Application of a differential evolution optimizer to the tank model

C.A.G. Santos[1], P.K.M.M. Freire[1], P.M. Arruda[1], S.K. Mishra[2]
[1]*Federal University of Paraíba, João Pessoa, Brazil;* [2]*North-Eastern Hill University, Shillong, India*

ABSTRACT

The difficulties involved in calibration of conceptual models have been partly attributable to the lack of robust optimization tools. This paper presents the essential concepts and application to optimize the tank model for a Japanese watershed, with a global optimization method known as Differential Evolution (DE), which grew out of Price's attempts to solve the Chebychev Polynomial fitting Problem that had been posed to him by Rainer Storn. A breakthrough happened, when Price came up with the idea of using vector differences for perturbing the vector population. The crucial idea behind DE is a scheme for generating trial parameter vectors. The optimization technique was tested with the field data from Ishite river dam, which is the reservoir that supplies water to the city of Matsuyama, Japan. On the basis of these results, the parameter values are given, which could serve as an initial estimate for other similar Japanese watersheds.

Keywords: Optimization, tank model, Japanese watershed, differential evolution, modeling.

INTRODUCTION

The major problem concerning the use of hydrological models is the need of parameters which cannot be directly measured in the field, especially in nonlinear models. However, it is difficult to assure that the final values for the parameters are not a result of either a local minimum or another trap. Therefore, more robust algorithms are required to estimate the parameter's final values (e.g, Santos et al., 1994 and Santos et al., 2003). A global optimization method known as Differential Evolution (DE) grew out of Price's attempts to solve the Chebychev Polynomial fitting Problem that had been posed to him by Rainer Storn, and a breakthrough happened when Price came up with the idea of using vector differences for perturbing the vector population (Storn, 1997). The crucial idea behind DE is a scheme for generating trial parameter vectors. Initially, a population of points (p in d-dimensional space) is generated and evaluated (i.e. $f(p)$ is obtained) for their fitness. Then for each point (p_i) three different points (p_a, p_b and p_c) are randomly chosen from the population. A new point (p_z) is constructed from those three points by adding the weighted difference between two points ($w(p_b - p_c)$) to the third point (p_a). Then this new point (p_z) is subjected to a crossover with the current point (p_i) with a probability of crossover (c_r), yielding a candidate point, say p_u. This point, p_u, is evaluated and if found better than p_i then it replaces p_i else p_i remains. Thus we obtain a new vector in which all points are either better than or as good as the current points. This new vector is used for the next iteration. This process makes the differential evaluation scheme completely self-organizing.

Thus, the objective of this work is to use Differential Evolution (DE) method of optimization, for application with the tank model, which is a hydrological model whose basic principle consists of representing the river basin as a set of tanks in which the outflows of each tank are proportional to the water height from the respective outlets for the reservoir of Ishite dam that supplies water to the city of Matsuyama, Japan. The paper presents the general details of the hydrological modeling with the DE method.

DIFFERENTIAL EVOLUTION

The Differential Evolution method is a population based algorithm like genetic algorithms using the similar operators; crossover, mutation and selection. The main difference in constructing better solutions is that genetic algorithms rely on crossover while DE relies on mutation operation. This main operation is based on the differences of randomly sampled pairs of solutions in the population. The algorithm uses mutation operation as a search mechanism and selection operation to direct the search toward the prospective regions in the search space. The DE algorithm also uses a non-uniform crossover that can take child vector parameters from one parent more often than it does from others. By using the components of the existing population members to construct trial vectors, the recombination (crossover) operator efficiently shuffles information about successful combinations, enabling the search for a better solution space. The DE method consists of three basic steps: (i) generation of (large enough) population with N individuals [$x = (x_1, x_2, \ldots, x_m)$] in the m-dimensional space, randomly distributed over the entire domain of the function in question and evaluation of the individuals of the so generated by finding $f(x)$; (ii) replacement of this current population by a better fit new population, and (iii) repetition of this replacement until satisfactory results are obtained or certain criteria of termination are met.

The crux of the problem lays in replacement of the current population by a new population that is better fit. Here the meaning of 'better' is in the Pareto improvement sense. A set S_a is better than another set S_b ***if*** **:** (i) **no** $x_i \in S_a$ is inferior to the corresponding member of $x_i \in S_b$; ***and*** (ii) **at least one** member $x_k \in S_a$ is better than the corresponding member $x_k \in S_b$. Thus, every new population is an improvement over the earlier one.

To accomplish this, the DE method generates a candidate individual to replace each current individual in the population. The candidate individual is obtained by a crossover of the current individual and three other randomly selected individuals from the current population. The crossover itself is probabilistic in nature. Further, if the candidate individual is better fit than the current individual, it takes the place of the current individual, else the current individual stays and passes into the next iteration. The crossover scheme (called exponential crossover, as suggested by Kenneth Price in his personal letter to the third author) is given below. This is coded for ncross $\geqslant$ 1 in the program.

The mutant vector is $\mathbf{v}_{i,g} = \mathbf{x}_{r1,g} + F(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g})$ and the target vector is $\mathbf{x}_{i,g}$ and the trial vector is $\mathbf{u}_{i,g}$. The indices r_1, r_2 and r_3 are randomly but different from each other. $U_j(0,1)$ is a uniformly distributed random number between 0 and 1 that is chosen anew for each parameter as needed.

Step 1: Randomly pick a parameter index $j = j_{\text{rand}}$.

Step 2: The trial vector inherits the j^{th} parameter (initially = j_{rand}) from the mutant vector, i.e., $u_{j,i,g} = v_{j,i,g}$.

Step 3: Increment j; if $j = D$ then reset $j = 0$.

Step 4: If $j = j_{\text{rand}}$ end crossover; else goto **Step 5**.

Step 5: If $C_r <= U_j(0,1)$, then goto Step 2; else goto **Step 6**.

Step 6: The trial vector inherits the j^{th} parameter from the target vector, i.e., $u_{j,i,g} = x_{j,i,g}$.

Step 7: Increment j; if $j = D$ then reset $j = 0$.

Step 8: If $j = j_{\text{rand}}$ end crossover; else goto **Step 6**.

There could be other schemes (as many as 10 in number) of crossover, including no crossover (only probabilistic replacement, NCROSS $\leqslant$ 0 that works better in case of a few functions.

TANK MODEL

Many models have been developed to simulate rainfall–runoff processes, and one of them is known as the "tank model". The model has been modified and some mathematical tools have been adapted in order to improve its efficiency (Lee & Singh, 1999). A tank model is a conceptual representation of a basin hydrological process. It simulates the wetness of the several soil layers by tanks arranged vertically in series, and each one adapted with one or more outlets to account for water flow and filtration to lower layers.

Precipitation is put into the top tank, and evaporation is subtracted from the top tank. If there is no water in the top tank, evaporation is subtracted from the second tank; if there is no water in both the top and the second tank, evaporation is subtracted from the third tank. The outputs from the side outlets are the calculated runoffs. The output from the top tank is considered as surface runoff, output from the second tank as sub-base runoff and output from the third tank as baseflow. Thus, the outflow or seepage from each tank is assumed to be proportional to the water height from the whole position of discharge or seepage. Water depth of the tank is assumed to be the storage in the basin. In this paper, the tank model was implemented with three tanks. Figure 1 shows how the tank model was implemented.

The parameters to be optimized are the runoff parameters a_1, a_2, a_3 and a_4; the infiltration parameters b_1, b_2 and b_3; and the height of the runoff outlets h_1, h_2, h_3 and h_4. These quantities and the tank model are defined by the following expressions:

$$y_1(t) = a_1[X_1(t) - h_1] \tag{1}$$

$$y_2(t) = a_2[X_1(t) - h_2] \tag{2}$$

$$y_3(t) = a_3[X_2(t) - h_3] \tag{3}$$

$$y_4(t) = a_4[X_3(t) - h_4] \tag{4}$$

$$z_1(t) = b_1 X_1(t) \tag{5}$$

$$z_2(t) = b_2 X_2(t) \tag{6}$$

$$z_3(t) = b_3 X_3(t) \tag{7}$$

$$Q(t) = y_1(t) + y_2(t) + y_3(t) + y_4(t) \tag{8}$$

$$X_1(t) = X_1(t-1) + P(t) - y_1(t) - y_2(t) - z_1(t) \tag{9}$$

$$X_2(t) = X_2(t-1) + z_1(t) - y_3(t) - z_2(t) \tag{10}$$

$$X_3(t) = X_3(t-1) + z_2(t) - y_4(t) - z_3(t) \tag{11}$$

where t is the day index; $y_1(t)$, $y_2(t)$, $y_3(t)$ and $y_4(t)$ are the runoffs from outlets at day t; $z_1(t)$, $z_2(t)$ and $z_3(t)$ are the values of infiltration of each tank at day t; $X_1(t)$, $X_2(t)$ and $X_3(t)$ are the storages in depth at day t; $Q(t)$ is the total runoff at day t; and $P(t)$ is the precipitation at day t.

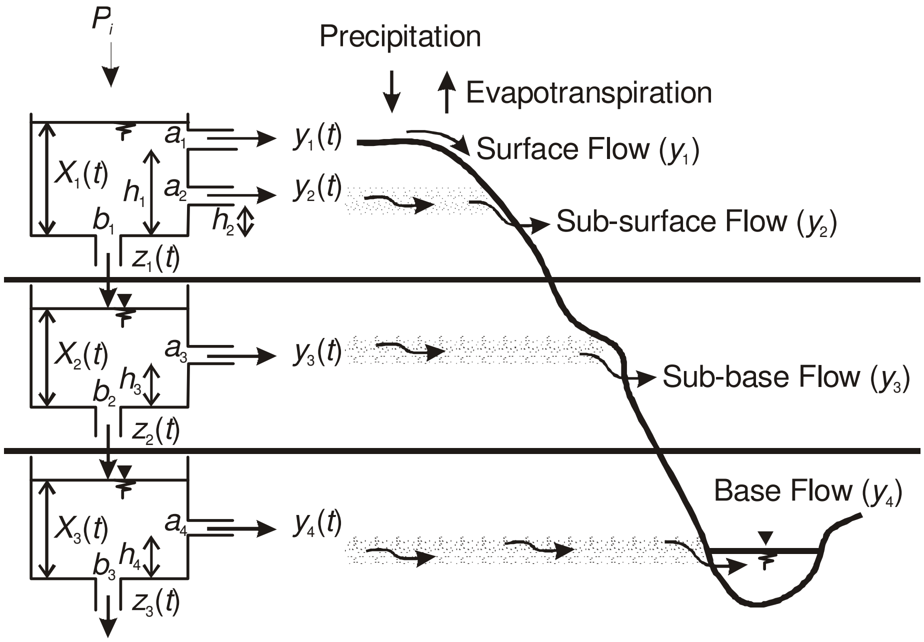

Fig. 1 Schematic design tank model with three tanks.

FIELD DATA

Ishite River basin is a sub-basin of Shigenobu River basin in Matsuyama city located in Shikoku Island, Japan (Fig. 2). The basin is 72.5 km^2, the river is 11 km long, and most of the basin is covered by pine forest. The daily rainfall and runoff data from January 1992 up to December 2003 at Ishite River dam were used.

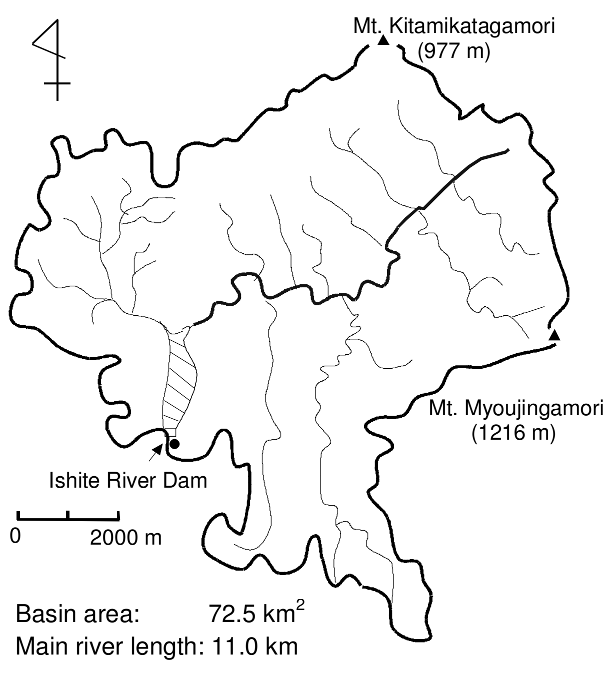

Fig. 2 Map of the Ishite River basin.

The annual mean precipitation depth is between 1300 and 1500 mm, and the rainy season is from the middle of June to the middle of July, with the typhoon season being from August to October. The selected period of observed data for daily rainfall and runoff was 1992 to 1993 for the calibration process, while the years from 1994 up to 2003 were used to validate the calibrated tank model.

APPLICATION AND RESULTS

Setting of the DE parameters

Before hand, it is necessary to setup certain parameters, which were fixed as: max number of iterations allowed, iter = 10000, population size, N = 110 (it should be 10 times of the dimension of the function or 100 whichever maximum); scheme of crossover, NCROSS = 1 (defined in the program); crossover probability, PCROS = 0.9 (suggested to be about 0.85 to 0.99); scale factor, FACT = 0.5 ($0.5 \leqslant$ FACT < 1.0); random number seed, IU = 1171 and all random numbers are uniformly distributed between -1000 and 1000; accuracy needed, which determines accuracy for termination, EPS = 10^{-8}. If x in $f(x)$ violates the boundary then it is forcibly brought within the specified limits through replacing it by a random number lying in the given limits of the function concerned.

Optimization of the rainfall–runoff model

There are 11 parameters in the tank model to be determined by optimization, which are a_1, a_2, a_3, a_4, b_1, b_2, b_3, h_1, h_2, h_3 and h_4. The ranges for each parameter are presented in Table 1. The initial storages for each tank were set as $X_1 = 0.00$ mm, $X_2 = 0.00$ mm and $X_3 = 100.00$ mm.

Table 1 Range for the tank model parameters.

Limits	Parameters										
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	h_1	h_2	h_3	h_4
Lower	0.001	0.001	0.01	0.01	0.15	0.01	0.0	10	10	10	10
Upper	0.1	0.1	0.3	0.7	0.9	0.04	0.3	90	120	120	120

The following objective function F, to be minimized, was chosen:

$$F = \sum_{t=1}^{n} \frac{|Q_o - Q_c|}{Q_o} \tag{12}$$

where Q_o and Q_c are the observed and calculated runoff (mm), respectively, and n is the number of days in the data set. As stated earlier, the period used for the calibration was from January 1992 to December 1993, while the period from January 1994 up to December 2003 was used to validate the calibrated tank model.

The correlation (r) and bias (B) statistical indexes were used as criteria for evaluating the model performance. The correlation computes the variability of a number of predictions around the true value. Different from correlation, the bias is a measure of systematic error and thus it calculates the degree to which the estimation is consistently below or above the actual value. High correlation alone does not mean high accuracy. For example, a significant constant bias in the estimations would provide the highest correlation ($r = 1$) but poor accuracy. As a result, the accuracy of estimations is better analyzed by using both bias and correlation. The perfect fit between observed and predicted values, which is unlikely to happen, would have $r = 1$ and $B = 0$. Salas (1993) provides the equations to calculate these indexes.

The DE method found the following parameter values $a_1 = 0.0658$ $a_2 = 0.093$, $a_3 = 0.042$, $a_4 = 0.013$, $b_1 = 0.175$, $b_2 = 0.043$, $b_3 = 0.007$, $h_1 = 11.562$ mm, $h_2 = 40.571$ mm, $h_3 = 64.814$ mm and $h_4 = 9.916$ mm. Figure 3 shows the comparison between observed and calculated reservoir inflows for this calibration data set ($r = 0.91$ and $B = 0.00$ mm). The optimized parameter values are used to validate the tank model using the period January 1994 up to December 2003, $r = 0.78$ and $B = -0.13$ mm, as shown in Fig. 4. These figures and indexes (high correlations and low biases) reveal that the calibrated tank model is very efficient for estimating reservoir inflows.

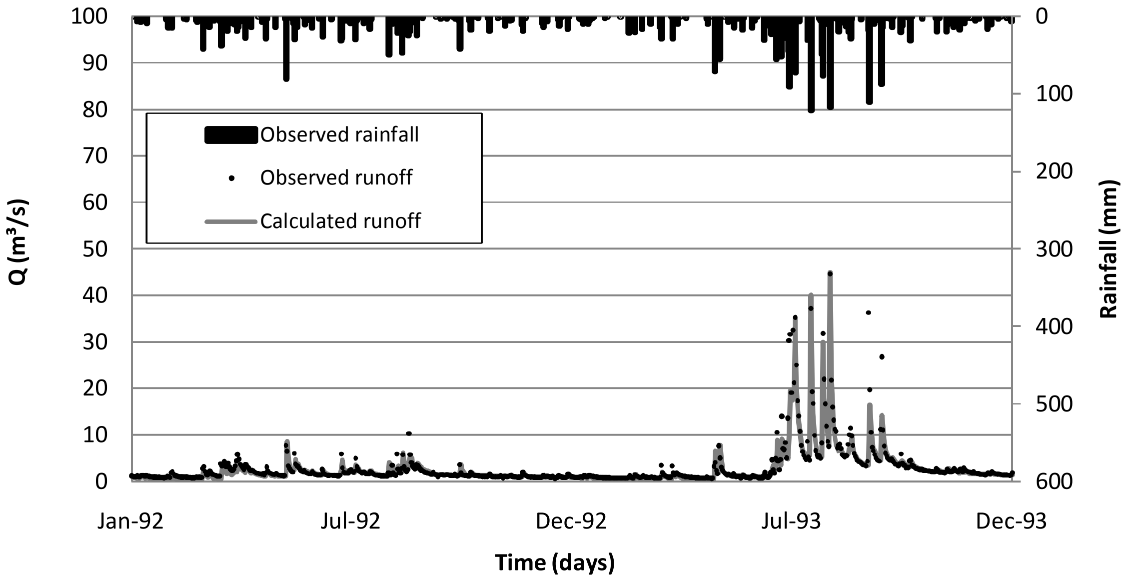

Fig. 3 Hyetograph, and observed (Q_o) and calculated (Q_c) inflows, January 1992–December 1993 (calibration).

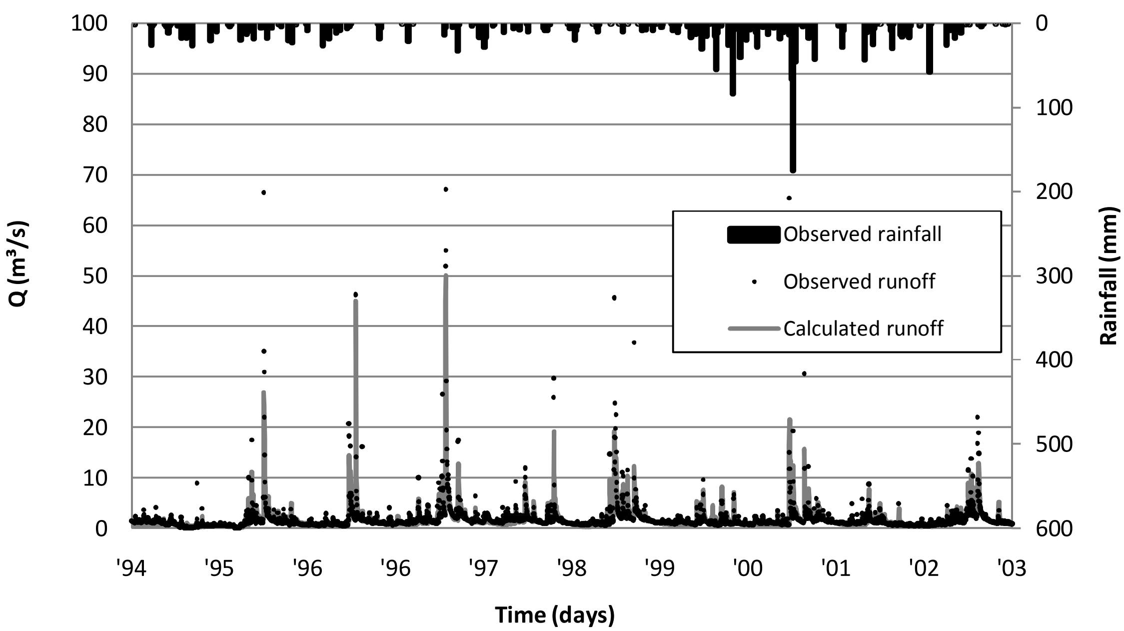

Fig. 4 Hyetograph, and observed (Q_o) and calculated (Q_c) inflows, January 1994–December 2003 (validation).

CONCLUSION

The tank model, a conceptual hydrological model, was used in order to simulate the daily runoff in Ishite River basin, Matsuyama city, Japan. It contains 11 parameters which should be set, thus global optimization method known as Differential Evolution (DE) was used to optimize such parameters. The main conclusions are as follows: (1) the tank model was shown to be useful for simulation in such a basin; (2) the DE method was proved to be robust to optimize its 11 parameters; however, in order to perform optimally, the probabilistic and deterministic components in the DE method must be chosen carefully, and as a first attempt, the following values was proposed: ITER = 10000, N = 110; NCROSS = 1; PCROS = 0.9; FACT = 0.5; IU = 1171; EPS = 10^{-8}; and (4) the found optimized parameters, which could be representative for the area, are as follows a_1 = 0.0658 a_2 = 0.093, a_3 = 0.042, a_4 = 0.013, b_1 = 0.175, b_2 = 0.043, b_3 = 0.007, h_1 = 11.562 mm, h_2 = 40.571 mm, h_3 = 64.814 mm and h_4 = 9.916 mm.

ACKNOWLEDGEMENT
The writers wish to thank Ehime University (Japan) for providing the field data, and the financial support provided by the National Council for Scientific and Technological Development (CNPq, Brazil) is gratefully acknowledged.

REFERENCES
Lee, Y. H. & Singh, V. P. (1999) Tank model using Kalman filter. *J. Hydrologic Engng* **4**(4), 344–349.

Salas, J. D. (1993) Analysis and modeling of hydrologic time series. In: *Handbook of Hydrology*, McGraw-Hill, New York, USA.

Santos, C.A.G., Srinivasan, V.S., Suzuki, K., Watanabe, M. (2003) Application of an optimization technique to a physically based erosion model. Hydrol. Processes 47, 989–1003, doi: 10.1002/hyp.1176.

Santos, C.A.G., Suzuki, K., Watanabe, M., Srinivasan, V.S. (1994) Optimization of coefficients in runoff-erosion modeling by Standardized Powell method, J. of Hydrosci. and Hydraul. Engng, **12**(1), 67–78. Japan. Soc. Civ. Engrs.

Storn, R. (1997) Differential Evolution, a simple and efficient heuristic strategy for global optimization over continuous spaces, J. Global Optimization, v. 11, Dordrecht, 341–359.