 10.1002/ird.2126.

NIMMO J. 2004. Porosity and pore size distribution. In: Encyclopedia of soils in the environment. London. Elsevier p. 295–303.

PATLE G.T., SIKAR T.T., RAWAT K.S., SINGH S.K. 2019. Estimation of infiltration rate from soil properties using regression model for cultivated land. Geology, Ecology, and Landscapes. Vol. 3. No. 1 p. 1–13. DOI 10.1080/24749508.2018.1481633.

PAWLICZ A., NAPIERALA T. 2017. The determinants of hotel room rates: An analysis of the hotel industry in Warsaw, Poland. International Journal of Contemporary Hospitality Management. Vol. 29. No. 1 p. 571–588. DOI 10.1108/IJCHM-12-2015-0694.

PHILIP J.R. 1956. The theory of infiltration: 1. The infiltration equation and its solution. Soil Science. Vol. 85. No. 6 p. 333–337.

RASHIDI M., HAJIAGHAEI A., AHMADBEYKI A. 2014. Prediction of soil infiltration rate based on particle size distribution, bulk density, organic matter and moisture content of soil. American-Eurasian Journal of Scientific Research. Vol. 4. No. 1 p. 24–29. DOI 10.5829/idosi.aerj.2014.4.1.1119.

RENAUD O., VICTORIA-FESER M.P. 2010. A robust coefficient of determination for regression. Journal of Statistical Planning and Inference. Vol. 140. No. 7 p. 1852–1862. DOI 10.1016/j.jspi.2010.01.008.

SINHA R., SINGH P. 2016. Evaluation of infiltration behaviour and soil characteristics in Dhanbad – Jharia Township Area, Jharkhand, India. Current World Environment. Vol. 11. No. 2 p. 584–591. DOI 10.12944/cwe.11.2.29.

SUN D., SUN W., XIANG L. 2010. Effect of degree of saturation on mechanical behaviour of unsaturated soils and its elasto-plastic simulation. Computers and Geotechnics. Vol. 37. No. 5 p. 678–688. DOI 10.1016/j.compgeo.2010.04.006.

VAN DE GENACHTE G., MALLANTS D., RAMOS J., DECKERS J.A., FEYEN J. 1996. Estimating infiltration parameters from basic soil properties. Hydrological Processes. Vol. 10. No. 5 p. 687–701. DOI 10.1002/(SICI)1099-1085(199605)10:5<687::AID-HYP311>3.0.CO;2-P.

YE X., WANG S. WANG Q., SLOAN S.W., SHENG D. 2018. The influence of the degree of saturation on compaction-grouted soil nails in sand. Acta Geotechnica. Vol. 3456789. DOI 10.1007/s11440-018-0706-x.

YORK R. 2012. Residualization is not the answer: Rethinking how to address multicollinearity. Social Science Research. Vol. 41. No. 6 p. 1379–1386. DOI 10.1016/j.ssresearch.2012.05.014.