GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 20 • Number 1 • 2026

https://doi.org/10.7494/geom.2026.20.1.69

Geologic Control of Soil-Infiltration Rate Based on Artificial Neural Network Models

Totok Sulistyo[1, 2]*, Sari Bahagiarti Kusumayudha[3], Tedy Agung Cahyadi[4], Reza Adhi Fajar[5], Mariatul Kiptiah[2]

[1] Universitas Pembangunan Nasional "Veteran" Yogyakarta, Geological Engineering Department, Doctoral Program in Geological Engineering, Yogyakarta, Indonesia
[2] Balikpapan State Polytechnic, Civil Engineering Department, Balikpapan, Indonesia
[3] Universitas Pembangunan Nasional "Veteran" Yogyakarta, Geological Engineering Department, Yogyakarta, Indonesia
[4] Universitas Pembangunan Nasional "Veteran" Yogyakarta, Mining Engineering Department, Yogyakarta, Indonesia
[5] Banjarmasin State Polytechnic, Civil Engineering Department, Banjarmasin, Indonesia
* Corresponding author

Abstract: The interconnected porosity of soil provides conduit channels for the downward infiltration of water into the subsurface; this occurs in soil layers and within soil-less areas or geologic formations. The lithology and geological structure significantly influence the infiltration capacity of soils and are crucial in determining whether the infiltration water continuously reaches an aquifer or becomes stagnant in the saturated soil. An artificial neural network (ANN) algorithm was employed to model the actual infiltration rate, incorporating soil texture and soil moisture along with geological scores as inputs and actual infiltration rates as outputs. This study aimed to quantify qualitative geological data and incorporate it into ANN model parameters. The development of the ANN infiltration model involved two serial trial-and-error experiments to determine the optimal number of nodes in the hidden layer, ranging from nodes c(4,2) to c(12,2), one serial experiment with geological input, and the other without geological input. Throughout the model testing, metrics such as MAE, RMSE, and MSE were recorded, and the first and second optimum models were identified when employing c(9,2) nodes of hidden layers. The resulting model can be used to predict actual infiltration and will be beneficial for hydrometeorological-disaster mitigation and city-development planning.

Keywords: infiltration, lithology, geological structures, ANNs, soil

Received: March 8, 2025; accepted: October 16, 2025

E-mails & ORCID iDs: totok.sulistyo@poltekba.ac.id, 311231002@student.upnyk.ac.id, https://orcid.org/0000-0001-7612-5959 (T.S.);
saribk@upnvy.ac.id, https://orcid.org/0000-0001-7809-1000 (S.B.K.);
tedyagungc@upnvy.ac.id, https://orcid.org/0000-0002-1773-2713 (T.A.C.);
reza@poliban.ac.id, https://orcid.org/0000-0001-5875-5586 (R.A.F.);
mariatul.kiptiah@poltekba.ac.id, https://orcid.org/0000-0002-1190-4940 (M.K.).

1. Introduction

Infiltration is governed by interconnected pores within soil; these are created primarily by the spacing between the grain packing, soil cracks, and soil piping [1]. Within the soil, water infiltrates and fills in these pores. Continuous infiltration leads to saturated soil; if the infiltrating water reaches the underlying strata, the physical properties of this stratum become crucial in maintaining a constant flow to recharge the aquifer. In contrast, infiltration cannot progress deeper if the underlying strata are aquicludes.

Infiltration is when water reaches the Earth's surface, where some runs off and some infiltrates the soil. The rainfall and runoff correlation has been previously researched and modeled as the R-R Model using the improvising ANN algorithm to obtain higher accuracy in prediction; the deep ANN model provides better results than the shallow ANN model does [2]. Previous studies have also revealed that soils with medium to low permeability, along with an underlying lithologic layer, causes water to infiltrate into the soil and flow horizontally [3].

Geology has never been used as a parameter in a machine-learning model, as it is characterized by qualitative descriptive data rather than quantitative data. The systematic literature review (which included 131 experimental and 22 theoretical articles selected from Scopus, ScienceDirect, and Google Scholar published between 1953 and 2019) was meta-analyzed; none of these articles mentioned any geology terms or phrases in their titles or abstracts as affecting parameters in soil infiltration [4]. Such a literature review was also confirmed by other research results about the effects of soil physical properties and vegetation on soil-infiltration rates [5, 6]. Moreover, there is still a gap in the research and publications of the correlation between soil infiltration and geologic parameters; this gap can be filled by developing an accurate model of the correlation among infiltration geologic parameters combined with other influencing parameters. Such an accurate infiltration model can be used to simulate and predict further work such as drainage planning, flood mitigation, and prevention.

Most residual soil profiles typically have limited depths, gradually transitioning downward to fresh lithology as part of a geological formation. Soil is absent in several areas, and geological formations are exposed on the region's surface. Sedimentary rocks (including breccia and sandstone) possess textures that resemble those of the soil, and their porosities are formed through grain packing as well as structures like joints, cracks, and faults. Subsurface geologic conditions are essential for directing water after its downward infiltration through soils, thus determining whether it will continuously move through the underlying strata [7] or flow parallel to the strata. Geologic faults significantly contribute to providing conduits in recharge areas [8]. Geologic formations also control water flows in the subsurface and surface runoffs on Earth [9].

The infiltration rate at a specific point on the surface determines whether rainwater runoffs, soil moisture, and saturated soil near the surface will flow or move horizontally to avoid a geohydrologic barrier [10]; alternatively, the rainwater will be recharged in the free aquifer as groundwater [8]. A low infiltration rate and other

prevailing parameters on the surface play significant roles in flood disasters. Moreover, the high infiltration rate and subsurface parameters in recharge areas (such as lithology, structural geology, and hydrogeology) also play vital roles in storing the water in aquifers. Aquifer-recharge processes are mainly controlled by geological conditions such as faults, and water recharge typically results from streambed infiltration [8, 11]. A recharge area can be analyzed via geology and geophysics (such as through geological maps and cross-sections, seismic stacks, and geoelectric data). The recharge area's location can also be traced using the elevation versus the $\delta^{18}O$ and $\delta^{2}H$ of the meteoric water isotope value [12]. In the soil, water infiltrates through the soil layer, moving both upward and downward due to heat and pressure. The capillary zone has been observed and studied in previous research [13, 14].

Nevertheless, there is a gap between the geology and the soil infiltration. The development of geologic control on the actual soil-infiltration model will fill the recent research gap between infiltration and the geology parameters. Moreover, obtaining data on the soil-infiltration distribution in certain areas is crucial for providing information for sustainable groundwater management [15], flood prevention, and regional development planning.

Hypothetically, each geological formation controls infiltration differently depending on the geologic structure and lithologic variation that are associated with it; in geohydrologic systems, the combination of its lithology and geological structure can include an aquifer, aquitard, or aquifuge depending on the permeability [16]. The geologic conditions control infiltration directly when alluvial deposits, residual soil, and colluvium do not cover the area. In contrast, a geologic formation occurs beneath the soil; this indirectly controls the soil-infiltration rate. The research question is how the underlying geologic formation controls the surface soil's infiltration. This approach aims to determine the underlying geologic formation and geological structure that control the soil-infiltration rate at the surface.

2. Research Significance

The novelty of this research lies in the development of a soil-infiltration model that incorporates the quantifications of geologic parameters. Machine-learning models often overlook these parameters due to their qualitative and descriptive nature. This study addresses this research gap by emphasizing the roles of lithology and geologic structures on water infiltration. The findings of this research have potential applications in evaluating the influence of geologic features on runoff calculations and predictions. This study contributes to foundational knowledge in identifying recharge areas to conserve aquifers from pollution and drought. It can also be applied to flood-disaster management in the study area (where certain catchments are highly prone to flooding). Conservation of soil infiltration during urban development emerges as a key strategy for flood mitigation [17]. Moreover, intense precipitation

combined with limited infiltration capacity in slope areas can trigger flood disasters [18] and landslides [19]. Changes in infiltration affect the soil moisture, unit weight, and Atterberg limit of slope soils, ultimately increasing the risk of slope failure in vulnerable slopes.

3. Method

3.1. Study Area

The research commenced by choosing a location based on its geographical, hydrological, and geological considerations. The research area was a growing city that consisted of 508.39 km^2 and was inhabited by 717,231 citizens. Administratively, it is part of Balikpapan City and East Kalimantan Province, Indonesia. The national location of the research area is vital, as it is the buffer area of the Indonesian New Capital City that needs solutions for the potential for hydrometeorological disasters and groundwater resources. There are a few river basins in the area, such as the Sepinggan, Ampal, and Wain River Basins; these river basins are characterized by flooding in some places during the rainy season [20].

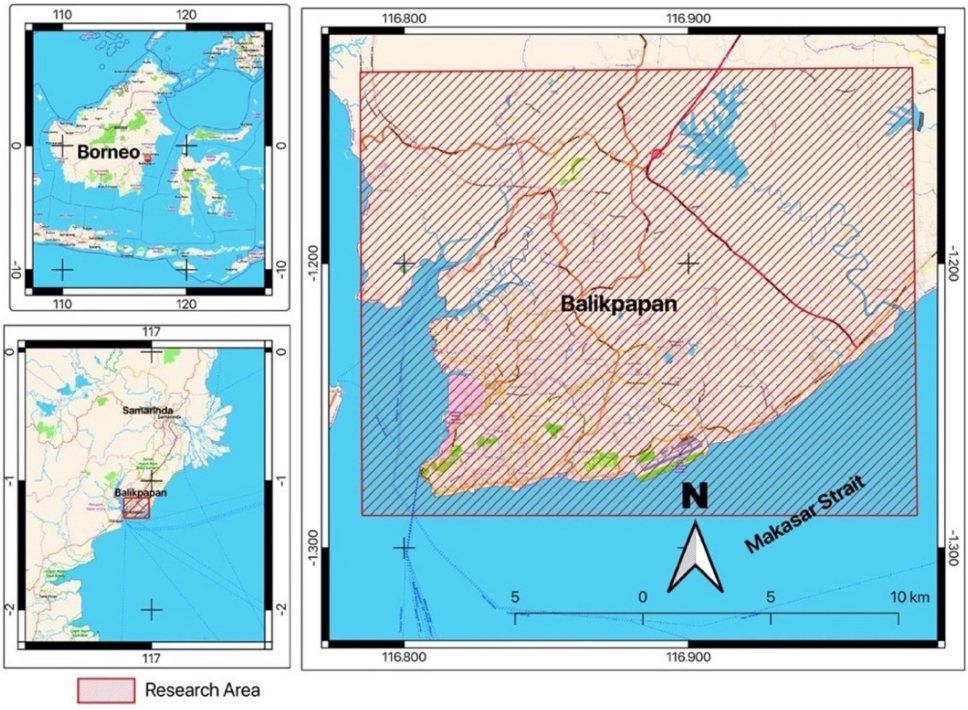

Fig. 1. Research area (adapted from OSM)
Source: OpenStreetMap OSM

The research area was located in Balikpapan City; physiographically, the city is located in the Kutei Basin [21] and administratively belongs to East Borneo Province, Indonesia (as shown in the location map in Figure 1). Geologically, three geological formations are extensively exposed on the surface: the Balikpapan Formation, Kampungbaru Formation, and recent alluvial deposits [1, 17–20].

3.2. Data Collection and Geologic Quantification

Based on a literature review that was related to the soil, infiltration, and geology of the research area, the soil, infiltration, and geologic data was available from previously published and unpublished research. The soil data consisted of soil moisture and percentages of sand, silt, and clay; it was mostly obtained from a database of the Civil Engineering Department of Balikpapan State Polytechnics. Direct field and laboratory tests were required to be conducted in order to fill the data gap in the area. The *in situ* test consisted of a double-ring infiltrometer test, and the laboratory tests consisted of soil water content as well as sieve and hydrometer tests. For interpretation and graphical presentation, the soil textures of the samples were classified using the USDA classification system. The geological data included a geological map of the area, which was published by the Indonesian Geological Agency, Ministry of Energy, and Mineral Resources (scale 1:100,000) [22], along with an unpublished report on the study geology that was attached to the geological map (scale 1:25,000) [25]. The geological map will be interpreted for its role in controlling the actual infiltration rate.

In this step, the available geological data on the lithology and geological structure were qualitative and quantified in tabular numerical format so that they could be read by machine learning. The numerical representation in the interpretation of a geological map is based on the domain of geological expertise [26]. In this research, the numerical values for both the lithological and geological structures were proposed in a scoring format that ranged from 0 to 5, with a rationalization to obtain higher scores; the lithology and geologic structures must give additional space as conduit channels that enable infiltration water to pass through spaces such as loose grain packing, intergrain spaces of clean sand, interconnected joints, and faults. Through this approach, the water infiltrates deep down. The lithologic scores were determined by the overall permeability of the geologic formation (which was made up of several lithologic units). When the formation consisted of massive intrusive igneous rock that could not store and was impermeable or an aquifuge or thick claystone and mudstone that could store but was impermeable and had a score of 0, the interbedding of claystone and other porous and permeable rocks could be scored as 1. Grains that were coarser than mud and clay had greater permeability [27] and had the potential to be an aquifer, such as sandstone, conglomerate, and breccia. The permeability of these lithologies can be scored as 2 through 5 depending on the permeability of the individual lithology; this was partially determined by the grain size. Such scoring was also based on the rank of the permeability of any common rocks [28].

The quantification of the geologic features relied on geologic formation and structural interpretation using available geological maps. The quantification was conducted by scoring the lithology and geological structure that were visible on the geological map.

3.3. ANN's Model Development and Assessment

This model development began with tabulating the data, which was headed by the actual infiltration rate [cm/h], soil moisture [%], sand [%], silt [%], clay [%], and geology. The data was normalized to avoid differences in the range of each column. The data was normalized using Equation (1), which is defined as follows:

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

After normalization, the data was randomly split into 75% for training and 25% for testing.

The next step involved modeling the relationships among the infiltration rate and the soil and geologic properties as inputs via the ANN algorithm using the neuralnet library in R Studio. An ANN is an algorithm that simulates the human nervous system; the system consists of artificially interconnected neurons (or nodes), and each node is weighted during training [29, 30]. The data processing occurs from the input layer to the hidden layer to the output layer, where the output (y) is defined by bias (b), aggregator ($\sum$), input (x), and synaptic weights (w); these processes are governed by Equation (1). In this study, y represents the actual infiltration rate, and x_1, x_2, x_3, x_4, and x_5 represent the soil moisture, sand, silt, clay, and geology, respectively:

$$y = b + \sum_{i=1}^{n} x_i \cdot w_i \tag{2}$$

The model was developed in two scenarios to determine the role of geologic control. The first scenario was a model with input-normalized soil moisture, sand, silt, clay, and geology. Moreover, the hidden layer column nodes were set by trial and error, and the output layer consisted of the actual infiltration rate. The second scenario was conducted similarly to the first, but the geologic parameters were excluded.

The prediction result of the model was then denormalized using Equation (3) as follows:

$$y = x_{\text{pred}} \cdot (x_{\max} - x_{\min}) + x_{\min} \tag{3}$$

where y is the denormalized output, x_{pred} is the output of the model prediction, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum of the actual infiltration data, respectively.

The assessment of the ANN model performance using standard statistical performance measurements consisted of the mean absolute error (MAE), the root mean squared error (RMSE), and the mean squared error (MSE) [31]. The optimum results of training and testing the ANN model were based on comparisons of the MAE, RMSE, and MSE of each model simulation; the MAEs, RMSEs, and MSEs were calculated via Equations (4), (5), and (6), respectively. The last step involved analyzing the results of all of the modeling and the conclusions.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_{pi} - Y_{oi}\right| \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{pi} - Y_{oi})^2} \tag{5}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_{pi} - Y_{oi})^2 \tag{6}$$

The first and second optimum models were determined based on the smallest MAE, RMSE, and MAE levels. The prediction results of both of the optimum models were further analyzed and compared using scatter plots of the actual-versus-predicted values to determine the roles of the geologic parameters in the soil infiltration.

4. Results

4.1. Geology and Soil Classes

The soil parameters included initial soil moisture and sand, silt, and clay percentages. Moreover, the geological properties of the lithology and structure were quantified. This research proposed the geologic scoring methods that are listed in Tables 1 and 2.

Table 1. Geologic formation in research area and proposed lithologic score

Geologic Formation	Lithologic Description	Proposed Score	Rational
Alluvium deposits (Qa)	gravel, sand, clay, mud as deposition of river, delta, and beach	depends on underlying formation	recent deposition: serving like soils above erosion plane, some parts of alluvial deposit areas are wetland

Table 1. cont.

Geologic Formation	Lithologic Description	Proposed Score	Rational
Kampungbaru (Tpkb)	sandy claystone, sand quartz, siltstone, coal interbed, limestone, and lignite in research area dominated by sandy claystone, very fine loose sand, and quartz	1	dominated by sandy claystone and siltstone, providing minimum interconnected pore space
Balikpapan (Tmbp)	research area is primarily dominated by sandstone, with alternating layers of marly clay, shale containing marl as interbed, limestone, and coal	2	dominated by sandstone, giving more interconnected intergrain pore space

Table 2. Geologic structures and proposed scores

Geologic Structure	Proposed Score	Rational
NA	0	no addition interconnected pore space in underlaying formation
Axis of syncline	0	in area of syncline axis, lithologic texture tended to have more dense grains, closed spaces, and compacted grains as result of past working stresses
Axis of anticline	1	in area of the anticline axis, will get more interconnected pore spaces, wider intergrain spaces, and even radial tension cracks in outer layer, as it had undergone strain
Joint/Fracture	2	joint or fracture will give extra interconnected space – both in permeable and impermeable strata
Fault	2	just like joint and fracture, faults also give deeper extra interconnected space – both in permeable and impermeable strata

The reason behind the geological interpretation judgment for the scoring of the geologic formations was as follows:

The Balikpapan Formation (Tmbp) consisted of sandstone, marly clay, and shale alternately, with marl as the interbed, limestone, and coal. This formation in the research area was dominated by sandstone, which could be observed visually in many outcrops located in South Balikpapan. The sand was mainly fine to medium in size; such lithology provides intergrain porosity as a conduit of infiltration water toward the groundwater table, so the lithology had better permeability than the overlying Formation of Kampungbaru (Tpkb); thus, it was given a score of 2 out of 5. On the surface, the Balikpapan Formation acted as a recharge area and functioned as an aquifer in the subsurface. The Kampungbaru Formation consisted of sandy claystone, sand quartz, siltstone, coal interbeds, limestone, and lignite [22]. Part of the Kampungbaru Formation in the research area was dominated by sandy claystone and loose sand quartz that was very fine. The sandy claystone (which was the most lithological bed and was classified as semi-impermeable and impermeable with a score of 1) allowed the Kampungbaru Formation to function as an aquitard

in hydrologic systems. Moreover, recent deposits of alluvium (Qa) were deposited above parallel and angular unconformities or erosional plains; these deposits consisted of gravel, sand, clay, and mud that were deposited by rivers, deltas, and beaches [23, 25]. As loose materials and well-sorted deposits, these deposits were additionally treated like soil, and the underlying geologic formations were used to determine the lithologic score.

Based on a direct observation, the Balikpapan Formation was more resistant to decomposition and disintegration processes than the other formations, so it had less soil coverage – especially in the South Balikpapan District and Balikpapan City Center District. Otherwise, the Kampungbaru Formation had thicker soil due to intense weathering. Based on the results of the sieve and hydrometer tests, 40 soil samples from the research area were found to be the dominant sand class (as depicted in Figure 2). Most of the samples from the Balikpapan Formation were categorized as a sand class. Moreover, the clay loam, sandy clay, and sandy loam classes were primarily associated with the Kampungbaru Formation. Alluvial deposits mostly overlaid the river and the southern line bank. This deposit stratigraphically overlaid the uppermost stratigraphic column of the area and was deposited above the unconformity; thus, the alluvial deposits acted similarly to residual soil or colluvium in terms of the actual infiltration rate.

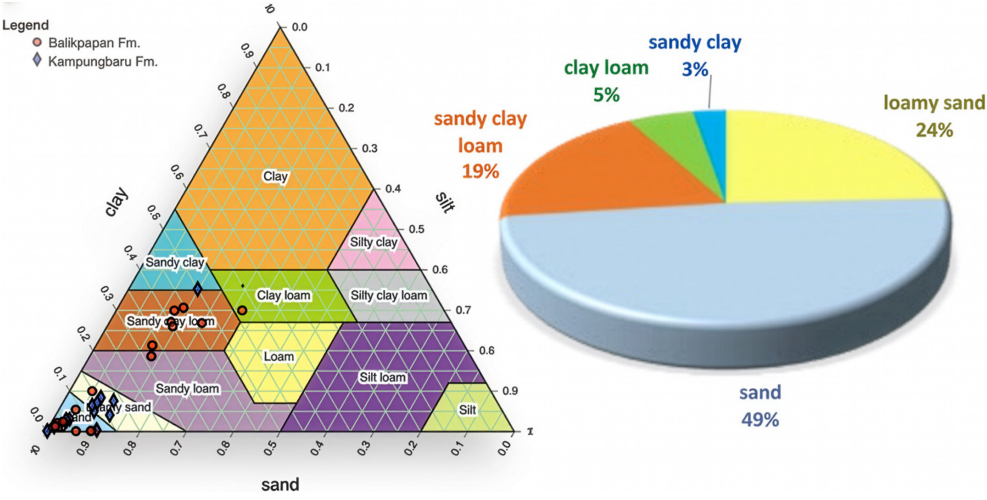

Fig. 2. Soil-texture classes of 40 samples

The scoring of geologic structures relied on the general interpretation of their natural lithologies (specifically for a monocline, which received a score of 0 because it was structural and did not contribute vertically to the waterway that passed through the formation). The syncline axis also provided no addition to the conduit even though its porosity tended to decrease as the grain packing became more

compacted; so, its geologic structure score was 0. Otherwise, the top layer of the lithology contained additional pore space on the axis of an anticline because it had experienced negative tension; so, the geologic structure score was 1. The infiltration in the soil increased the soil's moisture and ultimately saturated it; after this, the water inside the soil became stagnant. As the soil moisture increased, the soil-infiltration capacity decreased in parallel [32]. The infiltration characteristics of the saturated soil may have varied when the joints and faults existed beneath geological layers. The presence of the joints and faults gave extra conduit channels to the water that passed through geologic formations through channels that were under hydrodynamic infiltration pressure [33] through the aquifuge; when subjected, this water would serve as infiltration water passing through the barriers [34]. These conditions are illustrated in Figure 3; considering their addition to the conduits, these kinds of structures were scored as a 2. A summary of the proposed scores for the geologic structures is listed in Table 2.

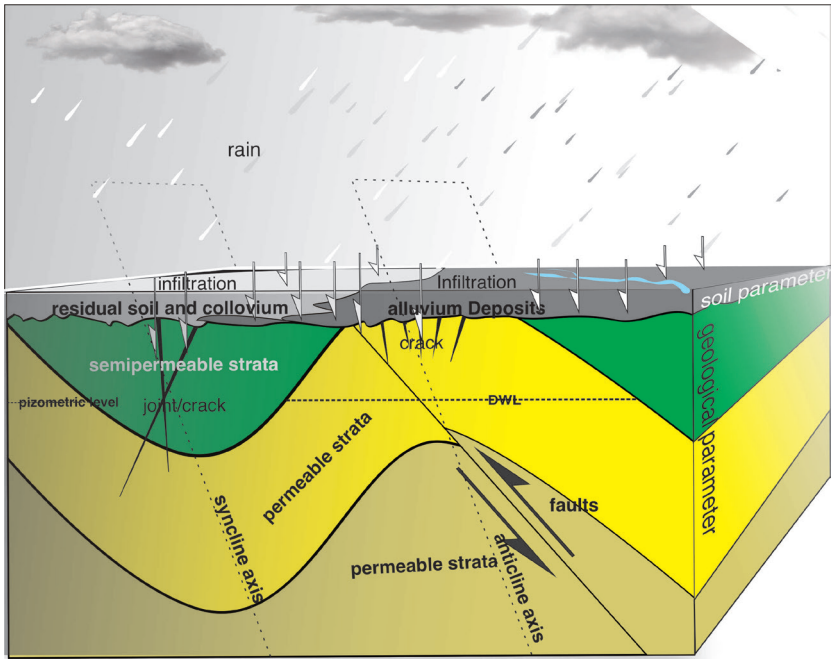

Fig. 3. The conceptual illustration of geological control on soil infiltration

The actual infiltration rate in the research area at 40 locations varied (as shown in the bar). This data represented the output variables in the ANN model. Moreover, the soil moisture, geologic score, and sand, silt, and clay contents are illustrated as colored lines; these values are treated as input variables. The code of each sample and location is written on the bottom horizontal axis of the chart in Figure 4 so it can be traced easily in the sampling location map.

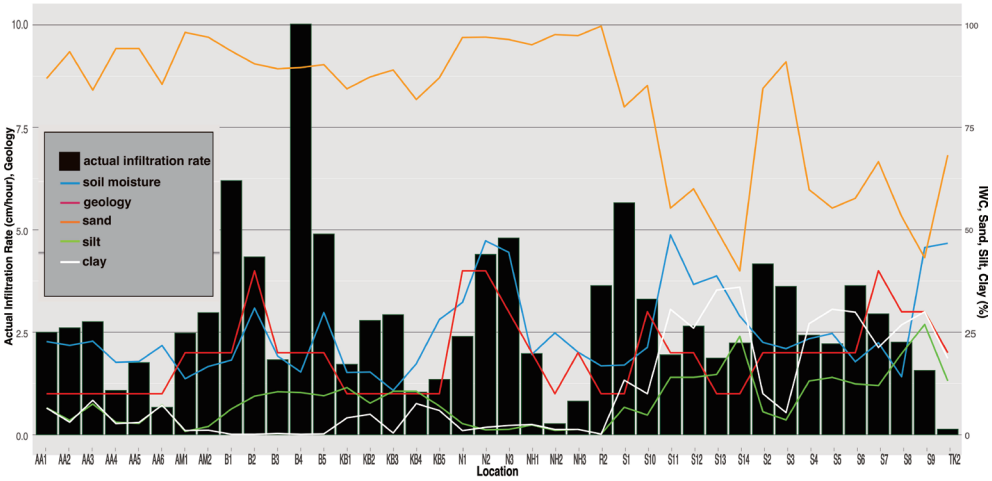

Fig. 4. Actual infiltration rates in research area

The overlay geological map and sampling location in Figure 5 were the basis of the permeability interpretation and geological scoring.

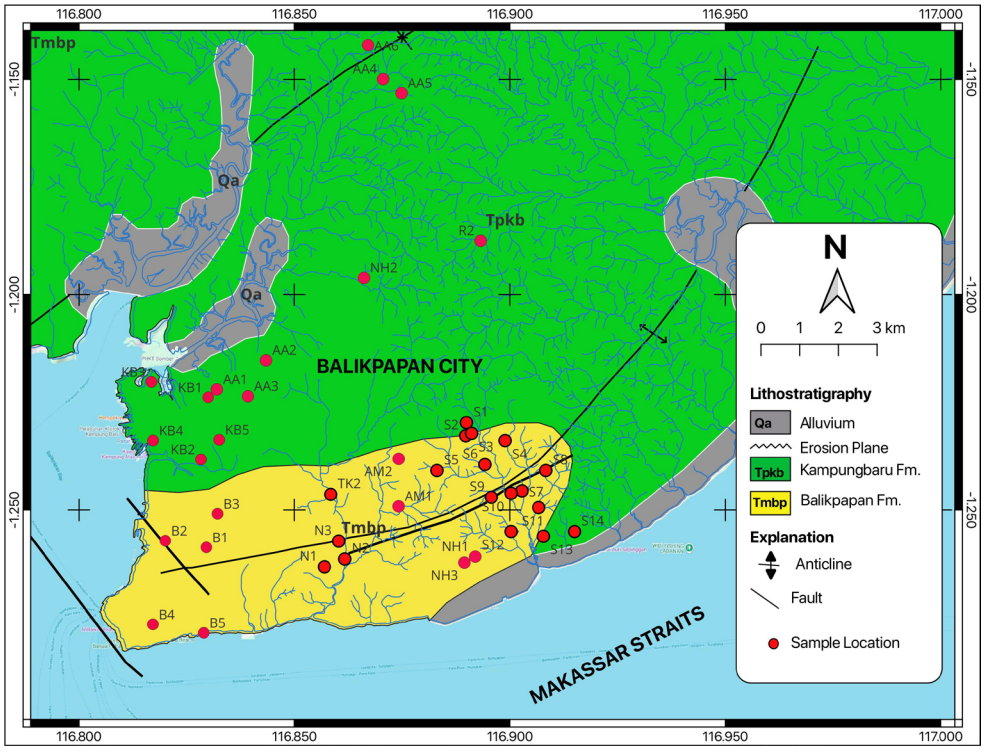

Fig. 5. Geological map of research area modified from geological map of Balikpapan City

Source: [25]

In this study, Equation (5) was used to determine that the geologic score (*Gs*) was the sum of the lithologic score (*ls*) and geologic structure (*gs*) score:

$$Gs = ls + gs \tag{5}$$

For instance, Location B2 was located in the Balikpapan Formation; based on Table 1, this location had an *ls* of 2. Structurally, the B2 was also located on the fault. According to Table 2, *gs* was 2, therefore *Gs* was 2 + 2 = 4, meanwhile *Gs* of B3 was 2 + 0 = 2, and so on.

4.2. Correlation

Figure 6 visualizes the correlation matrix of the variables. A larger dot indicates a more significant correlation between two variables on the horizontal versus vertical axes (both positively and negatively). The value inside the dot is the coefficient of correlation between the parameters in the diagonal axis. From such a figure, the coefficient of each parameter to the actual soil infiltration can be ordered from the biggest to smallest: geology 0.24, sand 0.19, silt –0.1, initial moisture content –0.13, and clay –1.022.

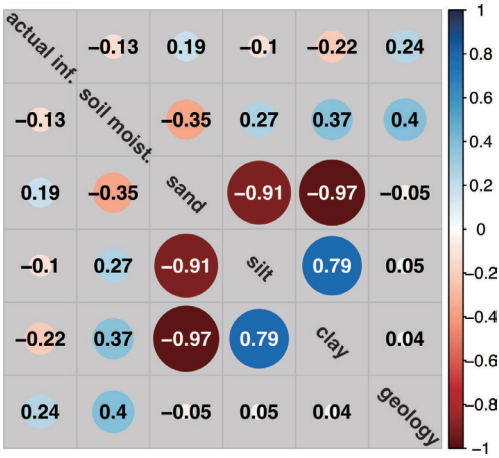

Fig. 6. Correlation matrix plot of all variables
that shows coefficient correlation of each variable to other variables

The three variables of sand, silt, and clay had greater exclusive correlations because they were the percentages of the soil grain size and the results of the sieve and hydrometer tests. The total sand, silt, and clay percentage in each sample was 100% (as required for the USDA soil-classification system). Moreover, the geologic score was strongly correlated with the actual infiltration rate (0.24). Only the sand and geologic score variables were positively correlated with the actual infiltration rate as independent variables, whereas the others (soil moisture, silt, and clay) were negatively correlated. This was a logical phenomenon in which increasing geologic coverage and

sand availability contributed to increased interconnected porosity as the channels for the infiltration of water deepened into the soil and underlying strata. Moreover, more silt and clay reduced the intergrain connected porosity [1]. Soil moisture also contributed negatively to the actual infiltration rate, as water was converted to soil moisture in the early stage of the actual infiltration; when the soil became saturated, infiltration became slower, more constant, or more stagnant. Generally, the correlations among the soil moisture content and the sand, silt, clay, and geologic data and the actual infiltration rate had no linear pattern (as shown in Figure 4).

In this study, the multivariate linear regression (MLR) method was tried, but the results were not really linearly correlated between the actual infiltration as the dependent variable and the soil moisture, sand, silt, clay, and geology as the independent variables. Such an MLR with the function formula code of in R (formula = actual_inf. ~ soil_moist. + sand + silt + clay + geology, data = training_set) had a coefficient of determination (R^2) of prediction versus actual of 0.1577; thus, the MLR model was not suitable for the complex correlations of soil moisture, soil texture, and geology as dependent variables versus actual infiltration as the independent variable. Therefore, this research proposed the use of the ANN model to solve these problems.

4.3. ANN's Model of Actual Infiltration Rate of Soil

The simulation of the ANN model involved 40 rows of infiltration data, starting with training the two ANN formula functions. The first model function was represented by formula code in R (actual_inf. ~ soil_moist. + sand + silt + clay + geology). To determine the geologic roles of the infiltration of the model, the first result was compared to the results of the second model function (actual_inf. ~ soil_moist. + sand + silt + clay). The hidden layers were activated using the logistic function (sigmoid), and the output layer was the identity function for the regression problems or the softmax function. To reach the minimum error in the training processes, the threshold was set to 0.001. To obtain the optimum ANN model for determining the actual soil-infiltration rate for each function, nine trial-and-error simulations were conducted using different hidden-layer-setting scenarios, starting with 4 and 2 nodes and going up to 12 and 2 nodes in the hidden layer. During the training and simulation processes, the errors, steps, MAE, RMSE, and MSE of each model were recorded (these are shown in Tables 3 and 4, and Figure 7). Based on the series of recorded values of MAE, RMSE, and MSE (Fig. 7), the optimum ANN model for the actual soil-infiltration rate could be determined. The optimum first model was the one with nine nodes and two nodes in the hidden layer [c(9,2)], as it had the highest levels of MAE, RMSE, and MSE. The training and testing results of the model with the hidden layer setting that used fewer nodes or more than c(9,2) nodes showed that the MAE, RMSE, and MSE levels increased slightly. The optimum ANN model architecture consisted of an input layer interlinked to the hidden layer of c(9,2) that was further interlinked with the output layer (as illustrated in Figure 8).

Table 3. Values of MAE, RMSE, MSE, Error, and Steps for each ANN model with geologic parameters

Error Parameter	Hidden Layer								
	c(4,2)	c(5,2)	c(6,2)	c(7,2)	c(8,2)	c(9,2)	c(10,2)	c(11,2)	c(12,2)
MAE	1.0448	0.9244	0.6834	0.6382	0.7047	0.4512	0.5314	0.7832	0.6819
RMSE	2.2994	1.8952	1.6490	1.5130	1.8955	1.1051	1.2377	1.8926	1.6870
MSE	5.2873	3.5920	2.7191	2.2891	3.5930	1.2213	1.5318	3.5819	2.8460
Error	0.0075	0.0261	0.0023	0.0006	0.0007	0.0065	0.0027	0.0005	0.0003
Steps	9,958	1,645	1,873	2,711	7,067	6,870	1,603	2,328	13,649

Table 4. Values of MAE, RMSE, Error, and Steps for each ANN model for which geology is excluded

Error Parameter	Hidden Layer								
	c(4,2)	c(5,2)	c(6,2)	c(7,2)	c(8,2)	c(9,2)	c(10,2)	c(11,2)	c(12,2)
MAE	1.1849	0.9244	0.7144	0.5929	0.7675	0.4740	0.6159	0.6820	0.5689
RMSE	3.0334	2.2098	1.5642	1.4989	1.6365	1.3467	1.5383	1.6969	1.2681
MSE	9.2015	4.8834	2.4467	2.2468	2.6782	1.8136	2.3664	2.8794	1.6082
Error	0.0093	0.0070	0.0071	0.0012	0.0061	0.0010	0.0116	0.0033	0.0022
Steps	21,281	10,293	3,503	7,222	3,418	10,032	3,103	2,373	3,669

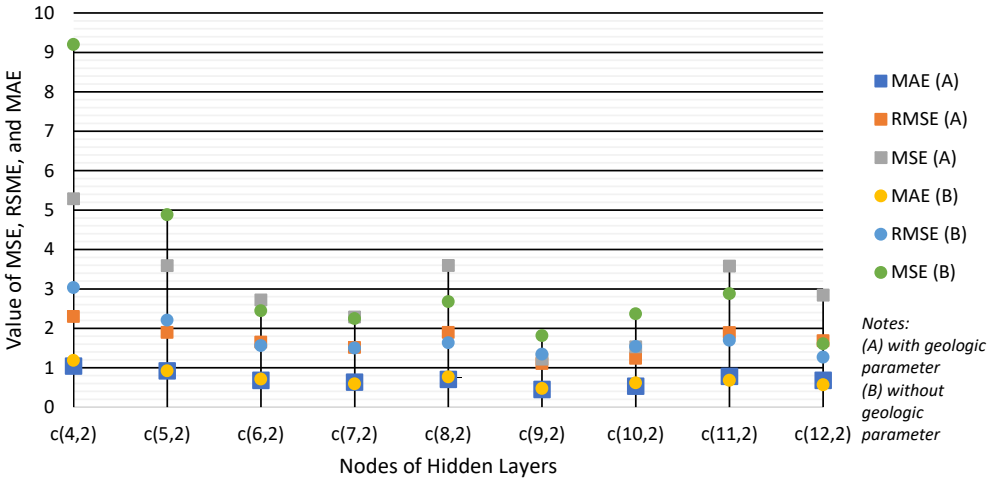

Fig. 7. Series plots of MAE, RMSE, and MSE for each model

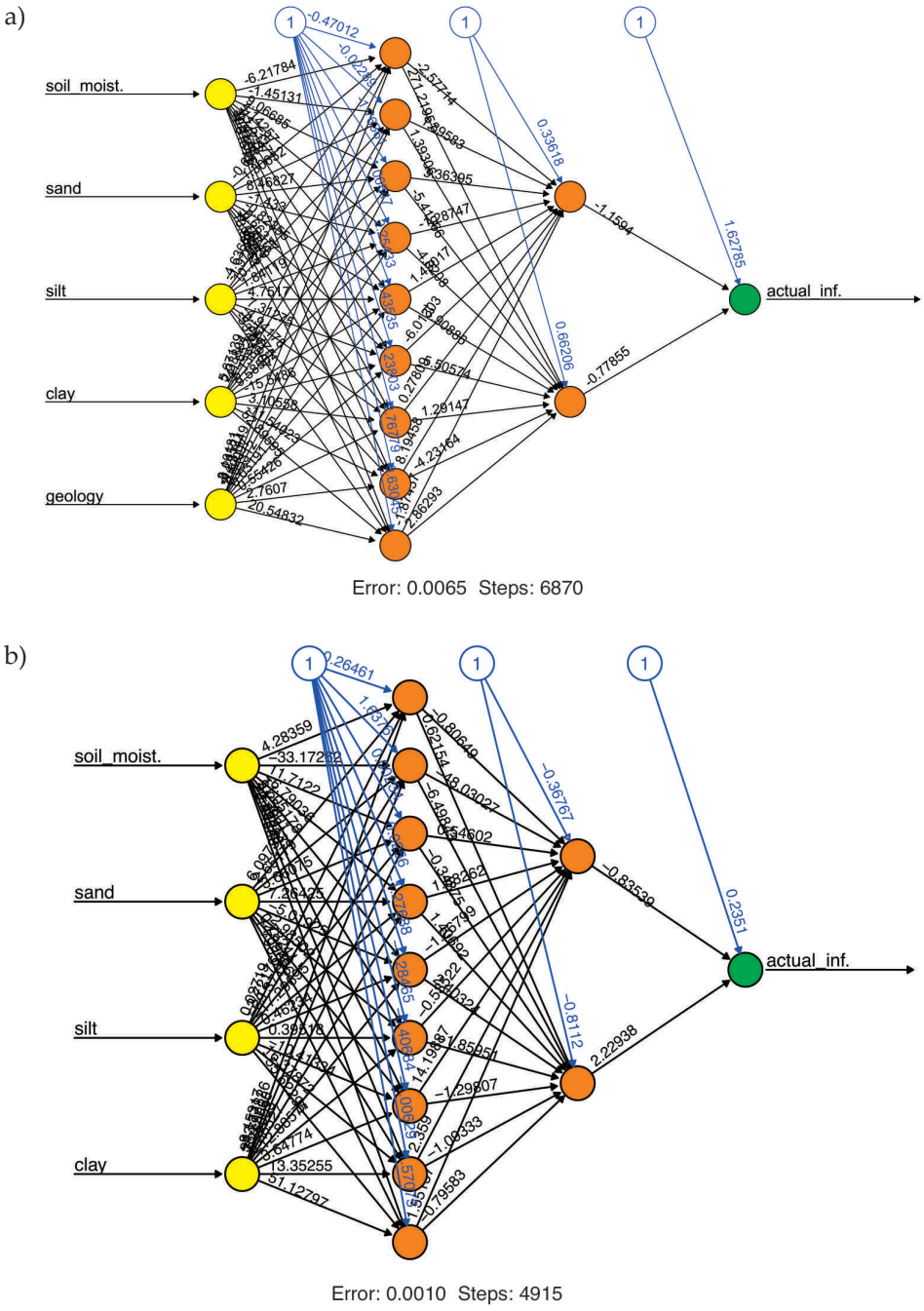

Fig. 8. Optimum ANN infiltration configuration models:
a) with 9 and 2 nodes of hidden soil textures and geology as inputs;
b) with 9 and 2 nodes of hidden layer geologic parameters excluded

In the second experiment, the geologic inputs were excluded, and the optimum model was coincidentally found in a configuration with a hidden layer setting of c(9,2); this had fewer nodes than the other configurations but resulted in higher MAE, RMSE, and MSE levels. The second optimum ANN model without geologic input results is listed in Table 4 and depicted in Figure 7. In this figure, the overlay of the line charts for the model metrics of both the first and second optimum models illustrates a comparison of their results.

4.4. Model Accuracy and Precision

The first optimum ANN model for the actual soil infiltration (using the geological parameters as input) was achieved with a hidden layer functional argument of c(9,2), resulting in an error of 0.0065 and a total of 6870 steps; its configuration is illustrated in Figure 8a. Moreover, the second optimum model without the geological parameters was reached when the hidden layer was set to c(9,2) with an error of 0.0010 and 10,032 steps; its ANN model configuration is illustrated in Figure 8b.

In addition to relying on the MAE, RMSE, and MSE levels of the optimum model, a scatter plot of the actual infiltration rate versus the prediction value of the infiltration rate was used to further assess the accuracy and precision of the optimum model in this research (as presented in Figure 9).

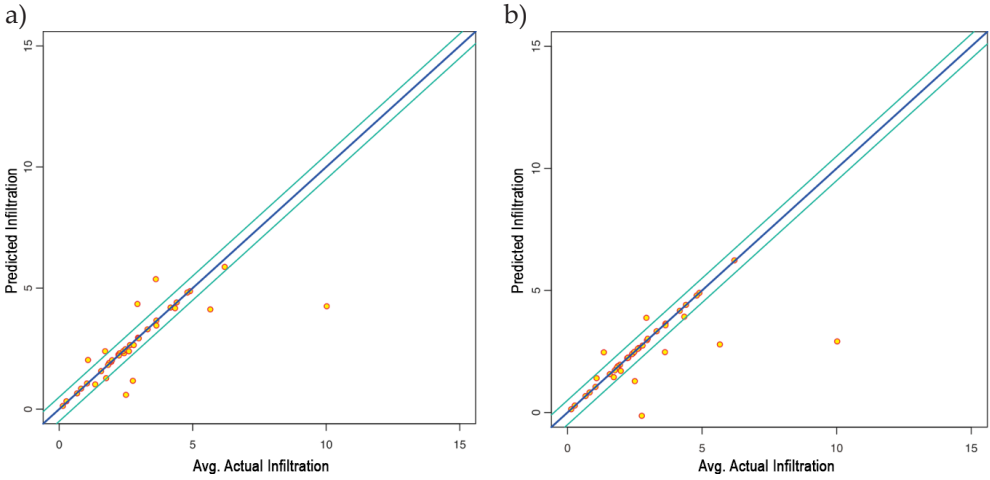

Fig. 9. Scatter plot with fitting lines of actual infiltration rate versus result of prediction of optimum ANN model: a) with geologic parameter; b) without geologic parameter

To visualize the fit between the prediction of the actual of the infiltration rate, the blue diagonal line represents the fitting line of the actual and prediction data, while the green line represents the line where the error was borderline more than 0.5; this could also help to judge the model's precision. The dots represent the actual and

predicted actual infiltration rates; a dot that is closer to the blue line indicates its higher accuracy. In the model with the geologic parameters, there are more red dots that are closer to the fitting line than in the second scatter plot (which is the prediction result of the optimum model without the geologic parameters).

5. Discussion

Based on the results of both the sieve and hydrometer tests on the soil samples (and the USDA soil classification system), the research area was dominated by relatively clean sand, which consisted of 49% sand and 24% loamy sand; the remaining 27% consisted of sandy clay loam, clay loam, and sandy clay (which is a mixture of clay, silt, and sand – as illustrated in Figure 2). These soil classes reflected the availability of connected intergrain pore spaces, as more than 73% of the soil tended to permeate and had greater infiltration rates. The remaining areas were impermeable and had lower infiltration rates [24].

To determine the role of geologic activity in actual soil infiltration, geological attribute scoring can be applied in geological hazard mapping [35]. The geological scoring was conducted to turn the descriptive and qualitative data into quantitative data based on Tables 1 and 2. Moreover, the geological expertise domain could be biased; however, the objectivity of the geological score was generally confirmed with a correlation plot, a logical explanation, and an analysis of the correlations among all of the variables. The geologic score focuses on the additional interconnected pore space, whether provided by the soil's lithological texture or geological structures; this increases the actual soil infiltration. The correlation matrix plot in Figure 6 shows a positive correlation between the geologic score and the actual soil-infiltration rate; its correlation coefficient was 0.24 out of 1, which demonstrated that the geological parameter played a significant role as compared to the others.

For instance, the soil moisture negatively correlated with the actual infiltration rate. This correlation could be further explained by how the infiltration of water filled the soil's porosity and increased the amount of soil pore water or moisture. The higher soil moisture decreased the infiltration rate [36]. This phenomenon was proven by the negative correlation between the soil moisture and the actual infiltration rate in the correlation plot in Figure 6; its correlation coefficient was –0.13 out of –1. The percentage of sand in the soil positively correlates with the actual soil infiltration rate because increased sand intergrain porosity allows water to flow through the intergrain spaces in relatively clean sand; higher sand content results in greater interconnected porosity, and the correlation coefficient, as shown in Figure 6, is 0.19. Moreover, silt and clay contents negatively correlate with the actual infiltration rate. The correlation coefficients are –0.1 and –0.22, respectively.

The correlation plot played an imperative role in determining the actual soil infiltration and the quantitative geological parameters, so these parameters needed to

be included as inputs in the development of an ANN model of the actual infiltration. The model development involved the initial soil moisture, soil texture, and geologic parameters as inputs and the actual soil infiltration rate as the output; despite the ANN model, the weights of the geological parameters could not be determined. However, the results of the optimum ANN model prediction with 9 and 2 hidden nodes [c(9,2)] yielded accurate prediction results, as was proven by the MAE levels of 0.4512 and 1.1051 and MSE levels of 1.2213 (listed in Table 3 and depicted in the overlay line chart of the model metrics in Figure 7). The accuracy of the first optimum prediction was also confirmed visually by the scatterplot in Figure 9a; this showed that the model accuracy was higher than that of the second optimum model, as indicated by intersection of dots and fitting line. The strengths of the first optimum model were that it was more accurate and has fewer errors, and its weakness was that it had less precision.

The analysis was also conducted by comparing the metrics of the first optimum model (which are listed in Table 3) and the second optimum model (which are listed in Table 4 and depicted in Figure 7). Table 4 shows the results of the series model without the geologic parameters. The optimum ANN model was reached with two hidden layers and 8 and 2 nodes. The second model-prediction result reached an optimum MAE of 0.5268, an RMSE of 1.3609, and an MSE of 1.8520; these were higher than those of the first optimum model. This meant that the second optimum model was less accurate than the first. According to Figure 9b, however, the second optimum ANN model c(9,2) was more precise than the first optimum ANN model c(9,1); in Figure 9b, more dots are closer to the fitting line and within the two green lines (although the fitting lines intersect fewer dots). Therefore, the strength of the second optimum model was more precise, and its weaknesses were less accurate and resulted in greater errors during the training (as shown in Figure 8b).

Further evaluation and assessment of the developed model were conducted using the Taylor diagram (which was generated via the open-air package in R Studio) [37]. All of the models in the first experiment were plotted in the Taylor diagram in Figure 10a, wherein the Taylor diagram model c(9,2) had the best performance because it had the most significant correlation coefficient and the smallest RMSE. However, it had a mediocre standard deviation, and the gray cross symbol represented the model. Moreover, the second experiment (Fig. 10b) showed that, while Model c(9,2) had the highest petite MAE, Model c(12,2) demonstrated the best performance in terms of the coefficient correlation, RMSE, and standard deviation (as indicated by the green dots). The overall model that was developed in the first experiment exhibited less variation in the results when compared to the second experiment.

With the use of the data on soil moisture, soil texture, and USDA soil class as well as a geologic map of the research area, the actual soil-infiltration rate at any point in the research area could be accurately predicted via the first optimum ANN model. The simulation could also be performed using the developed model. An accurate

model will reduce the time and costs of projects that measure field infiltration. This approach will benefit mitigation planning in hydrometeorological disasters such as floods and landslides. It will also be advantageous in other planning areas, such as drainage, agriculture, and groundwater management.

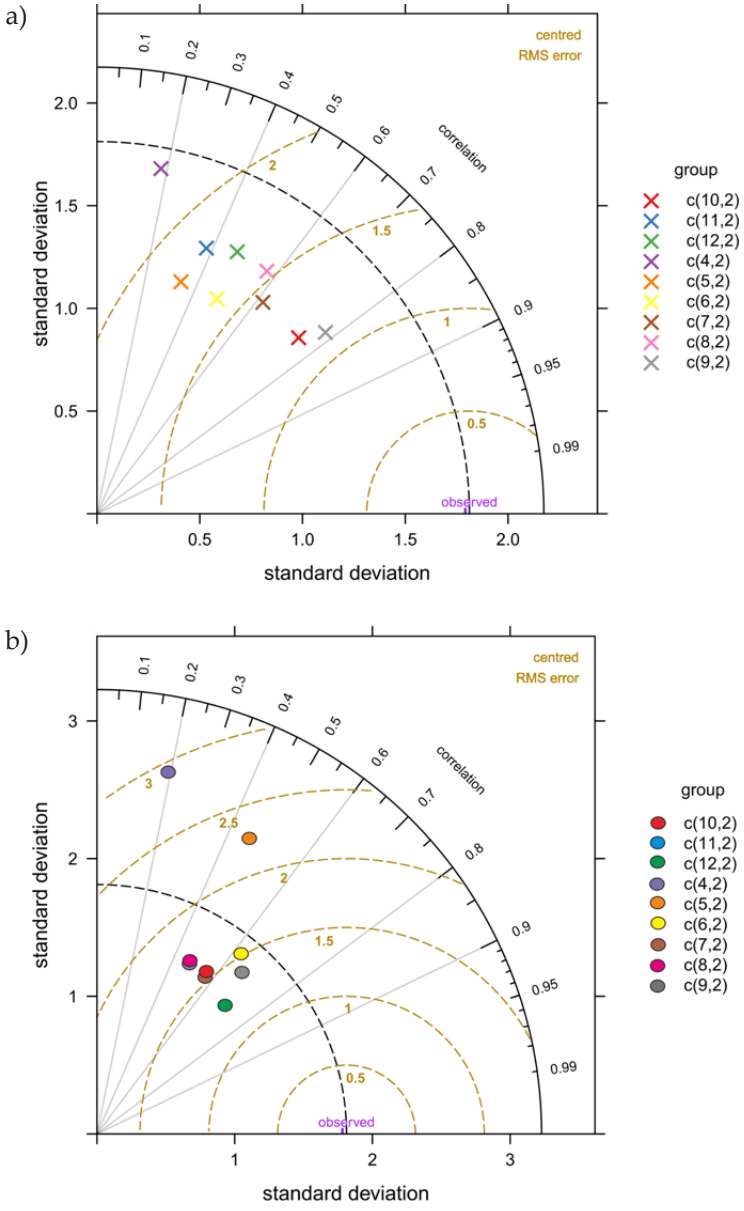

Fig. 10. Taylor diagrams:
a) models with geologic parameters; b) models without geologic parameters

The MAE, RMSE, and MSE levels of the first optimum model proved that the model was relatively accurate and that the geologic parameters played an essential role in determining the actual infiltration rate; this mainly controlled the continual movement of infiltration water under the soil layer. The weakness of this research was the lack of subsurface data, such as geophysical data, the RQD of the cored samples, and the coefficient of the permeability (k) of the jointed geologic strata [38, 39]; however, the subsurface data could be obtained by using the available geological map. Undeniably, the qualitative geology could be better quantified and scored when supported by subsurface data. In addition to the aim of this research, the mechanism through which the water infiltrated the geologic strata and moved continually should be further studied in future research.

The strength of the model is providing geological parameters, not merely soil parameters; in most places, these are not thicker than the underlying geologic strata. The weaknesses of this model are the 40 data points of soil infiltration, soil moisture, percentage of sand, percentage of silt, and percentage of clay. It also lacks subsurface geological data such as drilling, seismic data, and geoelectric data. The geology is merely based on the interpretation of the lithological and geological structure of a geological map; by providing more such geophysical data, it will provide quantitative and exact data as a parameter in the ANN model.

6. Conclusion

The soil texture showed a negative correlation with the actual infiltration rate (except for the percentage of sand). The geologic score, which reflected the lithology and geological structure, was positively correlated with the actual infiltration rate and played a more significant role than the other parameters (as indicated by its correlation coefficient of 0.24). The initial water content was also negatively correlated with the actual infiltration rate. The ANN infiltration model was developed with soil moisture, sand, silt, clay, and geology as inputs. The hidden layer consisted of two columns, which were determined through a series of trial-and-error tests; the node counts ranged from 4 to 12, while the second column contained two nodes. The output layer represented the actual infiltration rate. After the training, the first optimal model was achieved with nine and two nodes, yielding an error of 0.0065 and requiring 6870 steps during the training. This model attained an MAE of 0.4512, an RMSE of 1.1051, and an MSE of 1.2213. In contrast, the optimum second ANN model (where the geologic parameters were excluded) was reached with nine nodes and two nodes in the hidden layers, with 0.0010 errors and 4915 steps; additionally, the model had an MAE of 0.5268, an RMSE of 1.3609, and an MSE of 1.8520. The model proved that geology played a vital role in controlling the actual soil-infiltration rate at each sampling location. The implication of this research will open new research venues for other quantitative geologic parameters in infiltration and the water cycle.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Author Contribution

T. S.: writing – original draft, visualization, validation, methodology, investigation, formal analysis, data curation, conceptualization.

S. B. K.: review, supervision, methodology, data curation, conceptualization.

T. A. C.: review & editing, supervision, methodology, conceptualization.

R. A. F: review, supervision, methodology, conceptualization.

M. K.: writing (editing), data curation.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data will be made available upon request.

Use of Generative AI and AI-Assisted Technologies

No generative AI or AI-assisted technologies were employed in the preparation of this manuscript.

Acknowledgments

Special gratitude goes to the head of the Department of Civil Engineering of Balikpapan State Polytechnics, which permitted the use of its laboratory facilities.

References

[1] Sulistyo T., Fauzi R.: *Soil infiltration rate prediction using machine learning regression model: A case study on Sepinggan River Basin, Balikpapan, Indonesia*. Indonesian Journal on Geoscience, vol. 10(3), 2023, pp. 335–347. https://doi.org/10.17014/ijog.10.3.335-347.

[2] Deulkar A.M., Londhe S.N., Jain R.K., Dixit P.R.: *Rainfall-runoff modelling using artificial neural networks (ANNs) for Upper Krishna Basin, Maharashtra, India*, [in:] Timbadiya P.V., Patel P.L., Singh V.P., Mirajkar A.B. (eds.), *Geospatial and Soft Computing Techniques: HYDRO 2021*, Lecture Notes in Civil Engineering, vol. 339, Springer, Singapore 2023, pp. 439–450. https://doi.org/10.1007/978-981-99-1901-7_35.

[3] Alfaro I.A., Chavez J.A., Cuestas I.E., Mejía C.J., Landaverde M., Campos S.: *Study on infiltration and its relationship with the geology of the Metropolitan Area of San Salvador, El Salvador*. Revista Geológica de América Central, no. 63, 2020, pp. 40–57.

[4] Basset C., Abou Najm M., Ghezzehei T., Hao X., Daccache A.: *How does soil structure affect water infiltration? A meta-data systematic review*. Soil & Tillage Research, vol. 226, 2023, 105577. https://doi.org/10.1016/j.still.2022.105577.
[5] Cleophas F., Isidore F., Musta B., Mohd Ali B.N., Mahali M., Zahari N.Z., Bidin K.: *Effect of soil physical properties on soil infiltration rates*. Journal of Physics: Conference Series, vol. 2314(1), 2022, 012020. https://doi.org/10.1088/1742-6596/2314/1/012020.
[6] Cui Z., Wu G.L., Huang Z., Liu Y.: *Fine roots determine soil infiltration potential than soil water content in semi-arid grassland soils*. Journal of Hydrology (Amsterdam), vol. 578, 2019, 124023. https://doi.org/10.1016/j.jhydrol.2019.124023.
[7] Abdelkareem M., Al-Arifi N.: *The use of remotely sensed data to reveal geologic, structural, and hydrologic features and predict potential areas of water resources in arid regions*. Arabian Journal of Geosciences, vol. 14(704), 2021, pp. 2–15. https://doi.org/10.1007/s12517-021-06942-6.
[8] Beetle-Moorcroft F., Shanafield M., Singha K.: *Exploring conceptual models of infiltration and groundwater recharge on an intermittent river: The role of geologic controls*. Journal of Hydrology: Regional Studies, vol. 35, 2021, 100814. https://doi.org/10.1016/j.ejrh.2021.100814.
[9] Sulistyo T., Respati S.: *The estimation of flood area based on a few selected and weighted parameters: case study of the Nangka River basin, Balikpapan (Indonesia)*. Journal of the Geographical Institute Jovan Cvijic SASA, vol. 73(2), 2023, pp. 123–137. https://doi.org/10.2298/IJGI2302123S.
[10] Dahan O., Shani Y., Enzel Y., Yechieli Y., Yakirevich A.: *Direct measurements of floodwater infiltration into shallow alluvial aquifers*. Journal of Hydrology (Amsterdam), vol. 344(3–4), 2007, pp. 157–170. https://doi.org/10.1016/j.jhydrol.2007.06.033.
[11] Shanafield M., Cook P.G.: *Transmission losses, infiltration and groundwater recharge through ephemeral and intermittent streambeds: A review of applied methods*. Journal of Hydrology (Amsterdam), vol. 511, 2014, pp. 518–529. https://doi.org/10.1016/j.jhydrol.2014.01.068.
[12] Iqbal M., Al-Hassan M.A., Herdianita N.R., Juliarka B.R.: *Determining recharge area in ULUBELU geothermal field, LAMPUNG, Indonesia using stable isotope data*. Applied Geochemistry, vol. 156, 2023, 105763. https://doi.org/10.1016/j.apgeochem.2023.105763.
[13] Fajar R.A., Handayani G., Widodo L.E., Notosiswoyo S., Pamungkas T.C.: *Physical model of vertical water movement inside a soil-column apparatus for infiltration study with a two-way orientation approach*. Journal of Engineering and Technological Sciences, vol. 51(5), 2019, pp. 615–631. https://doi.org/10.5614/j.eng.technol.sci.2019.51.5.2.
[14] Brunner P., Cook P.G., Simmons C.T.: *Hydrogeologic controls on disconnection between surface water and groundwater*. Water Resources Research, vol. 45(1), 2009, W01422. https://doi.org/10.1029/2008WR006953.
[15] Megahed H.A., Farrag A.E.H.A., Mohamed A.A., D'Antonio P., Scopa A., AbdelRahman M.A.E.: *Groundwater recharge potentiality mapping in Wadi*

Qena, Eastern Desert Basins of Egypt for sustainable agriculture base using geomatics approaches. Hydrology, vol. 10(12), 2023, 0237. https://doi.org/10.3390/hydrology10120237.

[16] Tamea S., Butera I.: *Stochastic description of infiltration between aquifers*. Journal of Hydrology (Amsterdam), vol. 510, 2014, pp. 541–550. https://doi.org/10.1016/j.jhydrol.2013.12.007.

[17] Bergeson C.B., Martin K.L., Doll B., Cutts B.B.: *Soil infiltration rates are underestimated by models in an urban watershed in central North Carolina, USA*. Journal of Environmental Management, vol. 313, 2022, 115004. https://doi.org/10.1016/j.jenvman.2022.115004.

[18] Ramadhan R., Marzuki M., Suryanto W., Sholihun S., Yusnaini H., Muharsyah R., Hanif M.: *Trends in rainfall and hydrometeorological disasters in new capital city of Indonesia from long-term satellite-based precipitation products*. Remote Sensing Applications: Society and Environment, vol. 28, 2022, 100827. https://doi.org/10.1016/j.rsase.2022.100827.

[19] Jia N., Yang Z., Xie M., Mitani Y., Tong J.: *GIS-based three-dimensional slope stability analysis considering rainfall infiltration*. Bulletin of Engineering Geology and the Environment, vol. 74(3), 2015, pp. 919–931. https://doi.org/10.1007/s10064-014-0661-1.

[20] Balikpapan City Government: *Peraturan Daerah Kota Balikpapan Nomor 6 Tahun 2021 Tentang Rencana Pembangunan Jangka Menengah Daerah Tahun 2021–2026* [Local Government Regulation No. 6, 2021 Concerning to the Balikpapan City Development Plan 2021–2026]. https://jdih.balikpapan.go.id/dokumen/456/detail.

[21] Van Bemmelen R.W.: *The Geology of Indonesia*. V.F.A. Government Printing Office, The Hague 1949.

[22] Hidayat S., Umar I.: *Peta Geologi Lembar Balikpapan, Kalimantan* [Geological Map of the Balikpapan Quadrangle, Kalimantan]. GeoMap. https://geologi.esdm.go.id/geomap/pages/preview/peta-geologi-lembar-balikpapan-kalimantan [access: May 27, 2023].

[23] Sulistyo T., Abrar A.: *Characterization of thin alluvial bed aquifers in Manggar River Balikpapan East Kalimantan Indonesia*. Jurnal Teknologi Terpadu (JTT), vol. 5(1), 2017, p. 54. https://doi.org/10.32487/jtt.v5i1.212.

[24] Pristianto H., Bisri M., Suhartanto E.: *Soil textures-based evaluation of Horton and Philip's infiltration models for equatorial small watersheds*. Journal of Ecological Engineering, vol. 25(2), 2024, pp. 103–114. https://doi.org/10.12911/22998993/176319.

[25] Government of Balikpapan City, Regional Development Planning Agency (BAPEDALDA): *Laporan Akhir: Kajian Geologi Untuk Evaluasi Penataan Wilayah Dan Pengembangan Kota Balikpapan* [Final Report: Geological Study for the Evaluation of Spatial Planning and Development of Balikpapan City]. Balikpapan 2002. https://drive.google.com/file/d/1RJk52cwM0PC29Y-sMxDrG7mFr23SjxahA/view?usp=sharing.

[26] Deisenroth M.P., Faisal A.A., Ong C.S.: *Mathematics For Machine Learning*. Cambridge University Press, Cambridge, New York 2019.

[27] Yang Y., Aplin A.C.: *A permeability–porosity relationship for mudstones*. Marine and Petroleum Geology, vol. 27(8), 2010, pp. 1692–1697. https://doi.org/10.1016/j.marpetgeo.2009.07.001.
[28] Lewis M.A., Cheney C.S., Ó Dochartaigh B.É.: *Guide to Permeability Indices*. British Geological Survey, Keyworth 2006. https://nora.nerc.ac.uk/id/eprint/7457/1/CR06160N.pdf [access: September 3, 2025].
[29] Nunes da Silva I., Spatti D.H., Flauzino R.A., Bartocci Liboni L.H., Franco dos Reis Alves S.: *Artificial Neural Networks: A Practical Course*. Springer, Cham 2017. https://doi.org/10.1007/978-3-319-43162-8.
[30] Respati S., Sulistyo T.: *The effect of the number of inputs on the spatial interpolation of elevation data using IDW and ANNs*. Geodesy and Cartography (Vilnius), vol. 49(1), 2023, pp. 60–65. https://doi.org/10.3846/gac.2023.16591.
[31] Cheng S., Qiao X., Shi Y., Wang D.: *Machine learning for predicting discharge fluctuation of a karst spring in North China*. Acta Geophysica, vol. 69(1), 2021, pp. 257–270. https://doi.org/10.1007/s11600-020-00522-0.
[32] Ruggenthaler R., Meißl G., Geitner C., Leitinger G., Endstrasser N., Schöberl F.: *Investigating the impact of initial soil moisture conditions on total infiltration by using an adapted double-ring infiltrometer*. Hydrological Sciences Journal, vol. 61(7), 2016, pp. 1263–1279. https://doi.org/10.1080/02626667.2015.1031758.
[33] Ma J., Zeng R., Yao Y., Meng X., Meng X., Zhang Z., Wang H., Zhao S.: *Characterization and quantitative evaluation of preferential infiltration in loess, based on a soil column field test*. Catena, vol. 213, 2022, 106164. https://doi.org/10.1016/j.catena.2022.106164.
[34] Zheng Q., Wang C., Zhu Z.: *Research on the prediction of mine water inrush disasters based on multi-factor spatial game reconstruction*. Geomechanics and Geophysics for Geo-Energy and Geo-Resources, vol. 10, 2024, 41. https://doi.org/10.1007/s40948-024-00761-1.
[35] [Pachauri A.K., Pant M.: *Landslide hazard mapping based on geological attributes*. Engineering Geology, vol. 32(1–2), 1992, pp. 81–100. https://doi.org/10.1016/0013-7952(92)90020-Y.
[36] Yang M., Zhang Y., Pan X.: *Improving the Horton infiltration equation by considering soil moisture variation*. Journal of Hydrology (Amsterdam), vol. 586, 2020, 124864. https://doi.org/10.1016/j.jhydrol.2020.124864.
[37] Carslaw D.C., Ropkins K.: *openair – An R package for air quality data analysis*. Environmental Modelling & Software, vol. 27–28, 2012, pp. 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008.
[38] Cahyadi T.A., Syihab Z., Widodo L.E., Notosiswoyo S., Widijanto E.: *Analysis of hydraulic conductivity of fractured groundwater flow media using artificial neural network back propagation*. Neural Computing and Applications, vol. 33(1), 2021, pp. 159–179. https://doi.org/10.1007/s00521-020-04970-z.
[39] Berkowitz B.: *Characterizing flow and transport in fractured geological media: A review*. Advances in Water Resources, vol. 25(8), 2002, pp. 861–884. https://doi.org/10.1016/S0309-1708(02)00042-8.