

Optimal design of unit hydrographs using probability distribution and genetic algorithms

RAJIB KUMAR BHATTACHARJYA

Department of Civil Engineering, National Institute of Technology,
Silchar 788 010, India
e-mail: rajibkb@nits.ac.in, rajibkb@alumni.iitk.ac.in

MS received 5 August 2003; revised 31 May 2004

Abstract. A nonlinear optimization model is developed to transmute a unit hydrograph into a probability distribution function (PDF). The objective function is to minimize the sum of the square of the deviation between predicted and actual direct runoff hydrograph of a watershed. The predicted runoff hydrograph is estimated by using a PDF. In a unit hydrograph, the depth of rainfall excess must be unity and the ordinates must be positive. Incorporation of a PDF ensures that the depth of rainfall excess for the unit hydrograph is unity, and the ordinates are also positive. Unit hydrograph ordinates are in terms of intensity of rainfall excess on a discharge per unit catchment area basis, the unit area thus representing the unit rainfall excess. The proposed method does not have any constraint. The nonlinear optimization formulation is solved using binary-coded genetic algorithms. The number of variables to be estimated by optimization is the same as the number of probability distribution parameters; gamma and log-normal probability distributions are used. The existing nonlinear programming model for obtaining optimal unit hydrograph has also been solved using genetic algorithms, where the constrained nonlinear optimization problem is converted to an unconstrained problem using penalty parameter approach. The results obtained are compared with those obtained by the earlier LP model and are fairly similar.

Keywords. Unit hydrograph; rainfall-runoff; hydrology; genetic algorithms; optimization; probability distribution.

1. Introduction

One of the most common interests of hydrologists is the estimation of direct runoff from a watershed for specified distribution of rainfall. This can be achieved either by a system or a physical approach. The system approach is a lump approach and the spatial variation of rainfall and infiltration over the watershed is ignored. On the other hand, the physical approach considers spatial variation of rainfall and infiltration. In the lump approach, the

A list of symbols is given at the end of the paper

transformation of rainfall into direct surface runoff can be modelled mathematically using a transfer function. A unit hydrograph is the most widely used and accepted transfer function to model the rainfall-runoff process.

Unit hydrograph of a watershed under the assumption of linear system may be defined as the unit pulse response function of a linear system. The transfer function, unit hydrograph, of the linear hydrologic system follows two basic principles of linear system operation, i.e. the principle of proportionality and principle of superposition (Chow *et al* 1988). A unit hydrograph is defined as the direct runoff hydrograph resulting from one centimetre (or one inch) of rainfall excess uniformly distributed over a watershed at a constant rate for an effective duration (Sherman 1932). The main use of the unit hydrograph (UH) is to determine the direct runoff hydrograph (DRH) from the effective rainfall hyetograph (ERH). In this paper, both UH and DRH are considered on a unit watershed area basis. The discrete convolution relationship between the DRH, UH and ERH can be written as (Chow *et al* 1988)

$$Q_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1}, \quad (1)$$

where, P_m is the m th ordinate of the ERH, Q_n is the n th ordinate of the DRH, and U_{n-m+1} is the $(n - m + 1)$ th ordinate of UH. The total number of UH ordinates is $N - M + 1$, where M is the number of ERH ordinates, and N is the number of DRH ordinates. Knowing ERH and DRH, the corresponding UH can be obtained from (1). This process is known as deconvolution. On the other hand, knowing ERH and UH, one can determine the DRH of the watershed by using (1).

Many researchers derive unit hydrographs by using convolution relationships. These include the method of successive approximations (Collins 1939; Barnes 1959; Bender & Roberson 1961), method of least squares (Singh 1976; Bruen & Dooge 1984), optimization methods such as linear programming (Deinger 1969; Mays and Coles 1980; Morel-Seytoux 1982; Singh 1988; Prasad *et al* 1999) and nonlinear programming (Mays & Taur 1982). The main disadvantage of these methods is that the number of unknowns is equal to the number of unit hydrograph ordinates. Therefore, for larger time bases these methods may involve difficulties in estimating the unit hydrograph from the rainfall-runoff data as the number of unknowns is large.

This study presents a genetic algorithm-based optimization model to transmute unit hydrographs into probability distribution functions. Gamma and log-normal probability distributions are used in this study. The objective function is to minimize the sum of the squares of the deviation between predicted and actual direct runoff hydrographs per unit area of the watershed. The predicted direct runoff hydrograph per unit watershed area is estimated using a probability distribution function. From the principles of unit hydrograph drawn on a unit area basis, the depth of rainfall excess must be unity and the ordinates of the unit hydrograph must be positive. The incorporation of probability distribution function ensures that the depth of rainfall excess, and the area under the curve is unity, and also that the ordinates of the unit hydrograph are always positive. Therefore, the proposed methodology does not have any constraint, and the unconstrained nonlinear optimization formulation is solved using binary-coded genetic algorithms. The advantage of this method is that the number of unknowns to be determined by optimization is equal to the number of probability distribution parameters. It is also not dependent on the number of direct runoff hydrograph ordinates. A comparison with the previously published models is also presented.

2. Probability distribution

The probability distribution functions considered in this study are log-normal and gamma distributions.

2.1 Gamma distribution

A random variable x is said to follow gamma distribution if its probability density function (PDF) is given by

$$f(x) = (\lambda / \Gamma(\eta)) (\lambda x)^{\eta-1} e^{-\lambda x}, \quad \text{for } x \geq 0, \quad (2)$$

where $\lambda > 0$ and $\eta > 0$ are the parameters of the distribution, which is also known as scale and shape parameter respectively, and $\Gamma(\eta)$ is the gamma function given by

$$\Gamma(\eta) = \int_0^{\infty} x^{\eta-1} e^{-x} dx. \quad (3)$$

When η is an integer, then $\Gamma(\eta) = (\eta - 1)!$. When η is not an integer, then the gamma function has to be evaluated by numerical integration. In this paper, the gamma function is calculated by Simpson's method of numerical integration.

2.2 Log-normal distribution

A random variable x is said to be log-normally distributed if its probability density function is given by

$$f(x) = (1/x\sigma_n\sqrt{2\pi}) \exp[-(\ln x - \mu_n)^2/2\sigma_n^2], \quad \text{for } x \geq 0, \quad (4)$$

where μ_n and σ_n are the parameters of distribution, also known as scale and shape parameters respectively.

3. Genetic algorithms

The study of genetic algorithms originated in the mid-1970's from John H Holland (Holland 1975), and has become a powerful and robust tool for function optimization. It is a search technique based on the concept of natural selection inherent in natural genetics. Genetic algorithms combine the "survival of the fittest" principle with genetic operators abstracted from Nature. One of the advantages of genetic algorithms is that it does not require continuity or differentiability of either the objective function or the constraints (Deb 2001). The main difference of genetic algorithms with classical methods is that genetic algorithms work with coding of the parameter set, and not the parameters themselves and it starts from a population of points rather than a single point (Goldberg 1989).

Genetic algorithms start with randomly generating an initial population (p) of possible solutions. Each solution is known as a "chromosome". These chromosomes are evaluated based on their performance with respect to the objective function. The population is then operated by three basic operators in order to produce better offspring for the next generation. These operators are known as "reproduction", "crossover" and "mutation". Reproduction is

Table 1. Genetic algorithm parameters.

String length (l)	20* (no. of decision variables)
Population size (p)	1.5 * (l)
Crossover probability (p_c)	0.85
Mutation probability (p_m)	1/ l
Generation (g)	400
Elitism size (E)	5% of p

a process in which individual strings are copied according to their fitness (Goldberg 1989). Crossover is considered the partial exchange of corresponding segments between two parent strings to produce two offspring strings. The genetic algorithm picks up two strings from the population to perform crossover with probability p_c at a randomly selected point along the string. Mutation is the occasional introduction of new features into the population pool to maintain diversity in the population. Random bit-by-bit mutation is used in this study to generate new solutions. These genetic operators are designed in such a way that they produce better solutions after each generation. Sometimes, crossover and mutation may produce inferior solutions, but these solutions die out when passed through the selection operator the next time. Moreover, to increase the efficiency of the genetic algorithms, it is necessary to preserve some better solutions for the next generation without applying crossover or mutation operators. This process is known as elitism. Generally, the better E percent of the total population is preserved for the next generation without applying crossover and mutation operators. Elitism is important since it allows the solutions to get better over time. These procedures would continue till the better individual is better enough to suit the objective. The genetic algorithms parameters used in this study are listed in table 1.

4. Optimization formulation

4.1 Model I. Existing nonlinear optimization formulation

The main objective is to minimize the sum of the squares of deviation between the actual and predicted direct runoff hydrographs when drawn on a unit watershed area basis. Constraints are that the area under the unit hydrograph must be unity, and the UH ordinates must be positive. This may be written as

$$\text{Minimize } \sum_{n=1}^N e_n^2, \tag{5}$$

subject to

$$\left(1.0 - \Delta t \sum_{r=1}^{N-M+1} U_r \right) = 0.0, \tag{6}$$

$$U_r \geq 0 \text{ where } r = 1, 2, 3, \dots, N - M + 1, \tag{7}$$

where,

$$e_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1} - Q_n', \tag{8}$$

e_n is the deviation between the n th ordinate of the predicted and of the actual direct runoff hydrograph, Q'_n is the n th ordinate of the actual direct runoff hydrograph.

Equality constraints are difficult to handle using genetic algorithms. Therefore, the equality constraint, (6), is converted to an inequality constraint by imposing some allowable error (ε). This may be written as,

$$\varepsilon - \left(1.0 - \Delta t \sum_{r=1}^{N-M+1} U_r \right) \geq 0.0. \tag{9}$$

Here, the number of unknowns to be determined is equal to the number of unit hydrograph ordinates, i.e. $(N - M + 1)$. The constrained optimization problem is converted to an unconstrained one by penalty parameter approach. The unconstrained optimization problem is solved using binary-coded genetic algorithms.

4.2 Modified formulation to transmute UH into probability distribution

The main disadvantage of model I is that the number of unknowns are equal to the number of unit hydrograph ordinates, which is equal to $(N - M + 1)$. A modified formulation is presented here to transmute unit hydrographs into probability distribution functions. In this case, the number of unknowns is equal to the probability distribution parameters. Gamma and log-normal distributions are used in this model.

4.2a Model II. Transmute UH into gamma distribution: The optimization formulation to transmute a unit hydrograph into gamma distribution may be written as,

$$\text{Minimize } \sum_{n=1}^N e_n^2 \tag{10}$$

where

$$e_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1} - Q'_n, \tag{11}$$

$$U_{n-m+1} = f(x), \tag{12}$$

where $x = (n - m + 1) * \Delta t$, $f(x)$ is gamma probability distribution function and is written as

$$f(x) = [\lambda / \Gamma(\eta)] (\lambda x)^{\eta-1} \exp(-\lambda x)$$

4.2b Model III. Transmute UH into log-normal distribution: The optimization formulation to transmute a unit hydrograph into log-normal distribution may be written as,

$$\text{Minimize } \sum_{n=1}^N e_n^2 \tag{13}$$

where,

$$e_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1} - Q'_n, \tag{14}$$

$$U_{n-m+1} = f(x), \tag{15}$$

where $x = (n - m + 1) * \Delta t$, $f(x)$ is the log-normal probability distribution function and is written as

$$f(x) = (1/x\sigma_n\sqrt{2\pi}) \exp[-(\ln x - \mu_n)^2/2\sigma_n^2].$$

In model I and II, the number of unknowns, i.e. the probability distribution parameters is only two. The probability distribution parameters are η and λ in case of Gamma distribution, and σ_n^2 and μ_n in case of Log-normal distribution. The incorporation of probability distribution function ensures that the area under the probability distribution function is always unity and the ordinates are always positive. Therefore, this formulation does not have any constraint. The unconstrained optimization formulation is solved by using binary-coded genetic algorithms to estimate the probability distribution parameters.

5. Application of the model

Data derived from the North Potomac River near Cumberland, Maryland, USA corresponding to two separate storms (Singh 1976; Mays & Coles 1980) are used to evaluate the applicability of the proposed method. The first storm was of 24 hours duration between April 24 and 26, 1937 and the second was of 8 hours duration on March 15–16, 1942. These two datasets are given in table 2. These two sets of storm data are used to derive a 4-hour unit hydrograph for the above mentioned catchment.

Model I is an inverse problem. It optimizes the ordinate of the unit hydrograph by minimizing the difference between the predicted and actual direct runoff hydrograph. The number of variables to be estimated by optimization are equal to ordinates of the unit hydrograph,

Table 2. Storms for the North branch of the Potomac River, Cumberland, MD.

Dataset I (Storm A) April 24–26, 1937		Dataset II (Storm B) March 15–16, 1942	
<i>P</i> (inches)	<i>Q</i> (inches/hour)	<i>P</i> (inches)	<i>Q</i> (inches/hour)
0.36	0.001	0.12	0.0005
0.84	0.005	0.48	0.0017
0.92	0.023		0.0049
0.04	0.052		0.0151
0.12	0.083		0.0256
0.24	0.088		0.0238
	0.081		0.0174
	0.070		0.0136
	0.061		0.0111
	0.047		0.0091
	0.037		0.0073
	0.028		0.0060
	0.020		0.0046
	0.014		0.0035
	0.009		0.0024
	0.005		0.0016
	0.003		0.0011

Note: 1 inches = 25.4 mm; 1 inches/hour = 25.4 mm/h

Table 3. Probability distribution parameters.

Dataset	Model II (gamma distribution)		Model III (log-normal distribution)	
	λ	η	μ_n	σ_n^2
	4.2190	0.1740	3.2350	0.2590
	3.5635	0.1306	3.1280	0.2612

which is 12 for dataset I and 16 for dataset II. Model II and model III are also an inverse problem which optimizes the probability distribution parameter by minimizing the difference between the actual and predicted direct runoff hydrograph. In this case, two parameter gamma and log-normal probability distributions are used for models II and III respectively. Thus, the optimization problem becomes much smaller than the LP model (Mays & Coles 1980) and model I, as it needs to optimize only two variables. Table 3 shows the probability distribution parameters for both the datasets. The probability distribution parameters are optimized using binary-coded genetic algorithms.

The unit hydrographs derived by different models are shown in figure 1 for dataset I and in figure 2 for dataset II. The resulting direct runoff hydrographs are shown in figures 3 and 4 for dataset I and II respectively. Figures 1 and 2 show that the unit hydrograph obtained from model I is similar to that obtained by the LP model (Mays & Coles 1980). Models I, II, and III do not show any tail oscillations. The UH produced by the LP model has waves on the tail end of the unit hydrograph, and tail oscillation is not physically acceptable. Mays & Coles (1980) stated that the fluctuation at the tail of the unit hydrograph can be eliminated using some constraints in the LP model. They did not consider it as it would increase the computational complexity of the model. Therefore, models I–III are definitely better than the LP model, as monotonic declination at the tail of the unit hydrograph is achieved without imposing any constraints.

All the models have estimated the time-to-peak more or less perfectly for both the datasets. The unit hydrograph transmuted into gamma distribution (model II) failed to estimate the peak discharge for both the datasets. On the other hand, the log-normal distribution (model III) is

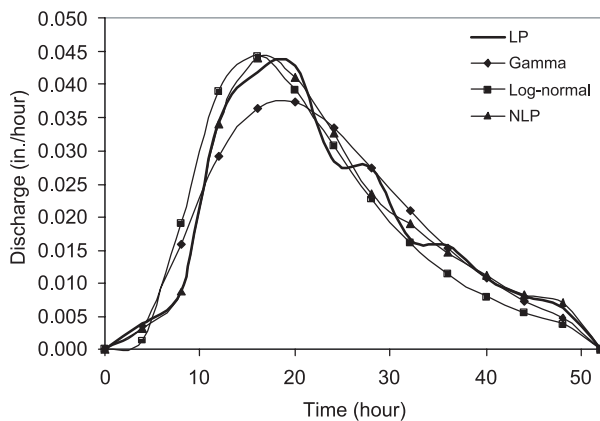


Figure 1. UH using different models (dataset I).

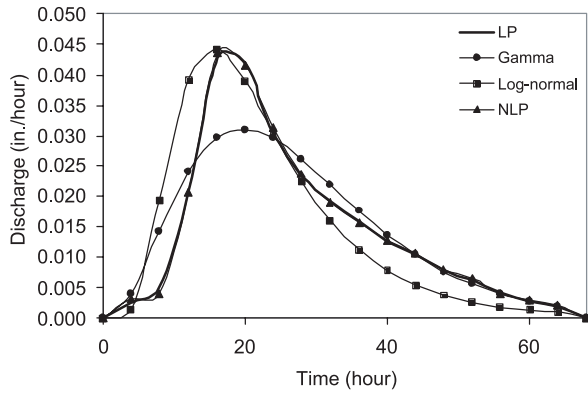


Figure 2. UH using different models (dataset II).

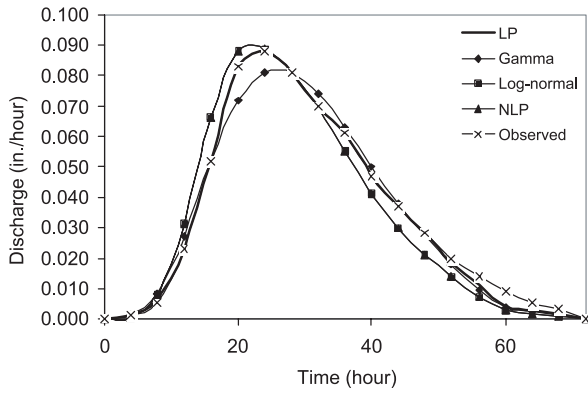


Figure 3. Comparison of DRH (dataset I).

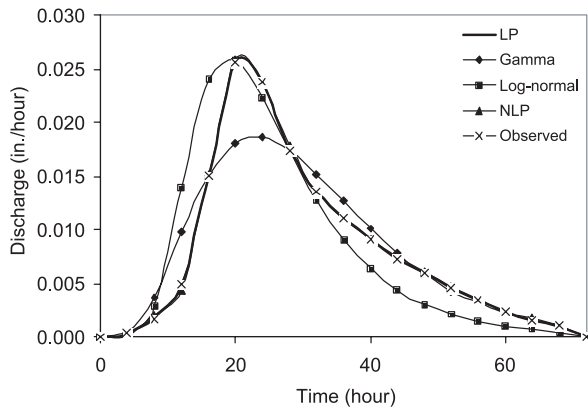


Figure 4. Comparison of DRH (dataset II).

Table 4. Objective function value.

Model	Dataset I	Dataset II
LP (Singh 1976)	0.015040	0.000612
Model I (existing NLP with GA)	0.020740	0.002733
Model II (gamma distribution)	0.053018	0.025232
Model III (log-normal distribution)	0.084061	0.040000

Table 5. Comparison of different optimization formulations.

Model	No. of decision variables	No. of constraints
LP (Singh 1976)	$3N - M + 1$	$N + 1$
Model I (existing NLP)	$N - M + 1$	1
Model II (gamma distribution)	2	Nil
Model III (log-normal distribution)	2	Nil

better in predicting peak flow for both the datasets, but it fails to predict the rising and recession limbs properly. Table 4 shows the objective function values for the different models. The lesser the objective function value the better would be the model performance. The objective function value of LP and model I is almost the same. The objective function value for a unit hydrograph transmuted into the gamma (model II) and log-normal distribution (model III) is not significantly better than the model I and the LP model, but is quite comparable. However, model II is better than model III.

Table 5 shows the comparison of different optimization formulations. Models II and III are much simpler than the LP model in terms of the number of decision variables, and number of constraints. In fact, there are no constraints on models II and III. On the other hand model I is simpler in term of number of decision variables and number of constraints than the LP model. The main advantage of models II and III is that the number of unknowns to be determined by optimization is only two, which can be estimated quickly by using any optimization technique, such as genetic algorithms. Moreover, the unknowns to be determined do not depend upon the number of DRH ordinates.

6. Conclusions

A simple nonlinear optimization model is developed to transmute a unit hydrograph into a probability distribution function. Gamma and log-normal probability distribution functions are used in this study. The nonlinear optimization problem is solved by using genetic algorithms. For the example problems, gamma and log-normal distribution estimate the time-to-peak properly. Gamma distribution fails to estimate the peak discharge, where as log-normal distribution predicts peak discharge more or less correctly. The main advantage of the model is that there are no constraints in the optimization model, and the number of variables to be estimated by optimization is equal to probability distribution parameters. The performance of the model is not better than the LP and existing nonlinear optimization model, but is quite comparable. Moreover, model II is much simpler in terms of the number of variables to be determined and number of constraints. Therefore, this model may be applied for quick and approximate estimation of unit hydrographs. However, more rigorous evaluation with different datasets is required before the applicability of this approach is fully established.

List of symbols

- E elitism size in percentage;
- $f(x)$ probability density function for variable x ;
- M total number of effective rainfall hyetograph ordinates;

m	effective rainfall hyetograph ordinate;
N	total number of direct runoff hydrograph ordinates;
n	number of the direct runoff hydrograph ordinate;
P_m	excess rainfall (L);
p	population size;
p_c	crossover probability;
p_m	mutation probability;
Q'_n	n th actual direct runoff hydrograph ordinate (L/T);
t	time (T);
U_r	r th ordinate of unit hydrograph (L/T);
ε	permissible error for each constraint;
λ	distribution parameter for gamma function (scale parameter);
η	distribution parameter for gamma function (shape parameter);
$\Gamma(x)$	gamma function for random variable x ;
μ_n	distribution parameter for log-normal function (scale parameter);
σ_n	distribution parameter for log-normal function (shape parameter).

References

- Barnes B S 1959 Consistency in unit hydrographs. *Proc. Am. Soc. Civil. Eng.* 85 (HY8): 39–63
- Bender D L, Roberson J A 1961 The use of dimensionless unit hydrograph to derive unit hydrographs for some Pacific basins. *J. Geogr. Res.* 66: 521–527
- Bruen M, Dooge J C I 1984 An efficient and robust method for estimating unit hydrograph ordinates. *J. Hydrol.* 70: 1–24
- Chow V T, Maidment D R, Mays L R 1988 *Applied hydrology* Int. edn (Singapore: McGraw Hill)
- Collins W T 1939 Runoff distribution graphs from precipitation occurring in more than one time unit. *Civ. Eng.* 9: 559–561
- Deb K 2001 *Multi-objective optimization using evolutionary Algorithms* (Chichester: John Wiley and Sons)
- Deinger R A 1969 Linear program for hydrologic analysis. *Water Resources Res.* 5: 1105–1109
- Goldberg D E 1989 *Genetic algorithms in search, optimization, and in machine learning* (Reading, MA: Addison Wiley)
- Holland J H 1975 *Adaptation in natural and artificial systems* (Ann Arbor, MI: University of Michigan Press)
- Morel-Seytoux H J 1982 Optimization methods of rainfall-runoff modeling. *Rainfall-runoff relationships* (ed.) V P Singh (Littleton, CO: Water Resources Pub.) pp 487–506
- Mays L W, Coles L 1980 Optimization of unit hydrograph determination. *J. Hydraul. Div.* 106 (HY1): 85–97
- Mays L W, Taur C K 1982 Unit hydrograph via nonlinear programming. *Water Resources Res.* 18(4): 744–752
- Prasad T D, Gupta R, Prakash S 1999 Determination of optimal loss rate parameters and unit hydrograph. *J. Hydrol. Eng. Am. Soc. Civil. Eng.* 4: 83–87
- Singh K P 1976 Unit hydrographs: A comparative study. *Water Resources Bull.* 12: 381–392
- Singh V P 1988 *Hydrologic systems* (Englewood Cliffs, NJ: Prentice Hall) vol. 1
- Sherman L K 1932 Stream flow from rainfall by the unit graph method. *Eng. News Rec.* 108: 501–505