Adequacy of two-parameter beta distribution for deriving the unit hydrograph

R. K. Rai, S. Sarkar and H. G. Gundekar

ABSTRACT

In the past, to derive the unit hydrograph (UH) various distribution functions have been utilized so far, though they had their own limitations. In this study, the applicability of two-parameter beta distribution has been explored for the derivation of UH. The parameters of the distribution function were estimated using the Genetic Algorithm which facilitates the minimization of global error. The suitability of the derived UH resulting from the two-parameter beta distribution was verified by comparing the UH derived by the two-parameter gamma distribution. The results obtained from both the distributions were almost similar and close to the observed UHs, which confirmed the applicability of the two-parameter beta distribution as an alternate approach for the derivation of the UH.

Key words | beta distribution, gamma distribution, genetic algorithm, unit hydrograph, watershed

R. K. Rai (corresponding author)
State Water Resources Agency,
Department of Irrigation,
Ground Floor WALMI Bhawan, Utrethia 226 026,
Lucknow Uttar Pradesh,
India
E-mail: *rai.rkhyd@gmail.com*

S. Sarkar
H. G. Gundekar
Department of Hydrology,
Indian Institute of Technology Roorkee,
Roorkee 247 667, Uttarakhand,
India

INTRODUCTION

The nonlinear and complex nature of the watershed's rainfall–runoff process involves a number of variables in the modeling. To describe these physical processes in mathematical equations is very difficult: nevertheless, the input–output mathematical models based on the linear theory of a hydrologic system attempt to establish a causal linkage between the input and output without a detailed description of the physical process under investigation and are widely accepted theories in hydrologic modeling. The unit hydrograph (UH) concept is one such theory and perhaps the most popular modeling tool for the computation of the flood hydrograph. The unit hydrograph or unit pulse response function is an important tool for system identification in hydrologic analysis and can be used as a mathematical description of a linear system (Sherman 1932; Dooge 1959, 1973; Singh 1988; etc.). Since the shape of the probability density function (PDF) is similar to the conventional instantaneous unit hydrograph (IUH), therefore, many attempts have been made in the past to use the PDFs of gamma and three-parameter beta distribution function for the derivation of IUH and UH (Gray 1961; Croley 1980; Haktanir & Sezen 1990; Singh 2000; Yue *et al.* 2002; Bhunya *et al.* 2003 etc.). However, the application of two-parameter beta distribution function still remains unexplored for the derivation of a UH. Furthermore, application of a PDF for the derivation of a UH requires fitting of their parameters. To that end, several approaches are available, such as, for example, least squares (LS), method of moment (MoM), maximum likelihood (ML), etc. (Singh 1988). However, the traditional optimization methods have limitations in finding global optimization results as they search from point to point for optimization. On the other hand, the genetic algorithm (GA) searches the entire population instead of moving from one point to the next and can overcome the limitation of the traditional method (Goldberg 1989; Kuo *et al.* 2000). Goldberg (1989) further stated that a GA is a search procedure that uses random choice as a tool to direct a highly exploitative search through the numerical coding of a given parameter space. GAs differ from conventional optimization and search procedures as: (i) GAs work with a coding of the parameter set, not the parameters themselves, (ii) GAs search from a population of points, not a single point, (iii) GAs use objective function formulation, not derivatives or other auxiliary knowledge and (iv) GAs uses probabilistic

doi: 10.2166/nh.2008.038

transition rules, not deterministic rules. Furthermore, Jain et al. (2005) reported that GA performs best in the estimation of the hydrograph parameters such as peak discharge, time to peak and volume of DRH. In recent years, many applications of GA have been reported in hydrologic modeling (Wardlaw & Sharif 1999; Kuo et al. 2000; Jain et al. 2005; Rabuñal et al. 2007; etc.). Looking into the above facts, in the present study, the following objectives have been set: (1) check the adequacy of the two-parameter beta distribution for UH derivation, (2) parameter estimation of using the GA, and (3) study the comparative performance of two-parameter beta distribution and gamma distribution in the reproduction of UH and DRH.

DISTRIBUTION FUNCTION AND IUH SHAPE

The IUH satisfies all the properties of a PDF. The two shape variables, namely shape mean and shape variance, are defined to represent the shape of the IUH. The shape mean (S_m) of the IUH is defined as the centroid of the IUH that represents the central tendency of a hydrograph, whereas shape variance (S_v) is the variance of a hydrograph that represents the spreadness of the IUH. These shape parameters S_m and S_v of the IUH can be estimated using the following formulae:

$$S_m = \sum_{i=1}^{n} t_i \cdot f(t_i) \tag{1a}$$

$$S_v = \sum_{i=1}^{n} (t_i - S_m)^2 \cdot f(t_i) \tag{1b}$$

On the other hand, the shape parameters of the PDF can be given as follows:

$$\mu = \int_{-\infty}^{+\infty} t f(t) \mathrm{d}t \tag{2a}$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (t - \mu)^2 f(t) \mathrm{d}t \tag{2b}$$

where $f(t)$ = probability density function (PDF) of a random variable t. Based on Equations (1) and (2), the shape of the IUH can be interpreted as a PDF. Thus the PDF can be used to represent the shape of the IUH.

As stated in the objective, the adequacy of the two-parameter beta distribution for the derivation of the IUH is explored and its performance was further checked by comparing the results with the gamma distribution. To start with, the explanation of the two-parameter beta distribution has been taken up first.

Two-parameter beta distribution

The pdf of the two-parameter beta distribution function is given as follows (Johnson & Kotz 1970):

$$h(t) = \frac{1}{B(a,b)} t^{a-1} \cdot (1 - t)^{b-1}, \quad (0 \le t \le 1) \tag{3}$$

where $B(\cdot)$is the beta function, and a and b are the parameters which define the shape of the IUH. The value of t must lie in the interval of [0, 1]. The beta function is defined as follows:

$$B(a,b) = \int_0^1 t^{a-1}(1 - t)^{b-1} \mathrm{d}t \tag{4}$$

The cumulative beta distribution function is given by the following relationship:

$$g(t|a,b) = \frac{1}{B(a,b)} \int_0^t t^{a-1}(1 - t)^{b-1} \mathrm{d}t \tag{5}$$

Equation (5) gives the step response function based on the two parameters.

The parameters of the beta PDF can be estimated using the MoM (Johnson & Kotz 1970):

$$\mu = a/(a + b) \tag{6}$$

$$\sigma^2 = \frac{ab}{(a + b)^2(a + b + 1)} \tag{7}$$

Therefore, given the aforementioned shape variable $S_m = \mu$ and $S_v = \sigma^2$, the parameters of the two-parameter beta distribution can be derived and therefore the shape of the UH can be represented by the PDF of the two-parameter beta distribution.

Since applicability of Equation (5) is within the range of $0 \le t \le 1$, therefore, for computational purposes, the time t has been normalized by dividing it by t_B/t_r, in which t_B is the timebase of the UH and t_r is the excess rainfall duration.

Two-parameter gamma distribution

Nash (1959) and Dooge (1959) proposed the mathematical derivation of IUH (i.e. unit impulse response function) in the form of a gamma function utilizing the concept of n-linear reservoirs of equal storage coefficient K as follows:

$$h(t) = \frac{1}{K\Gamma(n)}\left(\frac{t}{K}\right)^{n-1}\exp(-t/K) \tag{8}$$

In Equation (8), n and K define the shape of the IUH, $h(t)$ is the flow rate [T^{-1}] and $\Gamma(n)$ is the gamma function of n. The dimension of K is similar to the time [T]. For deriving the UH of a desired duration, the step response function is derived which is analogous to the S-hydrograph in surface hydrology. The step response is computed from the following equation:

$$g(t) = \int_0^t h(t)\mathrm{d}t = \int_0^t \frac{1}{K\Gamma(n)}\left(\frac{t}{K}\right)^{n-1}\exp(-t/K)\mathrm{d}t$$

$$= \frac{1}{\Gamma(n)}\int_0^{t/K}\left(\frac{t}{K}\right)^{n-1}\exp\left(-\frac{t}{K}\right)d\left(\frac{t}{K}\right) \tag{9}$$

where $g(t)$ is the step response function used to compute the UH of a desired duration.

As the step response $g(t)$has been computed, the Δt hour-UH is computed from the following relationship:

$$u(\Delta t, t) = \frac{10.0A}{3.6\Delta t}\times\{g(t) - g(t-1)\} \tag{10}$$

where u(Δt, t) is the UH of Δt duration and A is the drainage area (km^2). The following discrete form of convolution expression is used for the computation of the direct runoff hydrograph (DRH):

$$Q(t) = \sum_{j=1}^{t} u(\Delta t, t-j+1)\cdot r_e(j) \tag{11}$$

In the above relationship (Equation (11)), r_e is the excess rainfall depth (cm) and $Q(t)$ is the ordinates of the DRH (m^3/s) at a time step of t.

OPTIMIZATION USING GA

Optimization of the parameters of the distribution function used was carried out using the GA. The procedural steps of the GA are summarized as follows. The algorithm begins by creating a random initial population. The algorithm then creates a sequence of new populations, or generations. At each step, the algorithm uses the individuals in the current generation to create the next generation. To create the new generation, the algorithm performs the following steps: scores each member of the current population by computing its fitness value; scales the raw fitness scores to convert them into a more usable range of values; selected parents based on their fitness produce children from the parents either by making random changes to a single parent (i.e. mutation) or by combining the vector entries of a pair of parents (i.e. crossover). The next generation is formed by replacing the current population with the children. The algorithm stops when one of the stopping criteria is met.

Initial population

The algorithm begins by creating a random initial population (chromosomes). These chromosomes are evaluated based on their performances (fitness value) in terms of certain objective functions. Then the chromosomes compete for their survival in a tournament, in which one parent is selected as having the best fitness value among two or more randomly picked chromosomes. A second parent is selected by repeating the same process. This selection process is called natural selection. The chromosomes with optimum fitness are used to create the next generation while the remaining one dies off.

Creating the next generation

The GA creates three types of children for the next generation: (i) elite children are the individuals in the current generation with the best fitness values: these individuals automatically survive to the next generation, (ii) crossover children are created by combining the vectors of a pair of parents and (iii) mutation children are created by introducing random changes, or mutations, to a single parent.

Mutation and crossover

The GA uses the individuals in the current generation to create the children that make up the next generation. Besides elite children, which correspond to the individuals in the current generation with the best fitness values, the algorithm creates crossover children by selecting vector entries, or genes, from a pair of individuals in the current generation and combines them to form a child, whereas mutation performs the creation of children by applying random changes to a single individual in the current generation. Both processes are essential to the GA. Crossover enables the algorithm to extract the best genes from different individuals and recombine them into potentially superior children. Mutation adds to the diversity of a population and thereby increases the likelihood that the algorithm will generate individuals with better fitness values. Without mutation, the algorithm could only produce individuals whose genes were a subset of the combined genes in the initial population.

Stopping criteria for the algorithm

The genetic algorithm uses the following five stopping criteria: (i) generation criteria: the algorithm stops when the number of generations reaches the value of generations; (ii) time limit criteria: the algorithm stops after running for an amount of time in seconds equal to the time limit; (iii) fitness criteria: the algorithm stops when the value of the fitness function for the best point in the current population is less than or equal to the fitness limit; (iv) stall generations criteria: the algorithm stops if there is no improvement in the objective function for a sequence of consecutive generations of length stall generations; and (v) stall time limit criteria: the algorithm stops if there is no improvement in the objective function during an interval of time in seconds equal to the stall time limit. The algorithm stops as soon as any one of these five conditions is met.

Objective function

The GA needs to define an objective function for optimizing the parameters. The objective function used here is given as follows:

$$\min f(\cdot) = \sum_{i=1}^{t_B} \{w(i) \times |\hat{Q}(i) - Q(i)|\} \tag{12}$$

where $f(\cdot)$ is the function of the parameters, $\hat{Q}(i)$ and $Q(i)$ are the computed and observed direct runoff, respectively, and $w(i)$ is defined as follows:

$$w(i) = Q(i) / \sum_{i=1}^{t_B} Q(i) \tag{13}$$

To perform the optimization using GAs, the program was developed in the MATLAB platform.

RESULTS AND DISCUSSION

To test the adequacy of the two-parameter beta distribution, the following data are considered: (1) data having the UH derived from the methods proposed by Haktanir & Sazen (1990) and (2) the DRH data reported by Gaur and Mathur (2003).

In totality, six storm events have been picked to test the applicability of the methodology. Out of six events, two observed UH of different duration were taken from the Anatolia river basin reported by Haktanir & Sazen (1990) and four storm events were picked up from the four Indian watersheds reported by Gaur & Mathur (2003). The brief description of these watersheds is given in Table 1. To start with, the approach was first applied for the derivation of the UH using the data of two sub-basins of the Anatolia river basin, namely the Hamam and Inderesi watersheds.

Case I

Using GAs, the parameters of the two-parameter beta distribution (i.e. a and b) were estimated using Equations (3)–(5) and (10)–(13) for the Hamam and Inderesi watersheds and used to derive the UH. The derived UH along with the observed UH are shown in Figures 1 and 2. Further, the derived UH from the beta distribution function was compared with the UH obtained by the gamma distribution function as well as that reported by Haktanir & Sazen (1990). The parameters of the gamma distribution (i.e. n and K) were also estimated by the GA using Equations (8)–(13). It is clear from Figures 1 and 2 that the two-parameter beta distribution is equally good in deriving the UHs for the Hamam and Inderesi watersheds when compared with the gamma distribution with GAs. Furthermore, it is evident from Figures 1 and 2 that the GA works well in the reproduction of peak flow rate and time to

Table 1 | Description of watersheds and event information

S. no.	Watershed description	(A) Haktanir & Sazen (1990)		(B) Gaur & Mathur (2003)			
		Hamam, Goksu River	Inderesi, Inderesi Creek	W-8	Jhandoo Nala	B-319	B-719
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
1	Drainage area (km^2)	4300.0	141.5	0.012	0.177	0.82	14.0
2	Slope (%)	0.83	1.18	1.90	48.8	9.2	6.8
3	Length of main channel (m)	16.1×10^4	23.5×10^3	177.6	900.0	1650.0	7200.0
4	Event date	–	–	23/08/1995	08/08/1991	05/08/1964	11/08/1965
5	Excess rainfall duration (min)	480.0	120.0	20.0	10.0	50.0	60.0
6	Time to peak (min)	1440.0	360.0	15.0	20.0	50.0	150.0
7	Peak flow rate (m^3/s)	417.0	54.0	15.0	6.83	16.03	10.03
8	Base time (h)	100.0	28.0	1.42	1.167	2.67	6.5

peak when compared with the results reported by Haktanir & Sazen (1990). Haktanir & Sazen (1990) used the Newton iterative algorithm for the estimation of the parameters for the gamma distribution. Besides the visual comparison of the results, the comparative performance of both the approaches was assessed by using the following statistical measures.

(i) Coefficient of efficiency (CE) (Nash & Sutcliff 1970):

$$CE = 1 - \left\{ \frac{\sum_{t=1}^{n} [\hat{Q}(t) - Q(t)]^2}{\sum_{t=1}^{n} [Q(t) - \bar{Q}]^2} \right\} \tag{14}$$

where $\hat{Q}(t)$ and $Q(t)$ are the computed and observed UH ordinates, respectively, $\bar{Q}$is the mean UH and n is the total number of ordinates.

(ii) Mean absolute percent error (MAPE) (Rai *et al.* 2007):

$$MAPE = \frac{1}{n}\left(100 \times \sum_{i=1}^{n} ABS(\hat{Q}_i - Q_i)\right) \tag{15}$$

The estimated values of the CE and MAPE for the two-parameter beta and gamma distribution functions are given in Table 2. The values in the parenthesis (Table 2) are based on the results reported by Haktanir & Sazen (1990). The values of CE and MAPE for both the approaches again confirm the results reported in Figures 1 and 2, which suggests the usability of the two-parameter beta distribution function for the derivation of the UH. In addition to this, the characteristic UH parameters, viz., peak discharge, time to peak and volume, were very well preserved by both the approaches; conversely, the recession segment computed by both distributions was

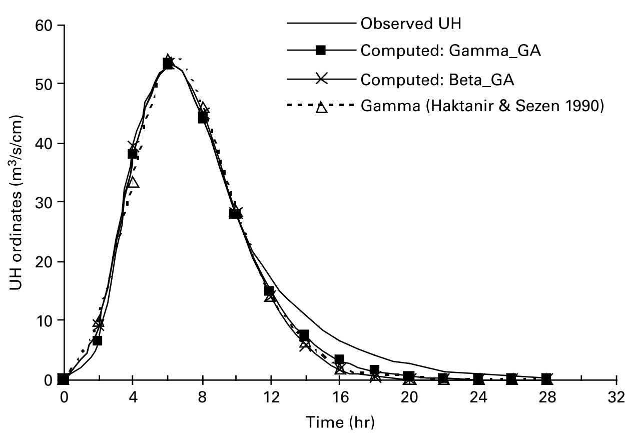

Figure 1 | Comparison of observed and computed UH for Hamam watershed.

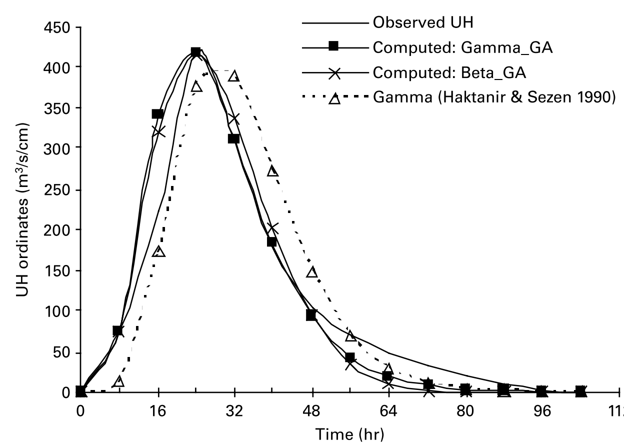

Figure 2 | Comparison of observed and computed UH for Inderesi watershed.

Table 2 | Parameters of the two-parameter beta and gamma distribution functions along with the estimated values of CE and MAPE

Watershed	Two-parameter beta				Two-parameter gamma			
	a	*b*	CE	MAPE	*n*	*K*	CE	MAPE
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(vii)	(ix)
Hamam	3.2478	10.794	0.9313	2107.5	3.8119 (5.6)	6.3882 (5.233)	0.9227 (0.8817)	1772.8 (3108.6)
Inderesi	3.2722	11.403	0.981	167.23	4.0991 (5.4)	1.6015 (1.358)	0.9888 (0.9782)	132.16 (199.53)
W-8	0.7883	3.8089	0.9346	0.2302	1.01	0.2602	0.9450	0.2265
Jhandoo	2.3618	6.5872	0.9361	1.925	3.4042	0.0975	0.9582	1.512
B-319	2.7915	9.5666	0.893	25.96	3.9815	0.1529	0.891	25.66
B-719	3.0148	4.7285	0.979	101.54	5.4986	0.4750	0.981	138.98

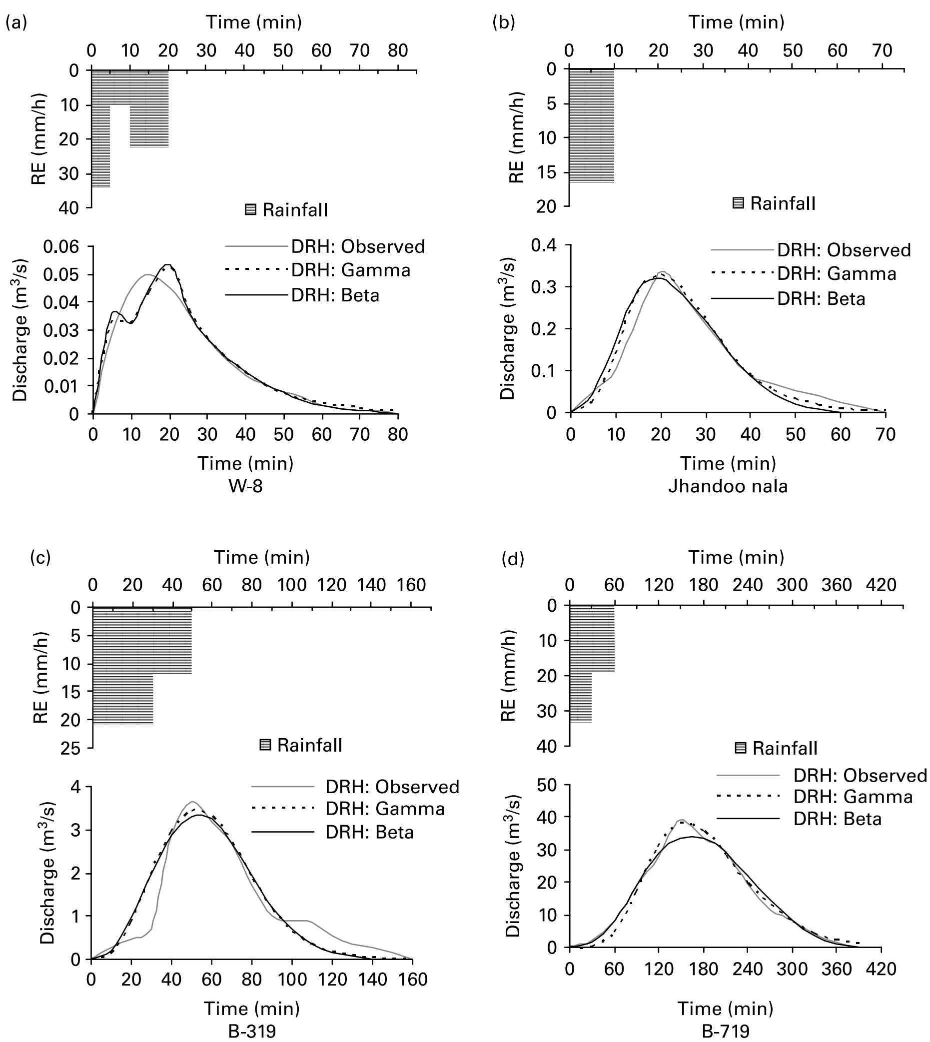

Figure 3 | Comparison of observed and computed DRHs for the four Indian watersheds.

under-predicted, which might be due to the characteristic shape of the applied distributions.

Case II

Now, so far, it is clear from case I that the two-parameter beta distribution is equally good for the UH derivation and, therefore, used in the estimation of the DRH. For this purpose, the excess rainfall-DRH data of four Indian watersheds, namely, W-8, Jhandoo Nala, B-319 and B-719 from Gaur & Mathur (2003), were used. Using the excess rainfall-DRH data, the parameters of the beta and gamma distribution functions were estimated using GAs (Table 2) and used to derive the corresponding UH. The resulting UH were further used to compute the DRH using the convolution technique. The ordinates of the DRH computed by using the beta and gamma distribution functions for the four watersheds along with the observed one are depicted in Figure 3. Besides this, the CE and MAPE were also estimated (Table 2). Based on Figure 3 and estimated values of statistical measures, it can be observed that the DRH computed by the beta distribution was more or less similar to the results obtained by the gamma distribution.

CONCLUSIONS

The present study analyzed the adequacy of the two-parameter beta distribution for the derivation of the UH and, in turn, the DRH. The study also presents the application of the GA for the estimation of the distribution parameters. The approach was verified on the two watersheds having the UH and four watersheds having the excess rainfall-DRH data. Further, the beta distribution was compared with the gamma distribution in the reproduction of the UH and DRH. Based on these findings, it is concluded that the fitted two-parameter beta distribution not only shows the proximity with the observed UH but also it gives closer agreement in the computation of the DRH. Based on the visual (Figures 1–3) and statistical criteria (Table 2), it is evident that the two-parameter beta distribution is equally good when compared with the gamma distribution. As far as the GA is concerned, the parameter estimation using GAs gives better reproduction of the hydrograph parameters, viz. peak discharge and time to peak. Finally, it may be concluded that the two-parameter beta distribution can be used as one of the alternate approaches for derivation of the UH.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their constructive suggestions and comments in improving the quality of the manuscript.

REFERENCES

Bhunya, P. K., Mishra, S. K. & Berndtsson, R. 2003 Simplified two-parameter Gamma distribution for derivation of synthetic unit hydrograph. *J. Hydrol. Eng. ASCE* **8**(4), 226–230.

Croley, T. E. 1980 Gamma synthetic hydrographs. *J. Hydrol.* **47**, 41–52.

Dooge, J. C. 1959 A general theory of the unit hydrograph. *J. Geophys. Res.* **64**(2), 241–256.

Dooge, J. C. 1973 *Linear Theory of Hydrologic Systems*. Technical Bulletin No. 1468, US Department of Agriculture, Agricultural Research Service, Washington, DC.

Gaur, M. L. & Mathur, B. S. 2003 Modeling event-based temporal variability of flow resistance coefficient. *J. Hydrol. Eng. ASCE* **8**(5), 266–277.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Learning*. Addison-Wesley, Reading, MA.

Gray, D. M. 1961 Synthetic hydrographs for small drainage areas. *J. Hydraul. Div. ASCE* **87**(4), 33–54.

Haktanir, T. & Sezen, N. 1990 Suitability of two-parameter Gamma and three-parameter Beta distribution as synthetic hydrographs in Anatolia. *Hydrol. Sci. J.* **35**(2), 167–184.

Jain, A., Srinivasalu, S. & Bhattacharjya, R. K. 2005 Determination of an optimal unit pulse response function using real-coded genetic algorithm. *J. Hydrol.* **303**, 199–214.

Johnson, N. L. & Kotz, S. 1970 *Distribution In Statistics: Continuous Univariate Distribution II*. Wiley, New York, NY.

Kuo, S.-F., Merkley, G. P. & Liu, C.-W. 2000 Decision support for irrigation project planning using a genetic algorithm. *Agric. Wat. Manage.* **45**, 243–266.

Nash, J. E. 1957 The forms of the instantaneous unit hydrograph. *Int. Assoc. Sci. Hydraul. Div. Proc. ASCE* **104**(HY2), 262–276.

Nash, J. E. & Sutcliff, J. V. 1970 River flow forecasting through conceptual models Part 3: A discussion of the principle. *J. Hydrol.* **10**, 282–290.

Rabuñal, J. R., Puertas, J., Suárez, J. & Rivero, D. 2007 Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. *Hydrol. Process.* **21**, 476–485.

Rai, R. K., Jain, M. K., Mishra, S. K., Ojha, C. S. P. & Singh, V. P. 2007 Another look at Z-transform technique for deriving unit impulse response function. *Wat. Res. Manage.* **21**(11), 1829–1848.

Sherman, L. K. 1932 Stream flow from rainfall by the unit-graph method. *Eng. News Rec.* **108**, 501–505.

Singh, S. K. 2000 Transmuting synthetic unit hydrographs into Gamma distribution. *J. Hydrol. Eng. ASCE* **5**(4), 380–385.

Singh, V. P. 1988 *Hydrologic Systems*, (vol. 1). Prentice-Hall, Englewood Cliffs, NJ.

Wardlaw, R. & Sharif, M. 1999 Evaluation of genetic algorithms for optimal reservoir system operation. *J. Wat. Res. Plann. Manage.* **125**(1), 25–33.

Yue, S., Taha, B. M. J., Bobee, B., Legendre, P. & Bruneau, P. 2002 Approach for describing statistical properties of flood hydrograph. *J. Hydrol. Eng. ASCE* **7**(2), 147–153.

First received 3 July 2007; accepted in revised form 1 November 2007