

EREM 78/3

Journal of Environmental Research,
Engineering and Management
Vol. 78 / No. 3 / 2022
pp. 39–55
DOI 10.5755/j01.erem.78.3.31482

The Effect of Training Data Series Length on the Performance of the Tank Model for Transforming Rainfall into Runoff Data Series

Received 2022/05

Accepted after revision 2022/08

<http://dx.doi.org/10.5755/j01.erem.78.3.31482>

The Effect of Training Data Series Length on the Performance of the Tank Model for Transforming Rainfall into Runoff Data Series

Sulianto*, Ernawan Setyono, Lourina Evanale Orfa

Department of Civil Engineering, Engineering Faculty, University of Muhammadiyah Malang, Indonesia

***Corresponding author:** sulianto1967@gmail.com

The Tank model by Sugawara is included in the lumped model category. As with other types of lumped models, the effectiveness of the application of the Tank model is largely determined by the parameter optimization method applied and the quantity of training data involved in the calibration process. This article proposes the Tank-DE model to transform rain data series into discharge in a watershed. The Tank-DE model is built from a combination of a simulation equation system based on the Tank model and a multi-parameter optimization equation system based on the Differential evolution (DE) Algorithm. This article also examines the sensitivity analysis of the model to study the effect of the length of the training data series involved in the calibration process on the predictive discharge quality generated by the Tank-DE model. Thus, the minimum length of the training data series can be recommended, related to the application of the model. The results of the analysis show that the Tank-DE model can present the relationship between rainfall data series and daily period discharge very well. The results of the sensitivity analysis show that there is an indication that the longer the training data series, the more quantitatively positive impact on the performance of the model. The calibration process involving a training data set for 1 year produces a very good value of the coefficient of determination ($r^2 = 0.94$), but the indicator decreases drastically at the validation stage. The calibration process involving a relatively long training data series produces a more consistent value of the coefficient of determination. This indicates that the Tank-DE model can be an alternative solution to solve the problem of scarcity of discharge data series which is a classic problem in water resource development activities.

Keywords: differential evolution, length training data, rainfall-runoff, tank models.

Introduction

In hydraulic building engineering activities, the dimensions and hydraulic characteristics of its components are largely determined by the discharge value which is used as a design benchmark. The design benchmarks can only be determined accurately if the building plan site provides information about discharge fluctuations in sufficient quantity and quality. Therefore, the effort to extend the historic discharge data series becomes a necessity, if at the planned site there is no sufficient observational data series available to determine the design benchmark size.

There are two classic problems in planning and designing hydraulic buildings on rivers. 1) There is scarcity of discharge data series on the building plan site. In many cases, discharge data series are obtained from observations during the course of the study, usually not more than 1 year. These data are only sufficient for the need for calibration of the method parameters applied to extend the discharge data series. 2) Practical methods for extending the discharge data series that are often used in Indonesia are empirical methods, including FJ Mock method, NRECA, and others. The application of this method is very limited by the dimensions of space and time, and its use is limited to analyzing monthly period discharge data. This method is not flexible in anticipating hydrological changes in a watershed, so the results obtained are often inaccurate. This study provides an alternative solution to solve this problem by utilizing the advantages of the Tank model which can be set for daily, weekly, and semi-monthly discharge analysis, even for hourly periods as needed. To be effective and applicable, the Tank model equation system is combined with the Differential evolution (DE) algorithm to calibrate the parameters. The Tank model is considered more flexible than the empirical method because it involves more complex watershed parameters in its analysis.

Efforts to increase the effectiveness of the Tank model by involving optimization methods to find the optimal value of its parameters have been proposed by many previous researchers. The application of the Kalman Filter with a recursive algorithm to calibrate the tank model parameters was proven to be accurate

in predicting flood events in the Wi-Chun Bazin 472.53 km², which is located in the middle of the Nakdong River basin in Korea (Lee and Singh, 1999). The application of the Powell Tank-Multi start model to the Gao-Ping Creek watershed in Taiwan (792 km²), involving the daily period data set for the 1 January 1991 to 31 December 1992 period, also showed accurate results (Chen et al., 2005). The Tank-Marquard Algorithm model successfully presented the daily period rain-runoff relationship at the Terauchi Watershed (5055 ha) in Fukuoka Japan and the Ciriung Watershed (120 ha) in the Banten Province of Indonesia (Setiawan, 2014).

Metaheuristics is an advanced method based on heuristics to solve optimization problems efficiently (Talbi, 2009). The development of a revolutionary and reliable metaheuristic method in solving large and complex systems of equations makes this method attractive to be applied to solve optimization problems to find the optimal value of the parameters of the hydrological conceptual model. The combination of metaheuristic methods with the Tank model for the transformation of rain data series into runoff has been proposed by many researchers around the world. Tank model-Particle Swam Optimization (PSO) algorithm has been successfully applied to the Shigenobu Watershed in Japan (Santos et al., 2011). Tank Model-Genetic Algorithm (GA) (Ngoc et al., 2013), Shuffled Complex Evolution (SCE), and PSO (Kuok et al., 2011) have also managed to show a very good performance. The combination of the Tank model with the SCE, GA, PSO, Artificial Immune System (AIS), and DE algorithm has also been successfully applied to the Yellow River watershed in China and the Reynold Creek Boise ID watershed, Mahantango Creek University Park watershed, Little River Tifton watershed in the United States of America. The five optimization methods work well, but the GA and PSO algorithms show better results in terms of accuracy and speed of convergence (Zhang et al., 2012). The combination of the Tank model with the PSO Algorithm for flood discharge analysis in urban areas in Taiwan has shown very good performance (Hsu, 2015). The combination of the Tank model with the SCE Algorithm has also been successfully applied

to flood simulations in terraced rice fields in Taiwan. The Tank-SCE model represents two major flood events caused by Plum Rain on 9–16 May 2005 and Typhoon Matsa on 1–6 August 2005 (Chen et al., 2014). The aims of this research are as follows: 1) to build a Tank-DE model for the transformation of rain data series into daily period discharge; the model application uses the MATLAB M-FILE program code compiled by the author himself; 2) to examine the effect of the length of the training data series involved in the calibration process on the quality of the discharge data series as predicted by the model; and 3) to examine the effect of the length of the training data series on the sensitivity level of the tank model parameters. Research related to items 2) and 3) was not found in the previous references. The Tank-DE model was developed from a combination of the Tank model simulation equation system and the Tank model parameter optimization method based on the DE Algorithm. The model testing uses a daily period data series, and as a case study is the Lesti Watershed in Malang Regency, East Java, Indonesia. The sensitivity level analysis is intended to examine the effect of the length of the training data series on the quality of the discharge data series from the model predictions, as well as to determine the level of consistency of the optimal values of the model parameters. The analysis was carried

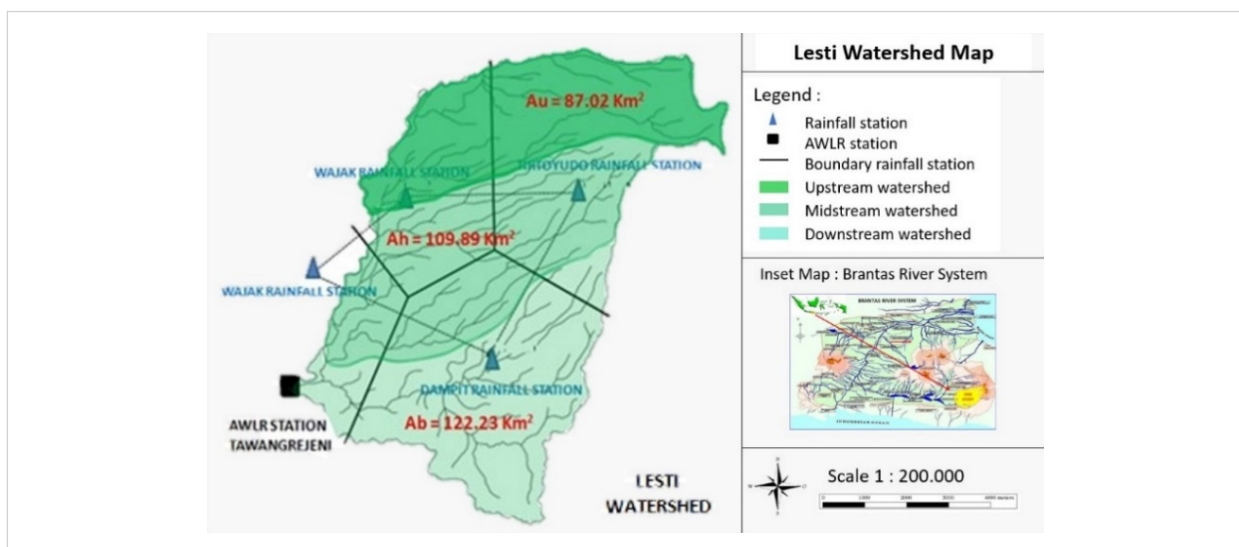
out through 5 scenarios with various input lengths of training data series, respectively, 1 year, 2 years, 4 years, 6 years, and 8 years. The results of the research are expected to increase the effectiveness of the application of the Tank model and become an alternative solution in solving the problem of limited discharge data series, which is often a classic problem in water resource development activities in developing countries, including Indonesia.

Materials and Method

Case study

The case study in this study is the Lesti watershed at the control point of the Tawangrejeni automatic water level record (AWLR) station as shown in Fig. 1. The Lesti watershed is included in the Upper Brantas watershed, with an area of 314.19 km². According to its physical characteristics, the Lesti watershed is divided into 3 zones, namely, up-land (Au = 87.02 km²), hill-slope (Ah = 109.89 km²), and bottom-slope (Ab = 122.23 km²). Lesti watershed is geographically located at 8°2'50"–8°12'10" South latitude and 112°42'58"–112°56'21" East longitude, and administratively is located in Malang Regency, East Java Province, Indonesia.

Fig. 1. Location of the case study (after Suliarto et al., 2020)



There are 3 types of hydro climatological data as input in the application of the Tank model, namely, evapotranspiration, rainfall, and discharge. The length of the hydro climatological data series involved in this study was 14 years, recorded from 1 January 1 2007 to 31 December 2020. Potential evapotranspiration data were obtained from the analysis using the Modified Penman Method. Climatic variables as input data in the Modified Penmann Method include air temperature, wind speed, humidity, and sunlight. The climate data series is the monthly average data from the recording results of the Karangates climatology station. The results of the analysis show that the average potential evapotranspiration of the Lesti watershed is 1056.25 mm per year, the minimum is 828.94 mm per year and the maximum is 1160.87 mm per year. Rainfall data involved is daily period data recorded at Dampit, Tirtoyudo, Wajak, and Turen rainfall stations. Regional rainfall was calculated with the Thiessen polygon method. The weighting factors for all four rainfall stations, respectively, are 0.38, 0.09, 0.19, and 0.34. The calculation results show the average rainfall of the Lesti watershed is 2417.31 mm per year, with a minimum of 1860.35 mm per year and a maximum of 3231.00 mm per year. The discharge data series at AWLR Tawangrejeni shows a mean value of 33.27

m^3/s , a minimum of $7.06 \text{ m}^3/\text{s}$, a maximum of $56.81 \text{ m}^3/\text{s}$, and a mean standard deviation of $10.03 \text{ m}^3/\text{s}$. The fluctuation of potential evapotranspiration data, rainfall, and daily period discharge are graphically shown in Fig. 2.

The hydro climatological data series is further divided into two groups, namely training data sets and testing data sets. The data series for the period 1 January 2007 to 31 December 2014 was used as a training data set for the model calibration process, and the hydro climatology data series from 1 January 2015 to 31 December 2020 as a testing data set for model validation. The comparison of the statistical parameters of the training data set and the testing data set is shown in Table 1. From the results of the two-way statistical test of the two data groups using the mean test (t-test) and variance test (F-test), it is concluded that the training data set and the testing data are homogeneous.

Tank model simulation

The number of tanks involved in the simulation process is highly dependent on the physical characteristics of the watershed under study. The higher the level of physical heterogeneity of a watershed, the variations in the use of tanks become more complex. However, of the many types of the Tank model, the

Fig. 2. Hydroclimatological data from 1 January 2007 to 31 December 2020 in the Lesti watershed

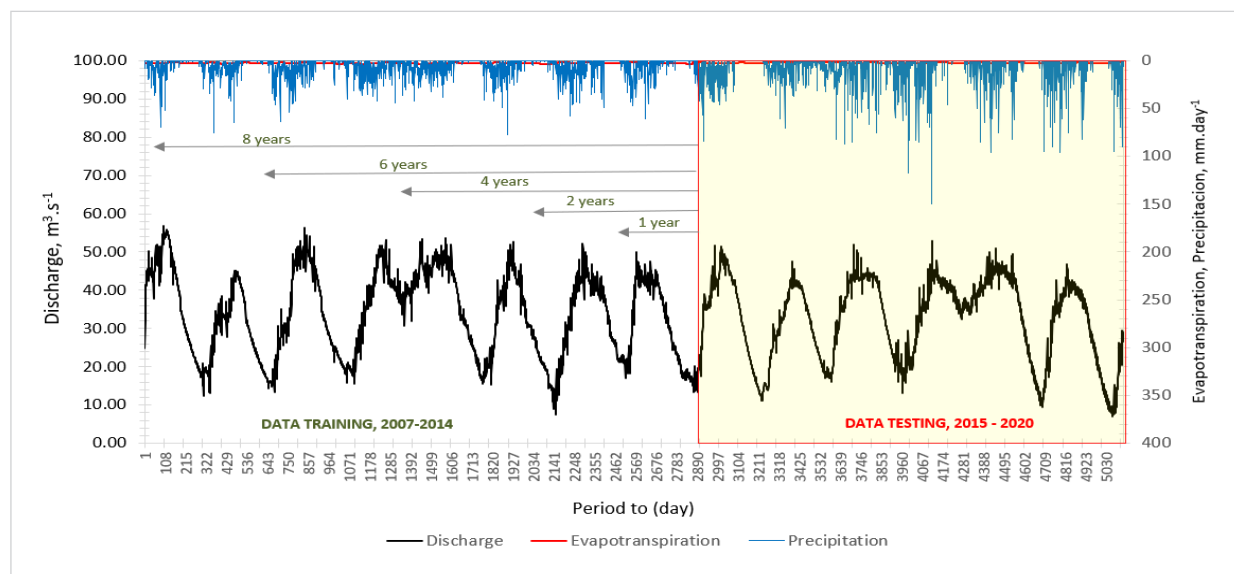


Table 1. Statistical characteristics of training and testing data sets

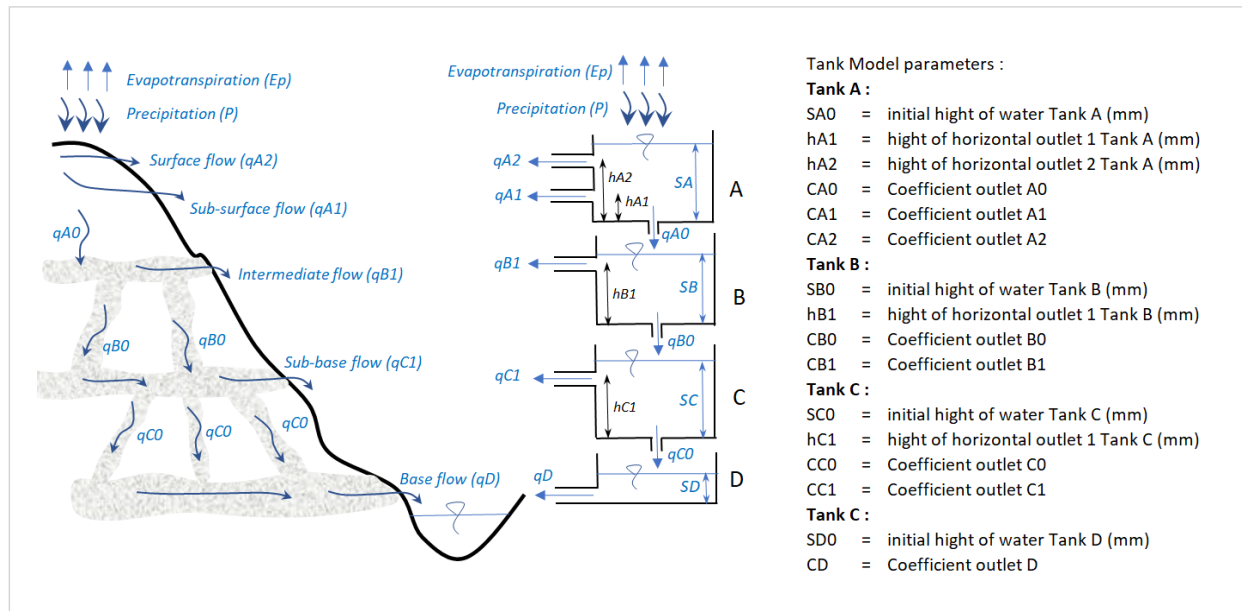
Data	Statistical parameters	Unit	Training data	Testing data
Period		day	1 January 2007 – 31 December 2014	1 January 2015 – 31 December 2020
Number of point data			2925	2194
Evapotranspiration	Mean	mm/year	897.78	856.78
Precipitation	Mean	mm/year	2235.73	2659.41
Discharge	Mean	m ³ /s	33.61	32.59
	Minimum	m ³ /s	7.55	7.06
	Maximum	m ³ /s	56.81	52.83
	Deviation standard	m ³ /s	10.99	10.61

application of a series tank arrangement consisting of four tanks (the Standard Tank model) is often the choice because of its simplicity and relevance to the physical characteristics of the watershed in general. The simulation scheme for the standard tank model is shown in Fig. 3. In the standard tank model, a watershed is represented by tanks arranged vertically with the assumption that each tank can represent a homogeneous sub-soil. The first tank to the fourth tank, respectively, will contribute to the occurrence of surface flow and sub-surface flow, intermediate

flow, sub-base flow, and base flow. The standard tank model has 16 parameters whose values must be relevant to the hydrological characteristics of the watershed under study.

In the Tank model simulation, water can fill the reservoir below it and can leave it if evapotranspiration is more dominant. As input variables in the application of the Tank model are regional rainfall (P), evapotranspiration (Et), and the value of the relevant parameters. The output variable is total discharge (Q), which is a superposition of surface flow (qA2), sub-surface

Fig. 3. Schematic of a standard Tank model simulation



flow ($qA1$), sub-base flow ($qC1$), and base flow (qD). Another outcome that can be explored is the fluctuation of the water level in each tank (SA , SB , SC , SD). This variable can be analogized as fluctuations in the groundwater level in each soil layer zone. The performance of the Tank model is largely determined by the accuracy in determining the value of each of its parameters. In this regard, the parameter calibration process becomes a very important part. The process of calibrating a large number of parameters simultaneously is certainly not effective if it is carried out by "trial and error"; therefore, the application of a reliable parameter optimization method is important to improve its performance.

In this study, the application of the Tank model uses the following assumptions: 1) each layer of soil in the watershed is considered to have uniform characteristics so that it can be represented by a tank, 2) the value of model parameters is considered constant during the analysis period, 3) the river discharge in one day analyzed from the input data of daily evapotranspiration and daily rainfall is a constant quantity, and 4) river discharge is a function of rainfall, evapotranspiration, and physical characteristics of the watershed. Other factors such as interception, snow, and others are not taken into account. So, the model developed is only relevant for watersheds in the tropics, including Indonesia.

The volume balance equation for each tank is expressed as:

$$\text{Tank A} : \frac{dSA}{dt} = P - Et - qA1 - qA2 - qA0 \quad (1)$$

$$\text{Tank B} : \frac{dSB}{dt} = qA0 - qB0 - qB1 \quad (2)$$

$$\text{Tank C} : \frac{dSC}{dt} = qB0 - qC1 - qC0 \quad (3)$$

$$\text{Tank D} : \frac{dSD}{dt} = qC0 - qD \quad (4)$$

The value of runoff and infiltration through the tank outlet is directly proportional to the water level at the outlet, expressed as:

$$\text{Tank A} : qA0 = CA0 * SA ; qA1 = CA1 * (SA - hA1) ; qA2 = CA2 * (SA - hA2) \quad (5)$$

$$\text{Tank B} : qB0 = CB0 * SB ; qB1 = CB1 * (SB - hB1) \quad (6)$$

$$\text{Tank C} : qC0 = CC0 * SC ; qC1 = CC1 * (SC - hC1) \quad (7)$$

$$\text{Tank D} : qD = CD * SD \quad (8)$$

The total runoff period t in mm/day is expressed as:

$$q(t) = qA1(t) + qA2(t) + qB1(t) + qC1(t) + qD(t) \quad (9)$$

In units of m^3/s , it is stated:

$$Q(t) = q(t) . A/86.4 \quad (10)$$

Although the tank components represent a simple linear system of equations, the series of combinations of tank components results in a completely non-linear integral operation.

Model calibration

The calibration process is directed at finding the optimal value of the tank model parameters. The calibration process can be analogized as an optimization process that has the objective function of minimizing the deviation of the observed discharge curve (Q^{obs}) and the simulation results from the Tank model (Q^{sim}). In the heuristic method, the objective function is expressed as a fitness function. The definition of the fitness function in the case of optimization of hydrological model parameters has been proposed by many previous researchers, including the minimization of root mean square error or RMSE (Hsu, 2015; Wang et al., 2012; Sulianto et al., 2018; Sulianto et al., 2020), minimization of sum square error (SSE) (Darikandeh et al., 2014; Paik et al., 2005), maximizing Nash-Sutcliffe model efficiency or NSE (Xuesong Zhang et al., 2008; Bao et al., 2010; Tolson and Shoemaker, 2007), minimization of mean square error or MSE (Ngoc et al., 2013), minimization of relative error or RE (Santos et al., 2011; Kuok et al., 2011). The fitness function in this article is to minimize the RMSE value, expressed as:

$$F = \min[RMSE] = \min \left[\sqrt{\frac{\sum_{t=1}^n (Q_t^{sim} - Q_t^{obs})^2}{n}} \right] \quad (11)$$

Where: F – fitness value;

Q_t^{sim} – discharge from model simulation in period t ;

Q_t^{obs} – discharge observation period t ;

n – number of data points.

As a constraints function in the optimization process are:

– The Tank model simulation equation system is expressed as

$$Q(t) = F(P(t), Ep(t), A, 16 \text{ parameters Tank model});$$

– Lower boundary (lb) and upper boundary (ub_j) of initial value height of water in tanks ($SA0, SB0, SC0, SD0$);

– Lower boundary (lb) and upper boundary (ub) of horizontal outlet height in tanks ($hA1, hA2, hB1, hC1$);

– Lower boundary (lb) and upper boundary (ub) of coefficient outlet value in tanks ($CA0, CA1, CA2, CB0, CB1, CC0, CC1, CD$).

Model validation

The model validation process uses a testing data set which is not involved in the model calibration process. The performance of the model is measured using deviation indicators, namely, RMSE, mean absolute error (MAE), standard error X (X), square standard error X (X^2), relative error (RE), square relative error (RR), and Nash-Sutcliffe model efficiency (NSE) are calculated using the following equation.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_t^{obs} - Q_t^{sim}| \quad (12)$$

$$X = \frac{1}{n} \sum_{i=1}^n \frac{|Q_t^{obs} - Q_t^{sim}|}{\sqrt{Q_t^{obs}}} \quad (13)$$

$$X^2 = \frac{1}{n} \sum_{i=1}^n \frac{|Q_t^{obs} - Q_t^{sim}|^2}{Q_t^{obs}} \quad (14)$$

$$RE = \frac{1}{n} \sum_{i=1}^n \frac{|Q_t^{obs} - Q_t^{sim}|}{Q_t^{obs}} \quad (15)$$

$$RR = \frac{1}{n} \sum_{i=1}^n \frac{|Q_t^{obs} - Q_t^{sim}|^2}{Q_t^{obs^2}} \quad (16)$$

$$NSE = 1 - \frac{\sum_{t=1}^n (Q_t^{sim} - Q_t^{obs})^2}{\sum_{t=1}^n (Q_t^{obs} - Q^{mean})^2} \quad (17)$$

Where: Q_t^{sim} – discharge from simulation model period t , m^3/s ;

Q_t^{obs} – observation discharge period t , m^3/s ;

Q_{t-1}^{obs} – observation discharge period $t-1$, m^3/s ;

Q^{mean} – mean of observation discharge, m^3/s ;

n – number of data points.

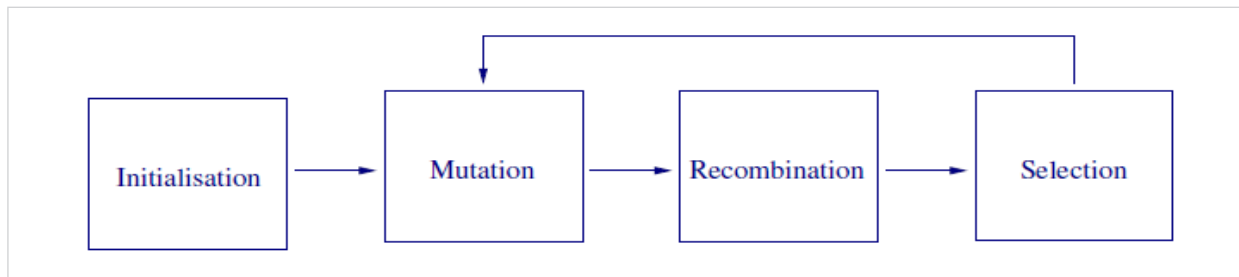
Differential evolution (DE) algorithm

The DE algorithm was developed by Reiner Storn and Kenneth Price in 1996. In the field of hydrological modeling, the DE algorithm has been successfully applied to the optimization of SWAT model parameters (Xuesong Zhang et al., 2008), and the optimization of HBV and GR4J model parameters (Piotrowski et al., 2017). It was also successfully applied in the case of multi-objective optimization of in-situ bioremediation of groundwater (Kumar et al., 2015), optimization of DISPRIN model parameters (Sulianto et al., 2018), and optimization of the Modified DISPRIN model (Sulianto et al., 2020). The analysis in the DE Algorithm contains 4 (four) components, namely, 1) initialization, 2) mutation, 3) crossover, and 4) selection. The relationship between the four components is shown in Fig. 4.

1 Initialization

In the case of the optimization of the Tank model parameters, the optimized variable vectors are 16 parameters of the Tank model as described in Fig. 3. Before initializing the searched vector variable, it is necessary to determine the lower limit (lb_j) and the upper limit (ub_j) of all optimized variables. lb_j and ub_j will be used as the first step to generate the values of the variables being searched for. For generating the initial value of the 0th generation variable, the j variable and the i vector can be represented by the following notation.

Fig. 4. Differential evolution component relationships (after Sulianto et al., 2020)



$$x_{j,i,0} = lb_j + rand_j(, 1)(ub_j - lb_j) \quad (18)$$

Random numbers are generated by the rand function, where the resulting number lies between (0,1). Index j indicates the variable to j .

2 Mutation

DE will mutate and combine the initial population to produce a population of size N experimental vectors. In DE, the mutation is done by adding the difference of two vectors to the third vector by:

$$v_{i,g} = x_{r_0,g} + F(x_{r_1,g} - x_{r_2,g}) \quad (19)$$

It can be seen that the difference between two randomly selected vectors needs to be scaled before being added to the third vector, $x_{r_0,g}$. The scale factor $F \in (0,1)$ has a positive real value to control the population growth rate. The base vector index r_0 can be determined in various ways, generally using a different random method than the index for the target vector, i . In addition to being different from each other and different from the indices for the base vector and target vector, the index of the difference vectors r_1 and r_2 is also selected once per mutant.

3 Crossover

At this stage, DE crosses each vector $x_{i,g}$, with a mutant vector $v_{i,g}$, to form the resulting vector, $u_{i,g}$ with the formula.

$$u_{i,g} = u_{j,i,g} = \begin{cases} v_{j,i,g} & \rightarrow \text{jika } (rand(0,1) \leq Cr \text{ atau } j = j_{rand}) \\ x_{j,i,g} & \rightarrow \text{jika } (rand(0,1) > Cr \text{ atau } j \neq j_{rand}) \end{cases} \quad (20)$$

4 Selection

If the trial vector $u_{i,g}$ has a value of the objective function that is smaller than the objective function of the target vector $x_{i,g}$, then $u_{i,g}$ will replace the position of $x_{i,g}$ in the population in the next generation. Otherwise, the target vector will remain in its position in the population.

5 Stopping criteria

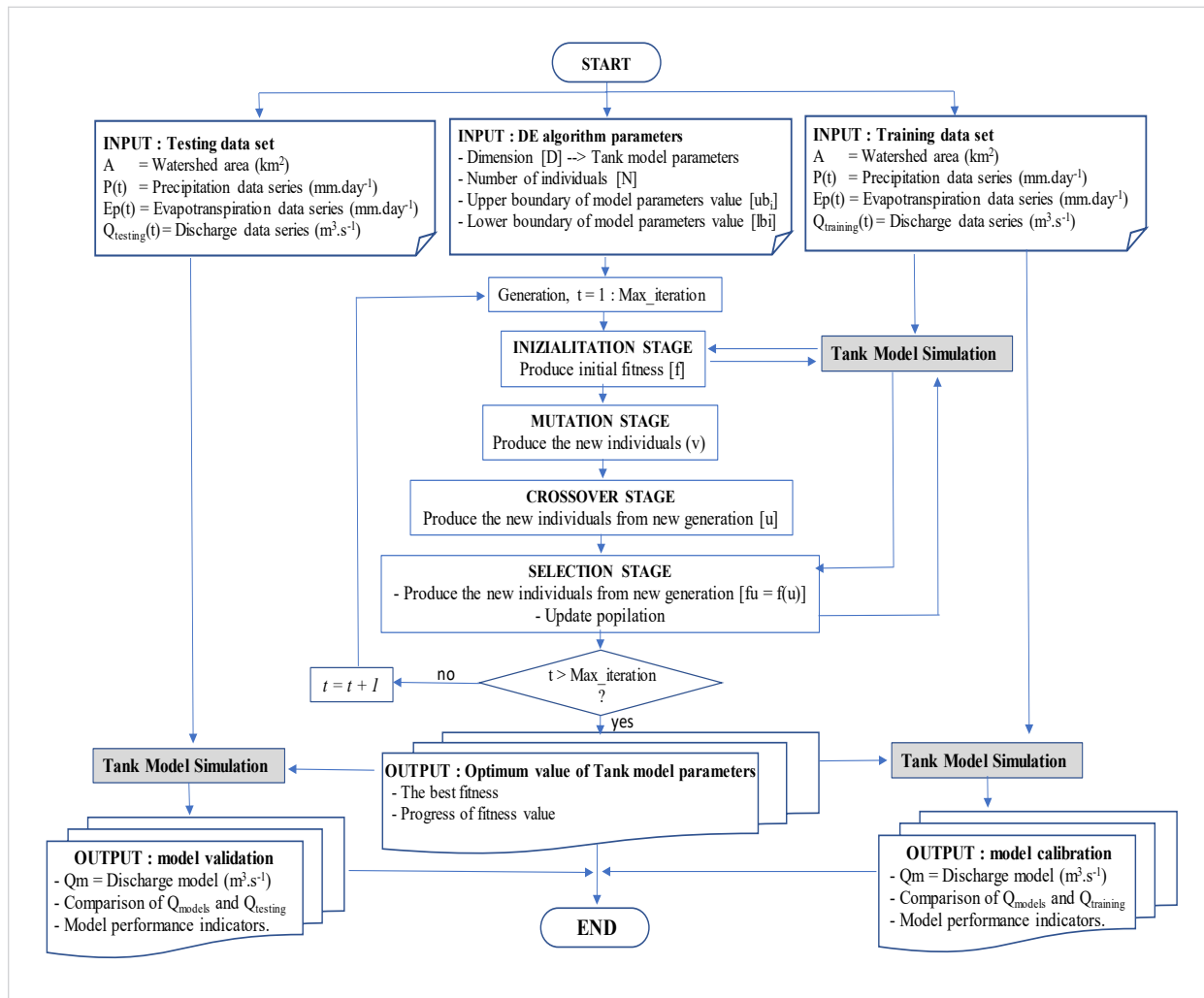
The iteration process will stop at the specified stopping criteria, namely the maximum number of generations given, or if the value of the optimized variable is constant from generation to generation.

In convergent conditions, the optimum value of the Tank model parameter has been obtained. Furthermore, by utilizing the simulation equation system according to equation (1) to equation (10), the output variables of the Tank model can be presented in the form of numerical or graphic data.

Tank-DE model algorithm

The model from the combination of the equation system of the Tank model simulation and the DE algorithm is hereinafter called the Tank-DE model. The Tank-DE model algorithm developed in this study is schematically shown in Fig. 5. The application of the Tank-DE model uses the MATLAB M-FILE program code which is composed of the main program and 9 sub-programs. The DE algorithm equation system is the main program, and 9 sub-programs include: 1) fitness function, 2) calibration process, 3) validation process, 4) rainfall data training, 5) evapotranspiration data training, 6) discharge data training, 7) rainfall data testing, 8) evapotranspiration data testing, and 9) discharge data testing.

Fig. 5. Tank-DE model algorithm



Result and Discussion

The criteria for the application of the Tank model related to the length limit of the training data series involved in the calibration process, as well as the minimum and maximum values of the parameters involved in the calibration process were not found in previous studies. All studies involved the length of the data series and the range of values of the various parameters. The analysis results obtained generally produce fairly good performance indicators, although the optimal values of the parameters also vary. This is caused by the non-linearity and the many

parameters of the Tank model equation system. Tank model calibration using the Marquard algorithm involved a 10-year training data set (years 1986–1995) at the Terauchi Watershed in Fukuoda in Japan and training data for 2 years (years 2002–2003) at the Cirung Watershed in Indonesia, both of which showed good performance satisfying. Analysis of the Fukuoda watershed has shown better performance in terms of X , X^2 , RE, and RR, although the difference is not significant, and the optimal value of the parameters shows a significant difference (Setiawan, 2014). The calibration of the Tank model using GA involved 1-year training data series (in 1995 and 2000) in Dau Tieng

River, Vietnam, indicating that the calibration with the 1995 training data set was slightly better, with values of NSE = 0.80, MSE = 0.2, RMSE = 1.39 and MAE = 1.70. The optimum value of the parameters also resulted in a significant difference between the two data analyses (Ngoc et al., 2013). The calibration of the Tank model in the Niulan sub-basin involved a 6-year training data series (2000–2006), and the Xining River basin for 8 years (2006–2013) resulted in an equivalent NSE value of 0.75 (Tayyab et al., 2015). These results indicate that the Tank model can represent the relationship between rainfall and discharge data series in the daily period quite well. The optimum value of the parameters of the resulting Tank model is very specific, the value is very dependent on the dimensions of space and time.

This study examines the effect of the length of the training data series on the performance of the Tank model and the level of consistency of the optimal values of the resulting parameters. In this regard, the data analysis uses 5 scenarios. Each scenario uses the input of a testing data set for 6 years (1 January 2015 – 31 December 2020), and a training data set for 1 year (1 January 2014 – 31 December 2014), 2 years (1 January 2013 – 31 December 2014), 4 years (1 January 2011 – 31 December 2014), 6 years (1 January 2019 – 31 December 2014), and 8 years (1 January 2007 – 31 December 2014). The characteristics of the training data set for each scenario are detailed in *Table 2*, and schematically shown in *Fig. 3*.

Data analysis for each scenario uses the DE algorithm

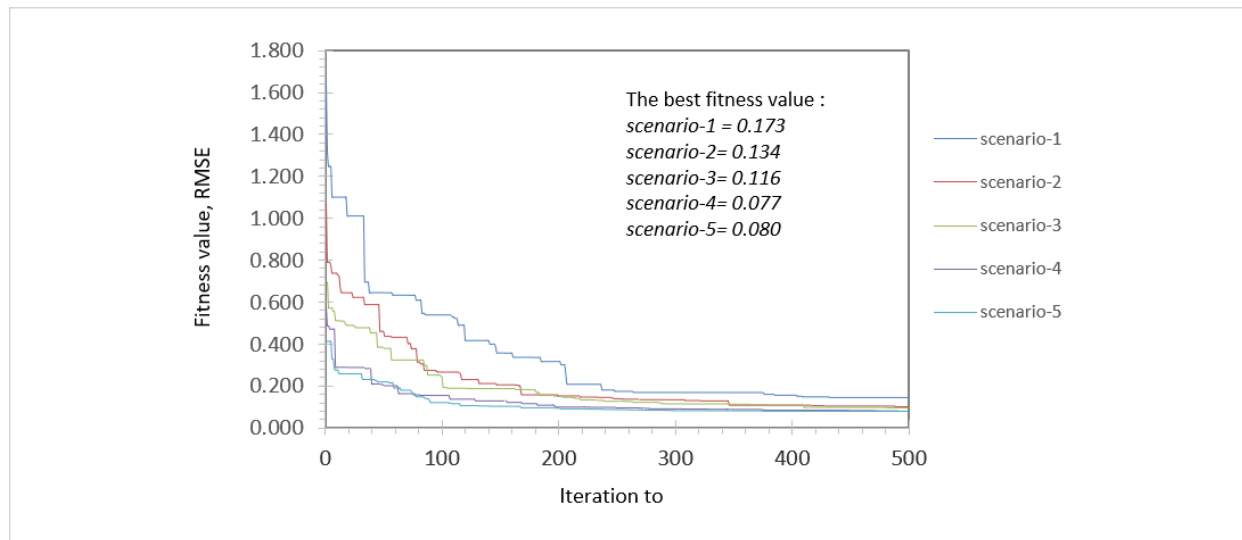
parameter set input with the same values, namely, the number of generations (N) = 500, the lower boundary of the tank hole coefficient value (lb_j) = 0.0001, the upper boundary of the tank hole coefficient value (ub_j) = 1.00, the lower boundary of the hole height and the initial water level in the tank (lb_j) = 0.0001, and the upper boundary of the hole height and initial water level in the tank (ub_j) = 250. The search space represented by the values of lb_j and ub_j is determined based on references from several previous papers and the optimum conditions obtained from several experiments that were carried out in this case study. The results of running programs from the 5 scenarios are briefly presented in *Figs. 6–11*, *Table 3*, and *Table 4*.

Fig. 6 shows the progress of achieving the best fitness value from generation to generation for all scenarios. In general, it shows that the developed model is quite consistent and has succeeded in finding the optimal conditions according to the formulated objective function. In optimal conditions, the best fitness in the 500th iteration shows a difference in the values of each scenario. It appears that the longer the training data set, the smaller the best fitness value tends to be, although the difference is not very significant. This is due to the difference in the statistical characteristics of the training data sets, which is not too conspicuous. Scenario-1 as a representation of a short training data series produces the largest RMSE value, which is 0.173, and scenario-5 as a representation of a long training data series produces the smallest RMSE value, which is 0.080.

Table 2. Characteristic data as input model

Scenario	Training data			Testing data		
	Period	Length	Data points	Period	Length	Data points
1	1 January 2014 – 31 December 2014	1 year	365	1 January 2015 – 31 December 2020	6 years	2194
2	1 January 2013 – 31 December 2014	2 years	735			
3	1 January 2011 – 31 December 2014	4 years	1465			
4	1 January 2009 – 31 December 2014	6 years	2195			
5	1 January 2007 – 31 December 2014	8 years	2925			

Fig. 6. The progress of the best fitness value in the iteration process



Under optimal conditions, the Tank model parameters values are generated as shown in Figs. 7–8. Fig. 7 shows the comparison of the parameter values for the height of the tank hole and the initial water level in the tank in each scenario. Parameters HA2 and SA0 showed significant differences in values, HA1 parameters were significantly different, and parameters HB1, SB0, SC0, SD0 were slightly different and even tended to have the same value for each scenario. The optimum value of the parameters HA2 and SA0 was more different than the other parameters, due to the non-linearity factor of the Tank model equation system. The variable values of the HA2 and SA0 parameters indicate that these parameters are not sensitive enough, meaning that any value does not have a major effect on the resulting RMSE value. Of course, this only applies to this case study. The value of the Tank model parameters is uncertain and the value will always be different for each case, depending on the physical properties of the watershed and the characteristics of the relationship between the rain data series and the discharge data series involved in the calibration process (Lee and Singh, 1999).

Fig. 8 shows that the outlet tank coefficient values in general are not significantly different, except for the CA0 and CB0 parameters. This indicates that the

smaller the difference in values, the higher the sensitivity of the parameter to the model's performance, and vice versa. Thus, it can be identified that the altitude parameters HB1, HC1, SB0, SC0, SD0 are very sensitive, HA1 is quite sensitive, and HA2, SA0 are not sensitive. The outlet tank coefficient parameters CA0, CB0, CB1 are not sensitive, and CA1, CA2, CC0, CC1, CD are very sensitive. Identification of the sensitivity level of this parameter will be useful in applying the Tank-DE model, especially in determining the feasibility value limit of the Tank model parameters.

Furthermore, by using the input set of optimum values of the Tank model parameters and the hydro climatology training data set, the output of the discharge model fluctuation is obtained. Fig. 9 shows the comparison of the discharge model curve and the training discharge data curve. Fig. 10 shows the comparison of the discharge model curve with the testing discharge data curve. Fig. 9 shows almost the same trend between the training discharge curve and the discharge model curve from all scenarios. This shows that the model being built is quite consistent. The difference in the model discharge value in each scenario is only influenced by the number of data points and the consistency of the relationship between climate data and training discharge data. Of course, if all data points

Fig. 7. Comparison of the optimum value of the the initial water level in the tanks and height of outlet tank

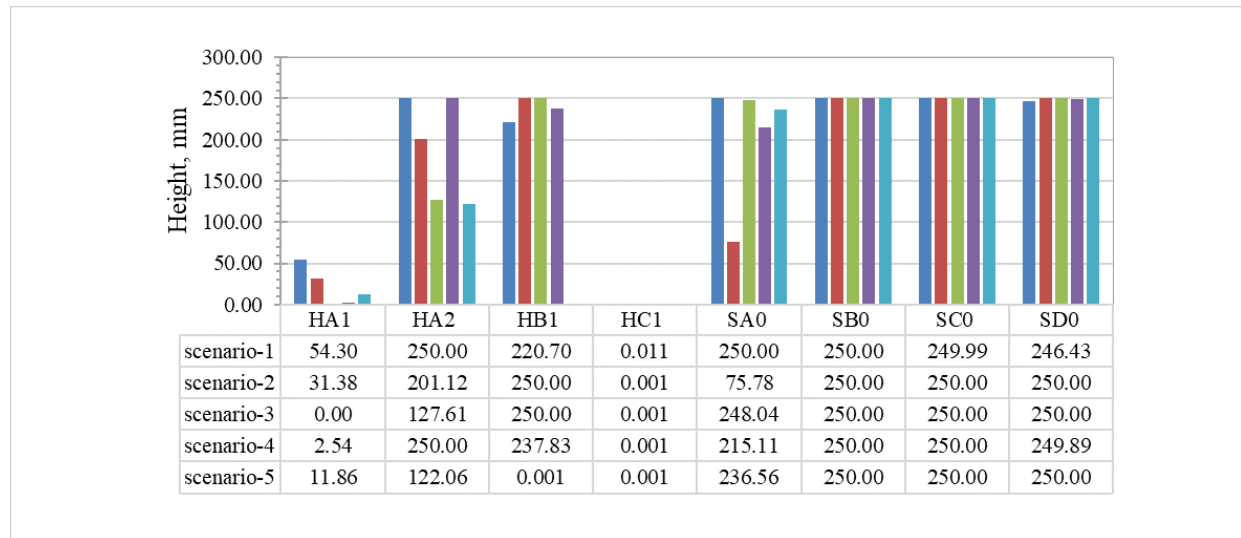
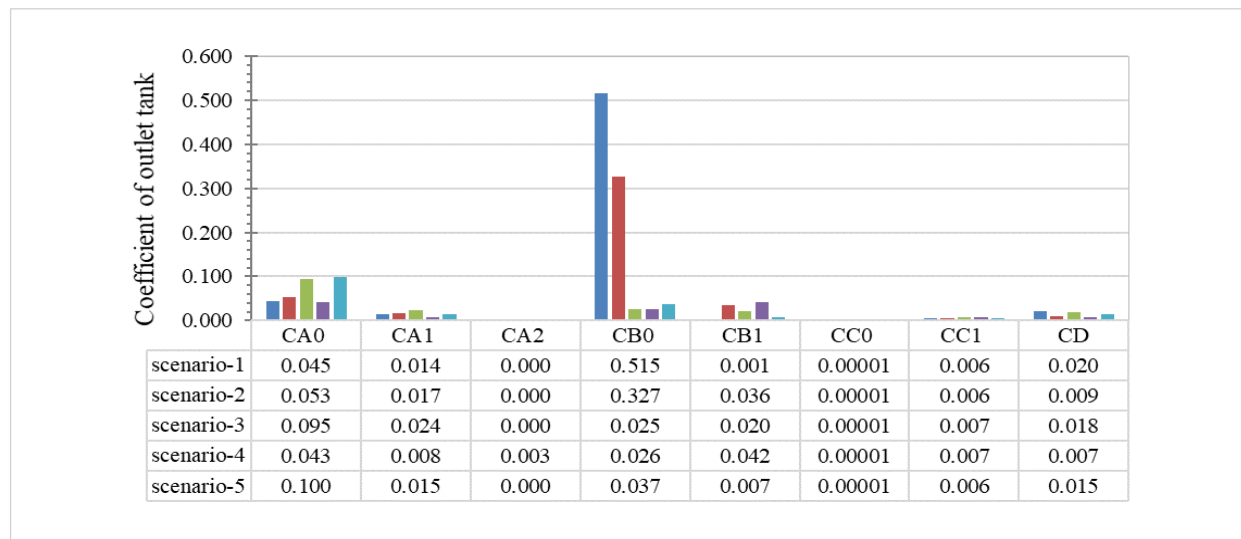


Fig. 8. Comparison of the optimum value of the outlet tank coefficient in the tanks



had a consistent correlation, the discharge curves of the model would coincide with each other. A comparison of model performance at the calibration stage for each scenario is shown in *Table 3*. The value of the model performance indicators shows various phenomena. RMSE tends to decrease with the length of the training data series involved in the calibration process. In scenario-4 and scenario-5, the RMSE value

appears smaller, as well as the NSE indicator. However, the indicators X , X^2 , RE, and RR do not show a certain trend. Their values tend to fluctuate with not too large a deviation. Based on the NSE values which are all greater than 0.8, it can be concluded that qualitatively the calibration process in all scenarios showed excellent performance, although quantitatively it appears that scenario-5 gives the most accurate results.

Table 3. Comparison of model performance indicator values at the calibration stage for all scenarios

Indicators	Symbol	Callibration Stage				
		scenario-1	scenario-2	scenario-3	scenario-4	scenario-5
Root mean square error	RMSE	0.173	0.134	0.116	0.077	0.083
Mean absolute error	MAE	2.495	2.790	3.662	2.874	3.594
Standard X	X	0.497	0.506	0.690	0.501	0.631
Square standard X	X ²	0.492	0.445	0.756	0.386	0.605
Relative error	RE	0.104	0.095	0.135	0.090	0.115
Square relative error	RR	0.026	0.017	0.033	0.013	0.022
Nash-Sutcliffe efficiency	NSE	0.908	0.862	0.830	0.888	0.834

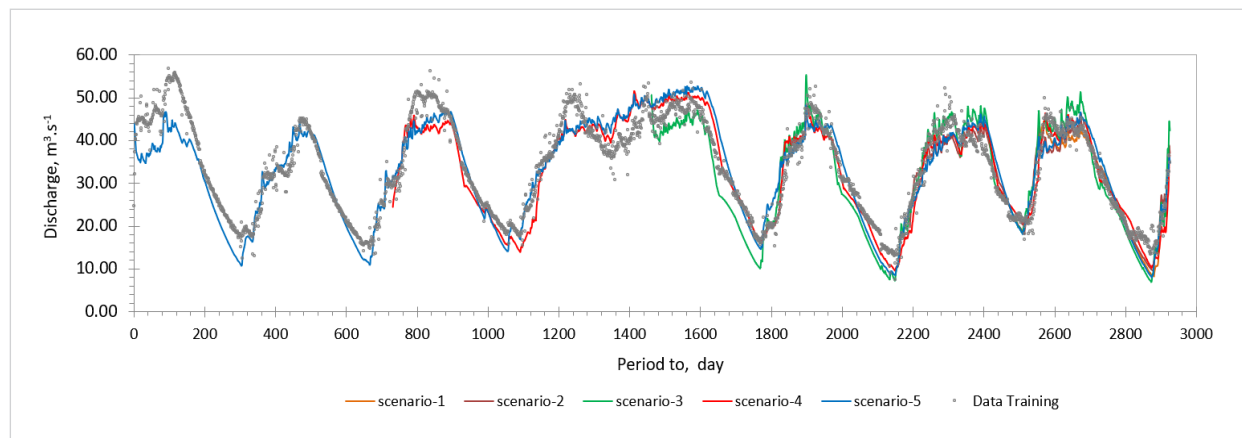
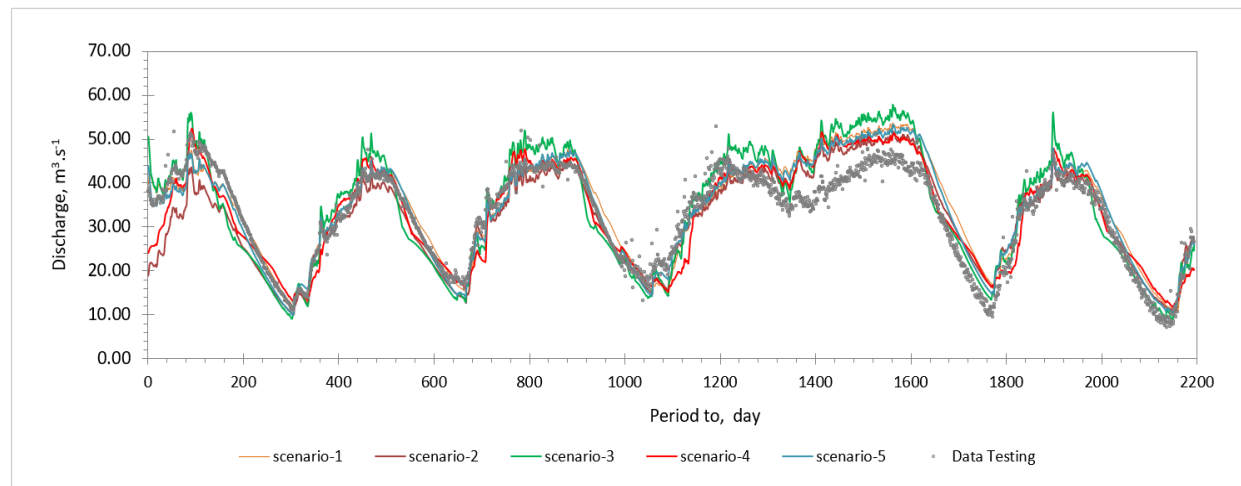
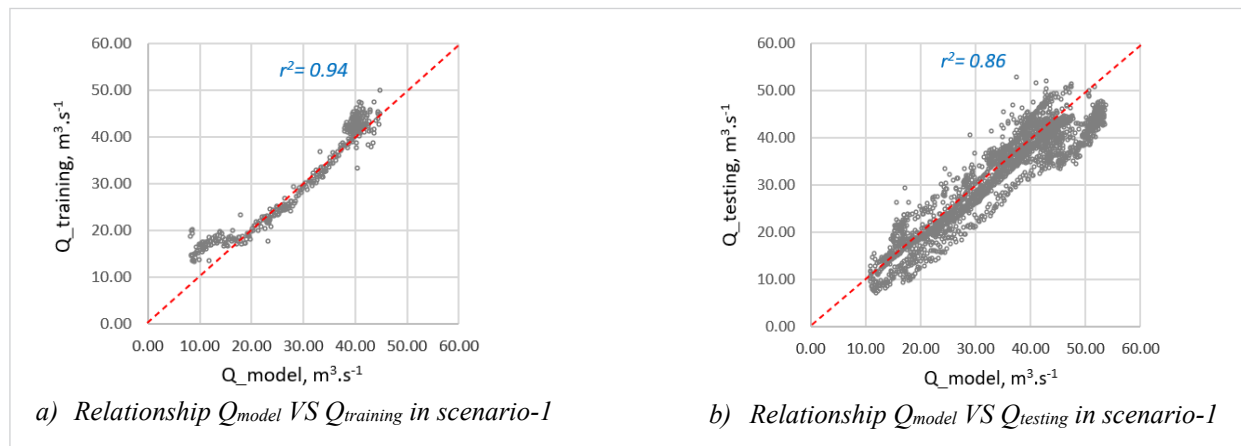
Fig. 9. Comparison of the model discharge fluctuation curve and the training data curve of all scenarios

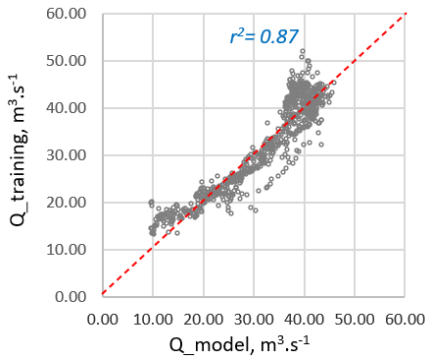
Fig. 10 also indicates that qualitatively the discharge model from all scenarios shows the same fluctuation trend. The discharge model curves from all scenarios with the discharge testing curves generally show a pretty good trend, except in the 1250th period to the 1700th period. In this period, it appears that the model output tends to be overestimated, although the deviation is not too large. This difference may be caused by an inconsistent correlation between rainfall data and discharge data. The NSE value generated in all scenarios is more than 0.75 indicating the validation process is also quite successful. Quantitatively, scenario-5 produces the best performance compared with other scenarios. This is indicated by the value of the RMSE, MAE, X, X², RE, and RR indicators, which tend to be the smallest, although the difference in values is not significant.

Fig. 11 presents the comparison of the values of the determination coefficient (r^2) from the relationship between the discharge data and the discharge model resulting from the calibration process and the validation process of all scenarios. In general, the calibration process shows better performance than the validation process, except for scenario-5. Scenario-1 shows the biggest difference in the value of r^2 in the calibration and validation processes. This condition indicates that the use of a minimum training data set (1 year) will have a risk of biasing the discharge predicted by the model. The use of a relatively long training data series (> 1 year) can reduce discharge prediction errors because the calibration process has provided training for models with various possible relationships between rain and discharge.

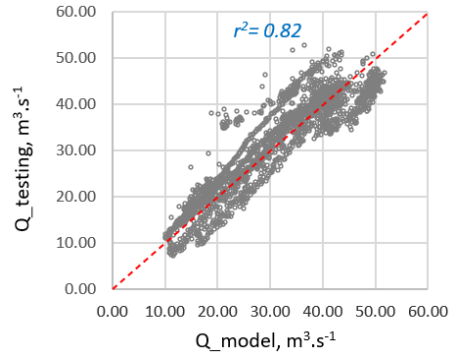
Table 4. Comparison of model performance indicator values at the validation stage for all scenarios

Indicators	Symbol	Validation Stage				
		scenario-1	scenario-2	scenario-3	scenario-4	scenario-5
Root mean square error	RMSE	0.099	0.099	0.112	0.094	0.087
Mean absolute error	MAE	3.611	3.562	4.260	3.373	3.185
Standard X	X	0.656	0.640	0.740	0.622	0.572
Square standard X	X ²	0.705	0.662	0.782	0.664	0.515
Relative error	RE	0.125	0.121	0.134	0.121	0.108
Square relative error	RR	0.030	0.026	0.025	0.030	0.020
Nash-Sutcliffe efficiency	NSE	0.810	0.810	0.754	0.829	0.853

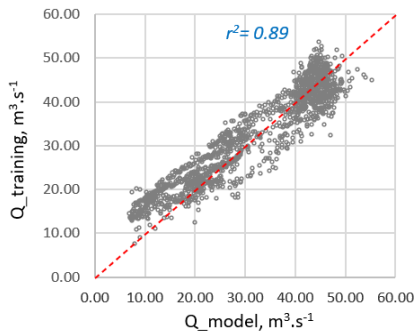
Fig. 10. Comparison of model discharge curves and testing data for all scenarios**Fig. 11.** Comparison of the determination coefficient of the relationship between Q_{model} and $Q_{observation}$ 



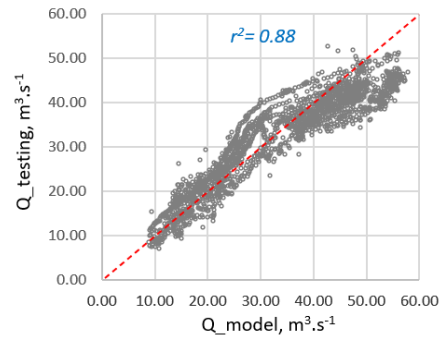
c) Relationship Q_{model} VS $Q_{training}$ in scenario-2



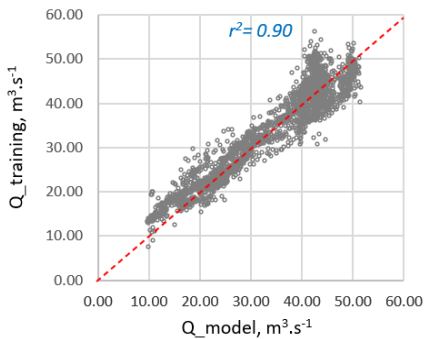
d) Relationship Q_{model} VS $Q_{testing}$ in scenario-2



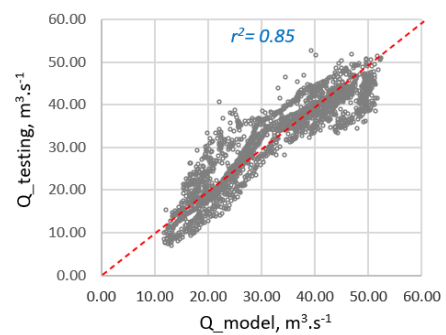
e) Relationship Q_{model} VS $Q_{training}$ in scenario-3



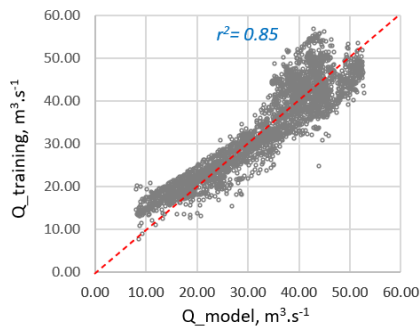
f) Relationship Q_{model} VS $Q_{testing}$ in scenario-3



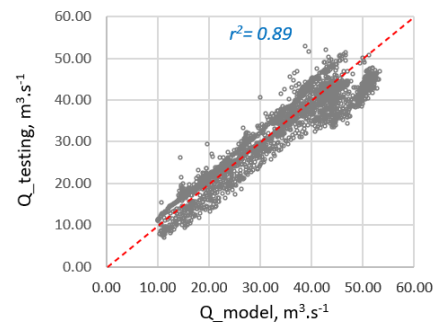
g) Relationship Q_{model} VS $Q_{training}$ in scenario-4



h) Relationship Q_{model} VS $Q_{testing}$ in scenario-4



i) Relationship Q_{model} VS $Q_{training}$ in scenario-5



j) Relationship Q_{model} VS $Q_{testing}$ in scenario-5

Conclusion

The Tank-DE model developed in this research can work very well in presenting the relationship of climate data series to a daily period discharge data series. At the calibration stage, the process of finding the best fitness value can run effectively. The results of the analysis of 5 scenarios indicate that the longer training data series involved in the calibration process tends to produce a smaller best fitness value, meaning that the longer training data series tends to be more accurate in terms of RMSE. In scenario-1, the value of r^2 at the validation stage decreased quite sharply, although r^2 from the calibration process showed the highest value. This indicates that the use of a short training data series (≤ 1 year) will have a risk of biasing the discharge predicted by the model. The use of relatively long training data series (> 1 year) can reduce discharge prediction errors because the calibration process has provided a wider training space.

The results of the analysis of the 5 scenarios show the equivalent RMSE values, but the optimum values of the resulting Tank model parameters are inconsistent. The optimum value of the parameters HA2 and SA0 is more different than the other parameters, due to the non-linearity factor of the Tank model equation system. The variable values of the HA2 and SA0 parameters indicate that these parameters are not sensitive enough, meaning that any value does not have a

major effect on the resulting RMSE value. The sensitivity level of the Tank model parameters is uncertain and the value will always be different for each case, depending on the characteristics of the relationship between the rain variable and the discharge variable involved in the calibration process.

The Tank Model is difficult to apply to solve practical problems, in conditions where training data sets are not available. Implementing the Tank-DE model requires a training data set. Although it showed a fairly good performance, the values of the resulting Tank model parameters are only specific and suitable for the solved case. For the application of the Tank model to be more practical and effective, further research is needed to examine the relationship between watershed physical properties, rainfall-discharge characteristics and the limit values of the Tank model parameters.

Acknowledgments

Researchers express appreciation and gratitude to DPPM University of Muhammadiyah Malang, which has provided financial support for the implementation of this research activity through the "PKID-UMM Scheme". Hopefully, the results of this research can make a positive contribution to the development of science, and become an alternative solution to the problem of the limited runoff data series that often becomes an obstacle in water resource management activities.

References

- Bao, H., Wang, L., Li, Z., Zhao, L., and Zhang, G. (2010). Hydrological daily rainfall-runoff simulation with BTOPMC model and comparison with Xin ' Anjiang model. *Water Science and Engineering*, 3(2), 121-131. <https://doi.org/10.3882/j.issn.1674-2370.2010.02.001>
- Chen, R., Pi, L., and Hsieh, C. (2005). Application of Parameter Optimization Method for CAalibrating Tank Model. *Journal of The American Water Resources Association*, 40(2), 389-402. <https://doi.org/10.1111/j.1752-1688.2005.tb03743.x>
- Chen, S., Chen, R., and Yang, T. (2014). Application of a tank model to assess the flood-control function of a terraced paddy field. *Hydrological Sciences Journal - Journal Des Sciences Hydrologiques*, 59(5), 1020-1031. <https://doi.org/10.1080/02626667.2013.822642>
- Darikandeh, D., Akbarpour, A., Bilondi, M. P., and Hashemi, S. R. (2014). Automatic Calibration for Estimation of the Parameters of Rainfall-Runoff Model. *Journal of River Engineering*, 2(8), 1-7.
- Hsu, P. (2015). Study on Flood Para-Tank Model Parameters with Particle Swarm Optimization. *Journal of Information Hiding and Multimedia Signal Processing*, 6(5), 911-923.
- Kumar, D., CH, S., Mathur, S., and Adamowski, J. (2015). Multi-objective optimization of in-situ bioremediation of groundwater using a hybrid metaheuristic technique based on differential evolution, genetic algorithms, and simulated annealing. *Journal of Water and Land Development*, 27(X-XII), 29-40. <https://doi.org/10.1515/jwld-2015-0022>

- Kuok, K. K., Harun, S., and Chan, C. P. (2011). Comparison of Particle Swarm Optimization and Shuffle Complex Evolution for Auto-Calibration of Hourly Tank Model's Parameters. *Int. J. Advance. Soft Comput. Appl.*, 3(3), 1-17.
- Lee, Y. H., and Singh, V. P. (1999). Tank Model using Kalman Filter. *Journal of Hydrologic Engineering*, 4(October), 344-349. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1999\)4:4\(344\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:4(344))
- Ngoc, T. A., Hiramatsu, K., and Harada, M. (2013). Optimizing Parameters for Two Conceptual Hydrological Models Using a Genetic Algorithm : A Case Study in the Dau Tieng River Watershed, Vietnam. *JARQ*, 47(1), 85-96. <https://doi.org/10.6090/jarq.47.85>
- Paik, K., Kim, J. H., Kim, H. S., and Lee, D. R. (2005). A conceptual rainfall-runoff model considering seasonal variation. *Hydrological Processes*, 3850(June), 3837-3850. <https://doi.org/10.1002/hyp.5984>
- Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., Osuch, M., Kundzewicz, Z. W., Piotrowski, A. P., Napiorkowski, M. J., and Napiorkowski, J. J. (2017). Are modern metaheuristics successful in calibrating simple conceptual rainfall - runoff models ? *Hydrological Sciences Journal*, 62(4), 606-625. <https://doi.org/10.1080/02626667.2016.1234712>
- Santos, C. A. G., Freire, P. K. M. M., Mishra, S. K., and Soares, A. (2011). Application of particle swarm optimization to the tank model. *Risk in Water Resources Management (Proceeding of Symposium H03 Held IUGG2011)*, 347, 114-120.
- Setiawan, B. I. (2014). Developing Procedures for Optimization of Tank Model's Parameters. *Agricultural Engineering International: The CIGR Journal of Scientific Research and Development.*, Manuscript(June), 1-14.
- Sulianto, S., Bisri, M., Limantara, L. M., and Sisinggih, D. (2018). Automatic calibration and sensitivity analysis of DISPRIN model parameters : A case study on Lesti watershed in East Java, Indonesia. *Journal of Water Land Development*, 37(V-VII), 141-152. <https://doi.org/10.2478/jwld-2018-0033>
- Sulianto, S., Bisri, M., Limantara, L. M., and Sisinggih, D. (2020). The Modified DISPRIN Model for Transforming Daily Rainfall-Runoff Data Series on a Small Watershed in Archipelagic Region. *Environmental Research Engineering and Management*, 76(2), 6-21. <https://doi.org/10.5755/j01.erem.76.2.20299>
- Talbi, E.-G. (2009). *Metaheuristics. Form Design to Implementation.* A John Wilwy and Sons, Inc., Publication. <https://doi.org/10.1002/9780470496916>
- Tayyab, M., Zhou, J., Zeng, X., Zhao, N., and Adnan, R. (2015). The integrated combination of the multi hydrological models by applying the least square method. *Research Journal of Applied Sciences, Engineering, and Technology*, 10(1), 107-111. <https://doi.org/10.19026/rjaset.10.2561>
- Tolson, B. A., and Shoemaker, C. A. (2007). The dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(November 2005), 1-16. <https://doi.org/10.1029/2005WR004723>
- Wang, W.-C., Cheng, C.-T., Chau, K.-W., and Xu, D.-M. (2012). Calibration of Xinanjiang model parameters using the hybrid genetic algorithm-based fuzzy optimal model. *Journal of Hydroinformatic*, 14(3), 784-799. <https://doi.org/10.2166/hydro.2011.027>
- Zhang, X, Hörmann, G., Fohrer, N., and Gao, J. (2012). Parameter calibration and uncertainty estimation of a simple rainfall-runoff model in two case studies. *Journal of Hydroinformatic*, 14(4), 1061-1074. <https://doi.org/10.2166/hydro.2012.084>
- Zhang, Xuesong, Srinivasan, R., Zhao, K., and Liew, M. Van. (2008). Evaluation of global optimization algorithms for the parameter calibration of a computationally intensive hydrologic model. *Hydrological Processes*, 23(3), 430-441. <https://doi.org/10.1002/hyp.7152>

