

Optimizing Parameters for Two Conceptual Hydrological Models Using a Genetic Algorithm: A Case Study in the Dau Tieng River Watershed, Vietnam

Trieu Anh NGOC^{1,3}, Kazuaki HIRAMATSU^{2*} and Masayoshi HARADA²

¹ Department of Agro-environmental Sciences, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University (Hakozaki, Fukuoka 812-8581, Japan)

² Department of Agro-environmental Sciences, Faculty of Agriculture, Kyushu University (Hakozaki, Fukuoka 812-8581, Japan)

Abstract

In recent years, many conceptual hydrological models have been constructed to calculate rainfall runoff for river watersheds, two of which, the Tank Model and the NAM Model, have been widely used in many Asian countries to forecast flooding and manage water resources because of their simplicity. However, obtaining good results can be time-consuming and costly because multiple model parameters must be calibrated. This requirement has led to an increased need for automated calibration. In this study, the two hydrological models had a genetic algorithm (GA) incorporated to model the rainfall runoff process and optimize model parameters. Calibration data were obtained at hydrological gauges of the river system upstream of the Dau Tieng River watershed, located along the upper Saigon River in Southeastern Vietnam. The GA optimization in this study concurrently adjusted eighteen of the Tank Model parameters and ten NAM Model parameters to improve modeling efficiency. The study concluded that both models showed good correlation between simulated and observed flows, with increased accuracy and convenience. The Tank Model produced better simulation results through error indicators such as root mean square error, efficiency, and relative error.

Discipline: Agricultural Engineering

Additional key words: NAM Model, Rainfall runoff, river watershed, Tank Model

Introduction

Many hydrological deterministic models have been developed to simulate the rainfall runoff process for river watersheds, but most have complex structures and require various observed data for calibration. Two models, the Tank Model and the NAM Model, have been widely used in many Asian countries not only because of their simple structures but also because of their simple data requirements^{12, 13}. However, these hydrological models still require extensive time and effort to calibrate various model parameters. Accordingly, parameter calibration has become the main challenge in developing hydrological models to represent the rainfall runoff process⁵, and the

demand for the application of optimization algorithms that automatically calibrate multiple model parameters has increased.

In recent years, various methods have been developed for the automatic calibration of hydrological model parameters, including the genetic algorithm (GA)^{1, 10}, Newton's method⁸, shuffled complex evolution⁴, and particle swarm optimization¹⁰.

In this study, one of these automated calibration techniques, the GA optimization search, was used to automatically determine the optimal parameters in each model. To calibrate and validate the models, several series of daily rainfall data sets were tested to determine the best parameters for each model in the simulation of the daily runoff.

³ Present address:

Faculty of Water Resources Engineering, Water Resources University–Second Base (No. 2 Truong Sa, Binh Thanh, Ho Chi Minh, Vietnam)

* Corresponding author: e-mail hiramatsu@bpes.kyushu-u.ac.jp

Received 17 February 2012; accepted 24 May 2012.

Material and Methodology

1. Study area

The selected study area, shown in Fig.1, is the Dau Tieng River watershed in Tay Ninh Province, approximately 90 km from Ho Chi Minh City, in southeastern Vietnam. It is located at the upstream portion of the Saigon River and is a main reach of the upper Saigon–Dongnai River system, outside the estuarine basin. The total watershed area is approximately 2,700 km², and most of the vegetation is brushwood, forest, and industrial cropland. The upper reach of the Saigon River connects with Cambodian river branches and discharges into the downstream portion of the Saigon–Dongnai River system. The watershed is a dendritic river system with a density of 0.39 km/km². The total length of the river in the watershed is approximately 130.5 km, and the average river slope is about 0.25%. The elevation of the watershed area varies from 24 to 100 m above mean sea level, and average annual rainfall is about 1800 mm.

2. Data

To evaluate the applicability of the two hydrological models incorporated with GA optimization in this study, daily rainfall, evaporation, and river discharge data were recorded at three gauges, Chon Thanh (CT), Tay Ninh (TN), and Dau Tieng (DT) respectively. The location and data obtained for each gauge are shown in Fig.1 and Table 1. These data were collected by the Dau Tieng Irrigation Exploitation and Management Company under the Ministry of Agriculture and Rural Development, Vietnam, and also provided by the Division of Applied Science & Technology, Water Resources University–Second Base, Vietnam.

The data used for model calibration included rainfall, evaporation, and observed runoff discharge data in 1995 and 2000. The year 1995 represented a hydrological year with rainfall less than the annual average, while the year 2000 represented a hydrological year with above-average rainfall. The calibrated models were validated in 2 years, 1998 and 2001, both of which represented typical hydro-

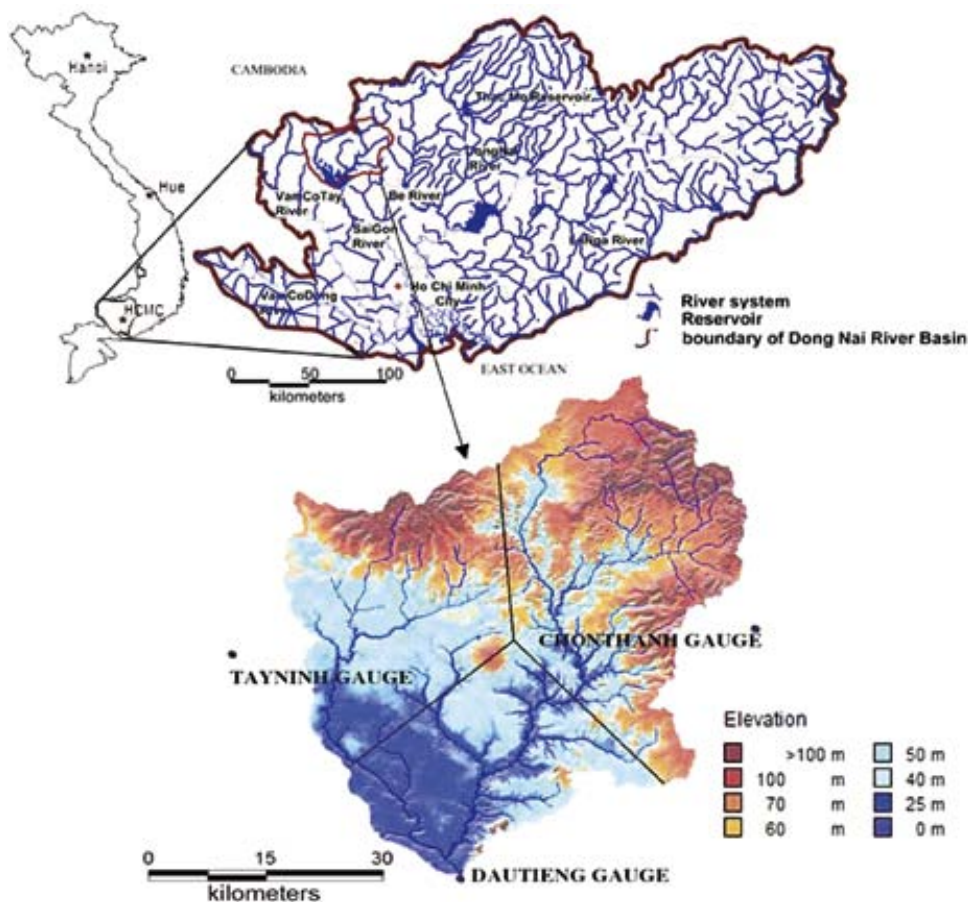


Fig. 1. Site location, 90-m DEM, and Thiessen polygon method for the Dau Tieng River watershed

logical years with high-quality observed data.

Because the daily rainfall data collected from each gauge did not reflect uniform rainfall distribution throughout the Dau Tieng River watershed, the Thiessen polygon method was used to obtain average watershed rainfalls, as shown in Fig. 1.

3. Hydrological NAM Model

NAM is an abbreviation for “Nedbor-Afstromings Model,” a Danish phrase meaning “precipitation runoff model.” The hydrological NAM Model simulates the rainfall runoff process that occurs at the watershed scale. The NAM Model forms part of the rainfall runoff module of

the MIKE 11 river modeling system and was originally developed at the Institute of Hydrodynamic and Hydraulic Engineering at the Technical University of Denmark^{7, 13}. Over the past decade, the NAM Model has been extensively applied and modified by the Danish Hydraulic Institute in many projects.

A lumped conceptual model of the NAM Model treats each subcatchment as a unit. The NAM Model simulates the rainfall runoff process in rural catchments and has 10 parameters: U_{max} , L_{max} , CQ_{OF} , CQ_{IF} , T_{OF} , T_{IF} , CK_1 , CK_2 , T_G , and CK_{BF} (snow storage was not considered in this study). The various components of the rainfall runoff process represent the average values for the entire subcatchment by continuously accounting for water contents in 4 different but mutually interrelated forms of storage, namely: snow, surface, lower zone, and groundwater. The routine for overland flow, interflow, and baseflow, as shown in Fig. 2, is also based on the linear reservoir.

Moisture intercepted on vegetation as well as water trapped in depressions and in the uppermost, cultivated part of the ground is represented as surface storage. U_{max} denotes the upper limit of surface water storage.

Evapotranspiration demand is initially met at the potential rate from the surface storage. If moisture content, U , in the surface storage is less than this requirement, the remaining fraction is assumed to be withdrawn by root activity from the lower zone storage at an actual rate, E_a . The value E_a is set to be proportional to potential evapotranspiration, E_p , according to:

$$E_a = E_p * \frac{L}{L_{max}} \quad (1)$$

where L and L_{max} are the actual and maximum possible moisture contents, respectively, in the lower zone storage.

When the surface storage spills, $U \geq U_{max}$, the excess maximum water, P_n , induces overland flow as well as infiltration. Q_{OF} denotes the portion of P_n that contributes to overland flow. Q_{OF} is assumed to be proportional to P_n and to vary linearly with the relative soil moisture content, L/L_{max} , of the lower zone storage.

Accordingly, overland flow, Q_{OF} , is determined as:

$$\begin{cases} Q_{OF} = CQ_{OF} \frac{L_{t-1}/L_{max} - T_{OF}}{1 - T_{OF}} P_n & \text{for } L_{t-1}/L_{max} > T_{OF} \\ Q_{OF} = 0 & \text{for } L_{t-1}/L_{max} \leq T_{OF} \end{cases} \quad (2)$$

where L denotes the soil moisture content of the lower zone storage, CQ_{OF} and T_{OF} are positive constants less than unity and without dimension, and t is time.

The interflow contribution, Q_{IF} , is assumed to be

Table 1. Data obtained at each gauge

Station	Sub-catchment Area (km ²)	Thiessen Weight	Data type (daily)
Tay Ninh	810	0.3	Rainfall
Chon Thanh	810	0.3	Rainfall
Dau Tieng	1,080	0.4	Rainfall Evaporation Discharge

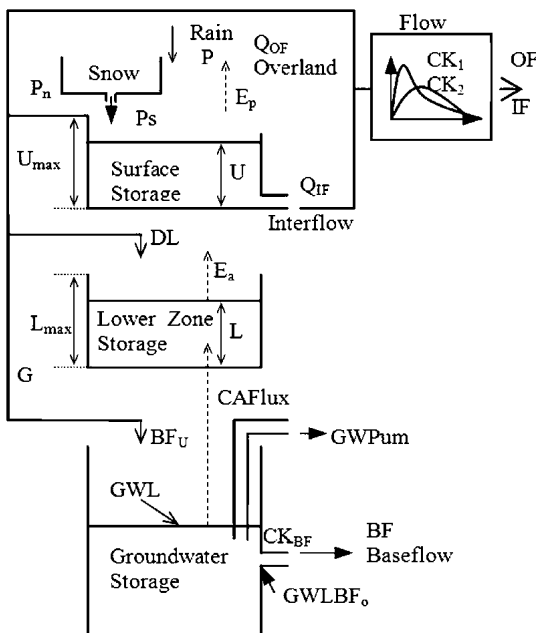


Fig. 2. Structure of the NAM model

proportional to U and to vary linearly with the relative moisture content, L/L_{max} , of the lower zone storage. Q_{IF} is determined as:

$$\begin{cases} Q_{IF} = CQ_{IF} \frac{L_{t-1}/L_{max} - T_{IF}}{1 - T_{IF}} U_t & \text{for } L_{t-1}/L_{max} > T_{IF} \\ Q_{IF} = 0 & \text{for } L_{t-1}/L_{max} \leq T_{IF} \end{cases} \quad (3)$$

where CQ_{IF} and T_{IF} are the time constant and root zone threshold value for interflow respectively.

The proportion of excess rainfall, P_n , that does not run off as overland flow infiltrates into the lower zone storage representing the root zone. A portion DL of the amount of infiltration, $P_n - Q_{OF}$, is assumed to increase soil moisture content, L , in the lower zone. G is assumed to percolate deeper and recharge groundwater storage.

$$\begin{cases} G = (P_n - Q_{OF}) \frac{L_{t-1}/L_{max} - T_G}{1 - T_G} & \text{for } L_{t-1}/L_{max} > T_G \\ G = 0 & \text{for } L_{t-1}/L_{max} \leq T_G \\ DL = (P_n - Q_{OF}) - G \end{cases} \quad (4)$$

where T_G is the root zone threshold value for groundwater recharge.

Percolation, G , is routed through a linear reservoir with the time constant, CK_{BF} , before reaching the groundwater table as recharge, BF_u .

The base flow is determined as:

$$BF_{u(t)} = BF_{u(t-1)} e^{-\left(\frac{t}{CK_{BF}}\right)} + G_t (1 - e^{-\left(\frac{t}{CK_{BF}}\right)}) \quad (5)$$

Based on meteorological data input, the NAM Model produces watershed runoff and other information concerning the land phase of the hydrological cycle such as temporal variation in evapotranspiration, soil moisture content, groundwater recharge, and groundwater levels. The resulting watershed runoff is conceptually divided into overland flow, interflow, and baseflow components^{2,3}.

4. Hydrological Tank Model

The Tank Model is a synthetic flow model based on rainfall in a watershed, which was developed and introduced in 1956 by a Japanese hydrologist, Dr. Masami Sugawara, who has authored many published works about his research and its practical applications. The model has been widely used worldwide and was positively evaluated by the World Meteorological Organization¹⁴. The Tank Model can be applied to reproduce streamflow from observed rainfall data from the watershed to plan, design and manage water resources. In Vietnam, the Tank Model has

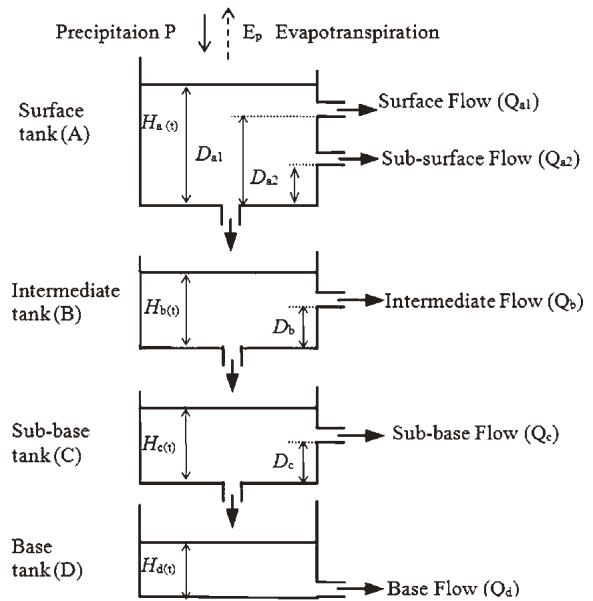


Fig. 3. Structure of the Tank model

been applied in many studies and is considered moderately suitable for river and stream systems.

The hydrological Tank Model used in this study has a simple structure with 4 tanks, a surface tank (A), an intermediate tank (B), a sub-base tank (C), and a base tank (D)⁶, laid vertically in a series, as shown in Fig.3. Precipitation on the watershed minus evapotranspiration is entered into the model. The 2 assumptions of the Tank Model are that (1) water can fill the storage that lies beneath and (2) water flows from a horizontal outlet in each tank and the total amount of water flowing represents the runoff. Each tank has a vertical outlet at the bottom (except Tank D) and one horizontal outlet at the side (except Tank A, which typically has 2 horizontal outlets). Rainwater falls into Tank A and then partly through the vertical outlet into the tank below. The remainder of the rainwater pours into the horizontal outlets to create flow when the water level in the tank exceeds the height of a horizontal outlet^{5,11}.

The total outflow, $Q_{(t)}$, at time t from the side outlets of all tanks represents the accumulation of the outflows from the river system in the watershed and can be expressed as follows:

$$Q_{(t)} = \{Q_{A1(t)} + Q_{A2(t)} + Q_{B(t)} + Q_{C(t)} + Q_{D(t)}\} \quad (6)$$

Given the initial conditions of the water levels in storage tanks A, B, C, and D at the initial time step, the storage in each tank is updated as follows:

$$H_{A(t+1)} = H_{A(t)} + P_{(t)} - E_{(t)} - Q_{A1(t)} - Q_{A2(t)} - I_{A(t)} \quad (7)$$

$$H_{B(t+1)} = H_{B(t)} + I_{A(t)} - Q_{B(t)} - I_{B(t)} \quad (8)$$

$$H_{C(t+1)} = H_{C(t)} + I_{B(t)} - Q_{C(t)} - I_{C(t)} \quad (9)$$

$$H_{D(t+1)} = H_{D(t)} + I_{C(t)} - Q_{D(t)} \quad (10)$$

where H is the water storage level (mm), P is rainfall (mm/day), E is evapotranspiration (mm/day), Q is total runoff (mm/day), and t is time step (day). I is the infiltration through the vertical outlet into the tank below (mm/day).

Although the Tank Model provides some indication of the lag time between rainfall and runoff, this lag time is often insufficient. In rainfall runoff, when discharge increases quickly, velocity also increases. Accordingly, lag time must decrease and is inversely proportional to velocity. In small watersheds, the lag time is often short; even if the velocity increases due to a change in discharge, we can assume this lag time is constant¹⁵. However, in large watersheds with a long lag time, we might need to consider an artificial lag time, TL , of watershed discharge, as follows:

$$Q_{E(t)} = (1 - D_{(TL)}) \times Q_{(t+TL)} + D_{(TL)} \times Q_{(t+TL)+1} \quad (11)$$

where $Q_{E(t)}$ is the calculated discharge, TL is the lag time, $[TL]$ is the integer part of TL , and $D_{(TL)}$ is the decimal part of TL .

5. Genetic algorithm method

The GA was originally developed and introduced in 1975 by John Holland^{1, 8}. It is a population-based optimization method that mimics the process of natural selection and natural evolution. The GA is used to search large, nonlinear spaces where expert knowledge is lacking or is difficult to encode⁹. The GA optimization search uses the idea of fitness to analyze various solutions and generate a new and better solution.

The GA begins with a randomly generated initial set of solutions called the population. Each individual in the population is called a chromosome, a string of symbols that is encoded into binary code, which represents the solution to a problem. The chromosome develops through consecutive repetitive revolutionary processes, called generations. During each generation, the chromosomes are assessed by fitness function, whereupon they undergo several main processes: selection, crossover, and mutation. To create

the new generation, parent chromosomes with higher fitness values are more likely to be selected, hence crossover and mutation processes are conducted to reproduce new offspring. These processes are repeated and stop only when the condition is satisfied. After several generations, the fitter chromosomes converge to the best chromosome, which represents the optimal solution to the problem.

6. Fitness function and error indicators

In this study, fitness function was based on the error indicator mean square error (MSE) and was used to evaluate GA optimization performance in calibrating model parameters. The equation of the fitness function, F_t , is given by equations (12) and (13):

$$\text{Maximum of } [F_t] = \text{Maximum of } \left[\frac{1}{MSE} \right] \quad (12)$$

$$MSE = \frac{\sum_{i=1}^N [Q_{Obs,i} - Q_{Sim,i}]^2}{\sum_{i=1}^N [Q_{Obs,i} - \overline{Q_{Obs,i}}]^2} \quad (13)$$

where obs, i is the observed discharge at the i^{th} time step, sim, i is the simulated discharge at the i^{th} time step, $\overline{Q_{Obs}}$ is the average of the observed discharge, i is the time step (day), and N is the total number of time steps.

According to national forecasting criteria in Vietnam, the percentage error of peak discharge, peak time, and total runoff volume are important indicators that evaluate the accuracy of simulated discharge. In this study, a comparison of simulated discharge accuracy with observed discharge was expressed by the error indicators coefficient of correlation (R), Nash-Sutcliffe coefficient (E_2), mean absolute error (MAE), root mean square error ($RMSE$), relative error (RE), and volume error (VE).

$$R = \frac{\sum_{i=1}^N [(Q_{Obs,i} - \overline{Q_{Obs}})(Q_{Sim,i} - \overline{Q_{Sim}})]}{\sqrt{\sum_{i=1}^N [Q_{Obs,i} - \overline{Q_{Obs}}]^2} \sqrt{\sum_{i=1}^N [Q_{Sim,i} - \overline{Q_{Sim}}]^2}} \quad (14)$$

$$E_2 = 1 - \frac{\sum_{i=1}^N [Q_{Obs,i} - Q_{Sim,i}]^2}{\sum_{i=1}^N [Q_{Obs,i} - \overline{Q_{Obs}}]^2} \quad (15)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Q_{Obs,i} - Q_{Sim,i}| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [Q_{Obs,i} - Q_{Sim,i}]^2} \quad (17)$$

$$RE_i = \left| \frac{Q_{Obs,i} - Q_{Sim,i}}{Q_{Obs,i}} \right| \times 100\% \quad (18)$$

$$VE_i = \frac{\left[\sum_i^N (Q_{Sim,i} \times 86400) - \sum_i^N (Q_{Obs,i} \times 86400) \right]}{\sum_i^N (Q_{Obs,i} \times 86400)} \times 100\% \quad (19)$$

Results and Conclusions

1. Results

The GA optimization search was combined into the parameter calibration of two hydrological models (GA–NAM and GA–Tank). Unlike other search techniques, GA optimization is generally conducted among a population using a coded parameter set and the probabilistic rules of roulette wheel selection. A GA optimization set consists of 4 parameters: crossover probability, mutation probability, population size, and maximum number of generations.

The crossover probability parameter controls the frequency of the crossover operation. If the crossover probability value is excessive, the structure of a high-quality solution could be damaged quickly; if the crossover probability value is too small, the search efficiency may be low. Generally, the crossover probability parameter is between 0.5 and 0.8. The mutation probability parameter is a critical factor that can lead to a new search direction in the solution space and extend population diversity. If this parameter is too small, new gene segments may not be inducted; if this parameter is too large, genetic evolution degenerates into a random local search. Generally, the mutation probability parameter is between 0.001 and 0.1. The population size parameter significantly affects solution quality and GA efficiency. If this parameter is too large, the computation time exceeds a tolerable limit and

the convergence time is prolonged. Generally, the population size parameter is between 150 and 300!

To select a set of GA parameters, their intercorrelations are considered, as mentioned above, and test runs for calibration data are conducted, as shown in Table 2. The lower and upper limits of each parameter for two hydrological models that define the GA search domain are also shown in Tables 3 and 4.

For model calibration, the GA operations shown in Table 2 were applied to the optimization search for 2 typical years, 1995 and 2000. NAM Model and Tank Model parameters calibrated by GA optimization are shown in Tables 5 and 6, respectively, and the results of error indicators by the GA optimization search are compared for the NAM and the Tank Models in Table 7.

For the GA–NAM Model, Table 7 and Figs. 4 and 5 indicate that the results of 2 simulated years showed similar performance with acceptable accuracy in simulated flow hydrographs. However, the calibrated parameters of the GA–NAM model differed between 1995 and 2000 (see Table 5); for example, the parameters U_{max} , L_{max} , CQ_{OF} , T_{OF} , T_{IF} , and T_G were 31.05, 233.94, 0.23, 0.22, 0.35, and 0.44, respectively, for 1995 and 22.99, 345.90, 0.75, 0.57, 0.07, and 0.001, respectively, for 2000. In addition, U_{max} was higher in 1995 than in 2000, indicating that surface-water storage capacity was larger in 1995 than in 2000.

Table 2. Parameters used for genetic operations

Bit-length per parameter	16
Population size	150
Generation	2,000
Crossover rate	0.7
Selection method	Roulette Wheel
Mutation rate	0.01

Table 3. NAM model parameters used for calibration

Parameter	Description	Lower limit	Upper limit
U_{max} (mm)	The maximum water content in surface storage	5	35
L_{max} (mm)	The maximum water content in root zone storage	50	350
CQ_{OF} (-)	Overland flow runoff coefficient	0	1
CQ_{IF} (h)	Time constant for routing interflow	500	1,000
T_{OF} (-)	Root zone threshold value for overland flow	0	0.9
T_{IF} (-)	Root zone threshold value for interflow	0	0.9
CK_1 (h)	The time constant for routing interflow	3	72
CK_2 (h)	The time constant for routing overland flow	3	72
T_G (-)	Root zone threshold value for groundwater	0	0.9
CK_{BF} (h)	Time constant for routing base-flow	500	5,000

Table 4. Tank model parameters used for calibration

Parameter	Functions	Description	Lower limit	Upper limit
C_{A1} (1/h)	$Q_{A1(t)} = C_{A1} \times (H_{A(t)} - D_{A1})$	Surface runoff coefficient	0	1
C_{A2} (1/h)	$Q_{A2(t)} = C_{A2} \times (H_{A(t)} - D_{A2})$	Sub-surface runoff coefficient	0	1
C_{A0} (1/h)	$I_{A(t)} = C_{A0} \times H_{A(t)}$	Infiltration coefficient	0	1
C_{B1} (1/h)	$Q_{B(t)} = C_{B1} \times (H_{B(t)} - D_B)$	Intermediate runoff coefficient	0	1
C_{B0} (1/h)	$I_{B(t)} = C_{B0} \times H_{B(t)}$	Infiltration coefficient	0	1
C_{C1} (1/h)	$Q_{C(t)} = C_{C1} \times (H_{C(t)} - D_C)$	Sub-base runoff coefficient	0	1
C_{C0} (1/h)	$I_{C(t)} = C_{C0} \times H_{C(t)}$	Infiltration coefficient	0	1
C_{D1} (1/h)	$Q_{D(t)} = C_{D1} \times H_{D(t)}$	Base runoff coefficient	0	0.1
D_{A1} (mm)		Height of surface outlet	0	500
D_{A2} (mm)		Height of sub-surface outlet	0	500
D_B (mm)		Height of intermediate outlet	0	500
D_C (mm)		Height of sub-base outlet	0	500
SA (mm)		Initial storage of Tank A	0	500
SB (mm)		Initial storage of Tank B	0	500
SC (mm)		Initial storage of Tank C	0	500
SD (mm)		Initial storage of Tank D	0	1,000
SM (mm)		Limit moisture threshold	0	10
TL (h)		Time lag	0	72

Table 5. Calibrated NAM model parameters

Parameter	Year 1995	Year 2000
U_{\max} (mm)	31.05	22.99
L_{\max} (mm)	233.94	345.90
CQ_{OF} (-)	0.23	0.75
CQ_{IF} (h)	329.78	200.00
T_{OF} (-)	0.22	0.57
T_{IF} (-)	0.35	0.07
CK_1 (h)	27.01	17.30
CK_2 (h)	64.01	40.90
T_G (-)	0.44	0.001
CK_{BF} (h)	4,875.10	4,069.52

Table 6. Calibrated Tank model parameters

Parameter	Year 1995	Year 2000
C_{A1} (1/h)	0.38	0.17
C_{A2} (1/h)	0.52	0.79
C_{A0} (1/h)	0.88	0.59
C_{B1} (1/h)	0.65	0.51
C_{B0} (1/h)	0.60	0.09
C_{C1} (1/h)	0.47	0.14
C_{C0} (1/h)	0.48	0.07
C_{D1} (1/h)	0.0018	0.0014
D_{A1} (mm)	269.60	242.80
D_{A2} (mm)	203.46	130.49
D_B (mm)	346.70	427.83
D_C (mm)	418.52	264.15
SA (mm)	87.36	139.68
SB (mm)	126.29	36.34
SC (mm)	252.60	12.55
SD (mm)	435.24	553.7
SM (mm)	3.00	1.62
TL (h)	8.15	3.17

Meanwhile, CQ_{OF} and T_{OF} were lower in 1995 than in 2000, showing that the rate of contributed overland flow was lower in 1995 than in 2000. Because 1995 represented a dry hydrological year with rainfall under the annual average, the amount of potential water in surface storage increased, overland flow contribution fell, and the root threshold capacity for groundwater recharge was high. Conversely, 2000 was a hydrological year with rainfall

slightly above average; therefore, the rate of contributed overland flow was higher (U_{\max} was lower, CQ_{OF} and T_{OF} were higher) and the amount of potential water in surface storage (L_{\max} was higher) was lower in 2000 than in 1995.

According to the calibration results of the GA–NAM Model presented in Table 7, the error indicators R , E_2 , and MSE were 0.86, 0.72, and 0.28, respectively, in 1995 and 0.93, 0.86, and 0.14, respectively, in 2000, showing that

Table 7. Results of error indicators for the GA–NAM and GA–Tank models

Error indicators	GA- NAM				GA-Tank			
	Calibration		Validation		Calibration		Validation	
	1995	2000	1998	2001	1995	2000	1998	2001
<i>R</i>	0.86	0.93	0.91	0.91	0.90	0.93	0.88	0.92
E_2	0.72	0.86	0.82	0.82	0.80	0.86	0.77	0.83
<i>MSE</i>	0.28	0.14	0.18	0.18	0.20	0.14	0.23	0.17
<i>RMSE</i>	1.66	2.27	1.68	0.18	1.39	2.29	1.88	1.67
max. <i>RE</i> (%)	34.80	19.07	31.71	41.59	28.22	30.44	29.73	28.87
<i>MAE</i>	3.13	6.90	3.23	1.20	1.70	0.47	1.71	7.05
Observed peak flow (m ³ /s)	464.37	1,078.40	579.12	409.34	464.37	1,078.40	579.12	409.34
Observed peak time	9/22	10/10	9/30	10/30	9/22	10/10	9/30	10/30
Simulated peak flow (m ³ /s)	421.00	981.73	459.33	348.60	429.89	965.63	529.10	381.95
Simulated peak time	9/22	10/10	9/30	10/30	9/22	10/10	9/30	10/30
Peak flow error (m ³ /s)	43.37	96.67	119.80	60.74	34.47	112.77	50.02	27.39
<i>RE</i> of peak flow (%)	9.34	8.96	20.69	14.84	7.42	10.46	8.64	6.69
Observed volume (10 ⁹ m ³)	1.49	2.52	1.88	2.88	1.49	2.52	1.88	2.88
Simulated volume (10 ⁹ m ³)	1.40	2.30	1.78	2.84	1.55	2.50	1.94	3.10
Volume error (%)	-6.61	-8.64	-5.41	-1.32	3.58	-0.59	2.85	7.73

all error indicators were lower in 2000 than in 1995. In particular, max.*RE* and *RE* of peak flow values were 19.07 and 8.96%, respectively, in 2000 and lower than the 1995 values (34.80 and 9.34%, respectively). These results demonstrate that the simulated discharge of the GA–NAM model was more appropriate and accurate in 2000 than in 1995, hence the calibrated parameters produced in 2000 were selected for validation.

The GA–NAM model validation results are shown in Figs. 6 and 7 and in Table 7. The GA–NAM model showed good correlation between the observed and simulated flow hydrographs in the validation as well as in the calibration. In the validation, the error indicators $R = 0.91$, $E_2 = 0.82$, and $MSE = 0.18$ were obtained in 1998 and 2001, showing that the calibrated parameters in 2000 ($R = 0.93$, $E_2 = 0.86$, and $MSE = 0.14$) provided a versatile model, although the accuracy was slightly lower. The volume errors in the 2 validated years, 1998 and 2001, were -5.41 and -1.32%, respectively, and were less than the volume error in 2000 (-8.64%). We concluded that the calibrated parameters obtained in 2000 provided stabilizing and versatile forecasts for the NAM model.

Table 6 shows the calibrated parameters of the GA–Tank model, while Figs. 4 and 5 show comparisons between simulated and observed flow hydrographs in the calibration. The calibrated parameters in 2000 resulted in the error indicators $R = 0.93$, $E_2 = 0.86$, $MSE = 0.14$, $RMSE = 2.29$, and $RE = 30.44$, as shown in Table 7, indicating that simulation results were more accurate in 2000 than in 1995. The calibrated parameters in 1995 also show good

correlation in a comparison between observed and simulated discharges. However, the parameters obtained in the 2 calibrated years differed, especially for C_{A1} , C_{A2} , C_{A0} , C_{B1} , C_{B0} , C_{C1} , C_{C0} , D_{A1} , D_{A2} , D_B , and D_C . The values for parameters C_{A0} , D_{A1} , and D_{A2} were 0.88, 269.60, and 203.46, respectively, in 1995 and exceeded the values obtained in 2000 (0.59, 242.80, and 130.49, respectively) because the potential amount of water in the surface Tank was relatively higher in 1995 than in 2000. C_{A2} was 0.58 in 1995 and 0.79 in 2000, which means that the subsurface flow from the Tank model in 2000 exceeded that in 1995 for reasons similar to those for the GA–NAM model. In 1995, which represented a dry hydrological year, the ratios of intermediate flow and sub-base flow to total outflow increased and the ratio of surface flow decreased compared to those in 2000. This result was reflected in the calibrated parameters $C_{A0} = 0.88$, $C_{B1} = 0.65$, $C_{B0} = 0.60$, $C_{C1} = 0.47$, $C_{C0} = 0.48$, $D_B = 346.70$, and $D_C = 418.52$ in 1995 and $C_{A0} = 0.59$, $C_{B1} = 0.51$, $C_{C1} = 0.09$, $C_{C0} = 0.14$, $D_B = 0.07$, and $D_C = 264.15$ in 2000. Based on the calibration results of the GA–Tank model in 1995 and 2000, shown in Table 7, the calibrated parameters in 2000, shown in Table 6, were selected to validate the model in 2 years, 1998 and 2001. Comparisons between simulated and observed flow hydrographs for 1998 and 2001 are shown in Figs. 6 and 7 as the validation results. The simulated flow hydrographs using the parameters obtained in calibration year 2000 correlated well with the observed hydrographs. The error indicators in the validation years 1998 and 2001 showed accuracy almost equivalent to those in the calibration year

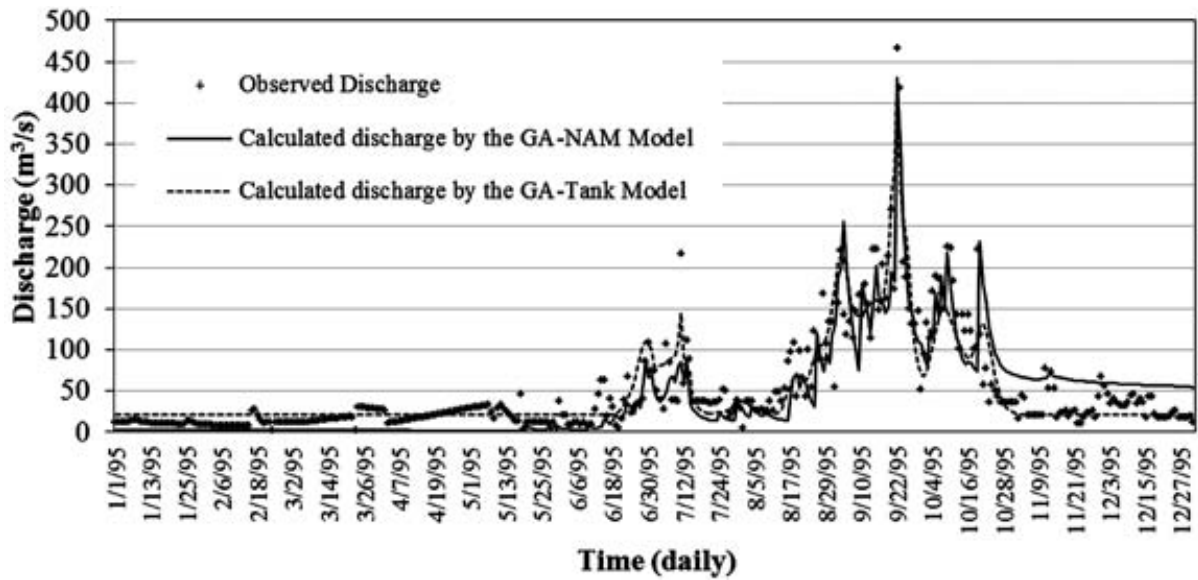


Fig. 4. Model calibration of daily discharges in 1995

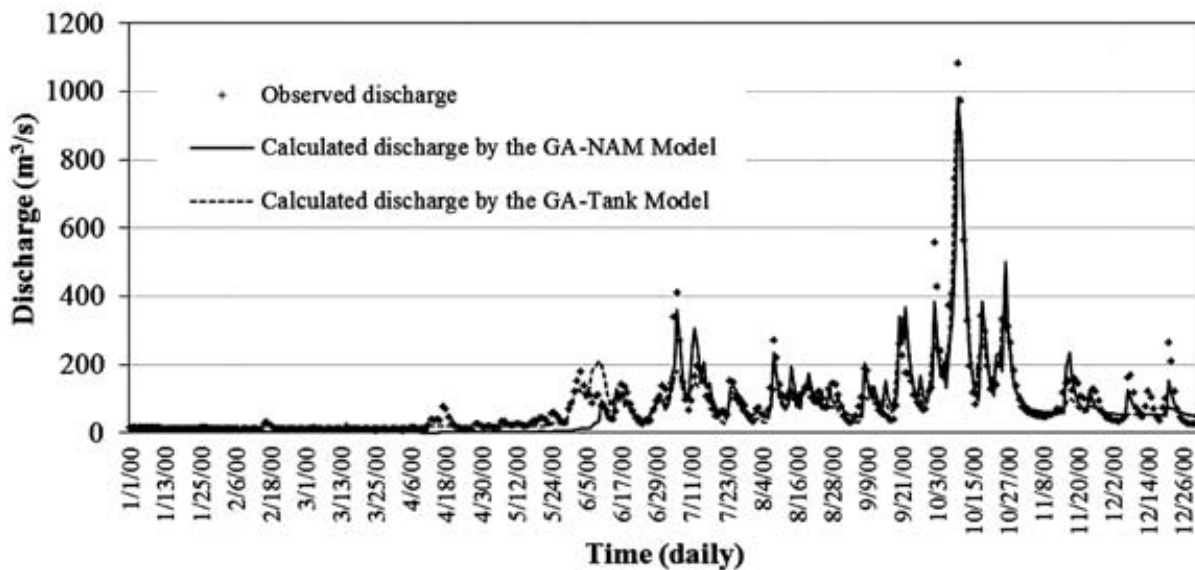


Fig. 5. Model calibration of daily discharges in 2000

2000. We concluded that the calibrated parameters in 2000 provided stabilizing and versatile forecasts for the Tank model.

2. Comparison of GA-Tank and GA-NAM models

GA-NAM and GA-Tank models have similar internal structures. In the GA-NAM model, overland flow is produced by the excess capacity of the upper storage representing initial abstraction and interception loss. In

the GA-Tank model, surface flow is the outflow from the side outlets of the surface tank. The parameter SM in the GA-Tank model shows some similarity with the parameter U_{max} in the GA-NAM model. These parameters are filled before infiltration is initiated.

The intermediate flow in the GA-Tank model is the outlet of the intermediate tank. It resembles the interflow in the GA-NAM model, which is proportional to the amount of water in surface storage and varies linearly

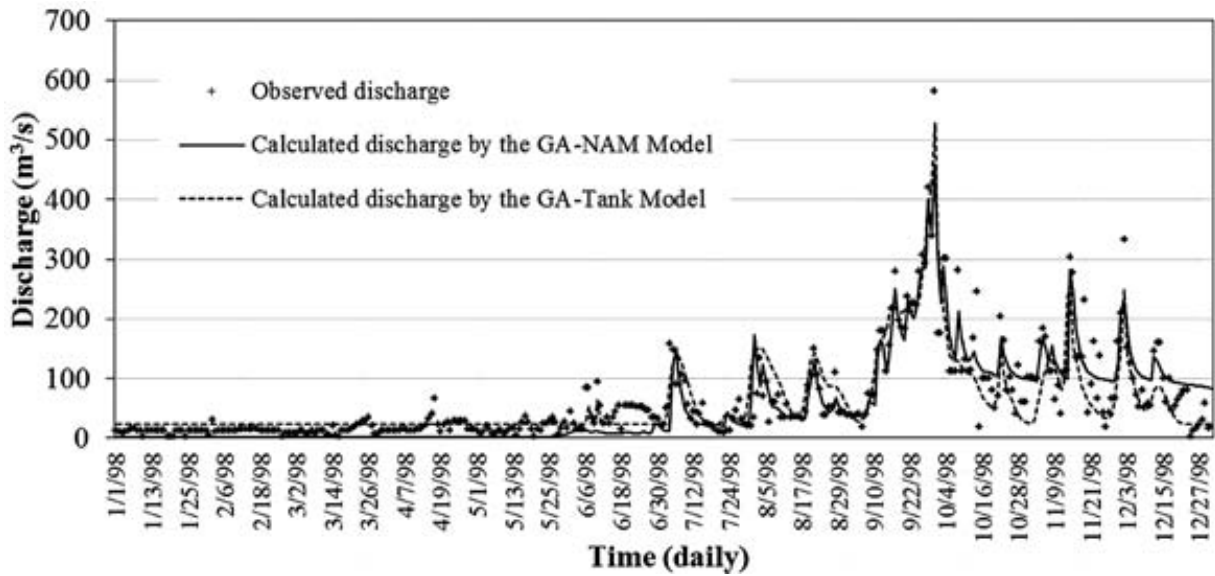


Fig. 6. Model validation of daily discharges in 1998 using calibrated parameters in 2000

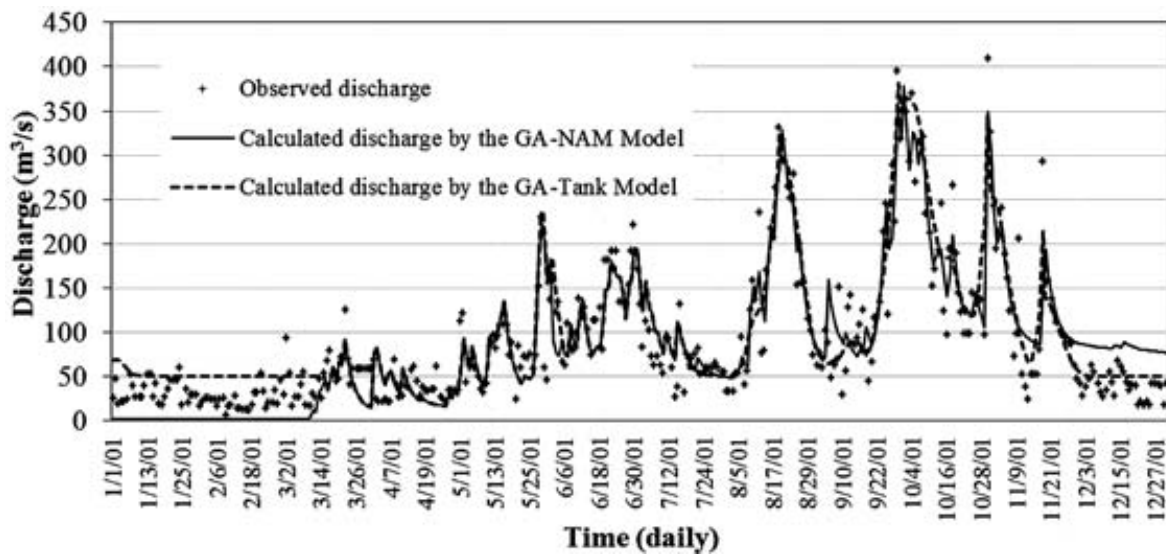


Fig. 7. Model validation of daily discharges in 2001 using calibrated parameters in 2000

with the relative soil moisture content of the lower zone storage. In this flow, the parameters C_{B1} in the GA-Tank model and T_{IF} in the GA-NAM model represent the same types of interflow thresholds.

In the routing of flow components, the total outflow in the GA-Tank model is the summation of outflows from the side outlets of all tanks. In the GA-NAM model, overland flow and interflow are routed through 2 linear reservoirs, while groundwater flow is routed through a single

linear reservoir. These flow components are represented by adding up and routing through a final single linear reservoir to produce a total runoff at the outlet point of the watershed.

Based on these structural similarities, the GA-NAM and GA-Tank models produced similar calibrated parameters. In 1995, the calibrated parameter values were $U_{max} = 31.05$ and $T_{IF} = 0.35$ in the GA-NAM model and $SM = 3.0$ and $C_{B1} = 0.65$ in the GA-Tank model and exceeded

the calibrated parameter values in 2000 ($U_{\max} = 22.99$ and $T_{IF} = 0.07$ in the GA–NAM model and $SM = 1.62$ and $C_{B1} = 0.51$ in the GA–Tank model).

The peak flow, peak time, and volume error indicators are also important for evaluating model performance. As shown in Table 7, the peak flow errors of simulated flow were higher in the GA–NAM model (119.80 and 60.74 in 1998 and 2001) than in the GA–Tank model (50.02 and 27.39 in 1998 and 2001). These results indicate that the tank model with the GA optimization search was better than the GA–NAM model for simulating daily runoff in the Dau Tieng River watershed.

The values of R and E_2 are essential coefficients to estimate model performance. Table 8 shows the average values of R and E_2 over the calibration and validation processes, and the GA–Tank model obtained mean values of R and E_2 slightly higher than in the GA–NAM model.

3. Conclusions

In this paper, GA, a powerful optimization technique, was integrated to the NAM and Tank models and applied to discharge simulations during discharge periods in the Dau Tieng River watershed. GA was thus enhanced to perform a search of optimal parameters for 2 hydrological models by comparing hydrograph shapes of simulated flow, observed flow, and error indicators.

Although the GA–NAM and GA–Tank models have some basic structural differences, they are similar in terms of basic conceptualization¹⁶. The GA–NAM model has fewer model parameters than the GA–Tank model, but the calibration process by the GA optimization search was conducted with multiple generations (2000 in this study). The calculation time of the GA–NAM model took around 75 minutes while that of the GA–Tank model was 85 minutes. The GA–Tank model has few parameters; thus, applying the limit of the GA optimization search was not important.

In a comparison of the 2 hydrological models, calibration and validation results were similar, although error indicators showed that performance was slightly higher in the GA–Tank model than in the GA–NAM model. Most

errors caused a high Max.RE value because flow peaks and volume error were not captured.

The results of this study showed the ability to simulate the models under each condition. However, the GA–Tank and GA–NAM model performance was highly dependent on input data quality and the specific characteristics of the rainfall periods, and model results control simulated output accuracy in terms of timeliness and magnitude.

Acknowledgments

We express special appreciation to Dr. Do Van Khiet, Head of the Division of Applied Science & Technology, Water Resources University–Second Base, Vietnam, and Dr. Nguyen Thai Quyet of the Institute for Water Resource & Environment for their assistance with the data collection. We would also like to thank Dau Tieng Irrigation Exploitation and Management Company for their research assistance. Last, but not least, we greatly appreciate the support under the FY2012–2014 JSPS Core-to-Core Program “Collaborative Project for Soil and Water Conservation in Southeast Asian Watersheds” by Japan.

References

1. Chu-Tian et al. (2006) Using genetic algorithm and TOPSIS for Xinanjiang model calibration with single procedure. *J. Hydrol.*, **316**, 129–140.
2. DHI (2007) DHI MIKE 11 user manual. DHI Water and Environment, Denmark, 229–244.
3. DHI (2007) DHI MIKE 11 reference. DHI Water and Environment, Denmark, 290–325.
4. Duan et al. (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Res.*, **28**, 1015–1031.
5. Franchini, M. (1996) Use of genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. *Hydrological Sci. J.*, **41**, 21–39.
6. Gunawan, S. (2010) Research on development of rainfall runoff model for Bogowonto River basin by using Tank model. International Centre for Water Hazard and Risk Management: <http://www.icharm.pwri.go.jp/training/master/publication/pdf/2010/gunawan.pdf>
7. Havnø, K. et al. (1995) MIKE-11 a generalized river modelling package. In *Computer Models of Watershed Hydrology*, ed. V. P. Singh, Water Resources Publications, Colorado, 733–782.
8. Hendrickson, J. D. et al. (1988) Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. *Water Resources Res.*, **24**, 691–700.

Table 8. Average efficiency coefficients of the GA–Tank and GA–NAM models.

Coefficiency	GA–NAM	GA–Tank
R	0.90	0.90
E_2	0.80	0.82

9. Holland, J. D. (1975) *Adaptation in natural and artificial systems: An introductory analysis with application to biology, control, and artificial intelligence*. University of Michigan Press, Ann Arbor, MI.
10. Kuok, K. K. et al. (2010) Global optimization methods for calibration and optimization of the hydrological Tank model's parameters. *Can. J. Civil Engineer.*, **1**, 1-14.
11. Madsen, H. (2000) Automatic calibration of a conceptual rainfall runoff model using multiple objectives. *J. Hydrol.*, **235**, 276-288.
12. Ngoc, T. A. et al. (2011) Parameter identification for two conceptual hydrological models of upper Dau Tieng River watershed in Vietnam. *J. Facul. Agric.*, Kyushu University, **56**, 335-341.
13. Nielsen, S. A. & E. Hansen (1973) Numerical simulation of the rainfall runoff process on a daily basis. *Nordic Hydrol.*, **4**, 171-190.
14. Russell, J. S. (2003) *Perspectives In Civil Engineering*. American Society of Civil Engineers, 347.
15. Sugawara, M. (1995) Tank model. In *Computer Models of Watershed Hydrology*, ed. V. P. Singh, Water Resources Publications, Colorado, 165-214.
16. Tawatchai, T. & M. R. Gautam (2000) Application of Tank, NAM, ARMA and neural network models to flood forecasting. *Hydrol. Process.*, **14**, 2473-2487.