CrossMark

ORIGINAL ARTICLE

# Probability distribution functions for unit hydrographs with optimization using genetic algorithm

Mohammad Ali Ghorbani[1] · Vijay P. Singh[2] · Bellie Sivakumar[3,4] ·
Mahsa H. Kashani[1] · Atul Arvind Atre[5] · Hakimeh Asadi[1]

**Abstract** A unit hydrograph (UH) of a watershed may be viewed as the unit pulse response function of a linear system. In recent years, the use of probability distribution functions (pdfs) for determining a UH has received much attention. In this study, a nonlinear optimization model is developed to transmute a UH into a pdf. The potential of six popular pdfs, namely two-parameter gamma, two-parameter Gumbel, two-parameter log-normal, two-parameter normal, three-parameter Pearson distribution, and two-parameter Weibull is tested on data from the Lighvan catchment in Iran. The probability distribution parameters are determined using the nonlinear least squares optimization method in two ways: (1) optimization by programming in Mathematica; and (2) optimization by applying genetic algorithm. The results are compared with those obtained by the traditional linear least squares method. The results show comparable capability and performance of two nonlinear methods. The gamma and Pearson distributions are the most successful models in preserving the rising and recession limbs of the unit hydrographs. The log-normal distribution has a high ability in predicting both the peak flow and time to peak of the unit hydrograph. The nonlinear optimization method does not outperform the linear least squares method in determining the UH (especially for excess rainfall of one pulse), but is comparable.

**Keywords** Genetic algorithm · Least squares method · Mathematica · Nonlinear optimization · Probability distribution function · Unit hydrograph

## Introduction

Prediction of flow hydrographs is important for undertaking water emergency measures and management strategies. A large number of methods have been proposed for flow prediction. The unit hydrograph (UH) is one of the most popular and widely used methods, especially in developing countries. A unit hydrograph (Sherman 1932) is defined as the hydrograph of direct runoff resulting from a unit depth of effective rainfall (ER) occurring uniformly over the basin and at a uniform rate for a specified duration. When the duration of ER becomes infinitesimally small, the UH is known as the instantaneous unit hydrograph (IUH). The hydrograph obtained with the use of UH is the direct runoff hydrograph (DRH). Because UH represents a linear response of the basin, the DRH is obtained by convoluting UH with the effective rainfall hyetograph (ERH). The discrete form of convolution can be written as follows (e.g. Chow et al. 1988; Singh 1988):

$$Q_n = \sum_{m=1}^{n \le M} P_m U_{n-m+1}, \tag{1}$$

✉ Mahsa H. Kashani
  mahsakashani2003@yahoo.com

[1] Department of Water Engineering, University of Tabriz, Tabriz, Iran

[2] Department of Biological and Agricultural Engineering, Texas A and M University, Scoates Hall, 2117 TAMU, College Station, TX 77843-2117, USA

[3] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, Australia

[4] Department of Land, Air and Water Resources, University of California, Davis, USA

[5] Mahatma Phule Krishi Vidyapeeth, Rahuri, India

Springer

where $Q_n$ is the DRH ordinate at a discrete time step $n$, $P_m$ is the effective rainfall pulse at a discrete time step $m$, and $U_{n-m+1}$ is the ordinate of the UH at any discrete time step $n - m + 1$. If the number of effective rainfall pulses is $M$ and the number of DRH ordinates is $N$, then there will be $N - M + 1$ ordinates in the UH of the watershed. On the other hand, when effective rainfall pulses ($P_m$'s) and DRH ordinates ($Q_n$'s) are known from observations, Eq. (1) can be used to determine the ordinates of UH through a reverse process. This reverse process of determining the UH ordinates is sometimes referred to as the "de-convolution" process.

There are many methods to solve Eq. (1) for determining the UH. These methods include successive substitution method (Dooge and Bruen 1989), Collins method (Collins 1939), successive approximation method (Newton and Vinyard 1976), Delaine method (Raghavendran and Reddy 1975), harmonic analysis (O'Donnell 1960), Fourier method (Levi and Valdes 1964), Meixner method (Dooge and Garvey 1978), least squares method (Bruen and Dooge 1984), linear programming method (Deininger 1969), and nonlinear programming method (Unver and Mays 1984), among others; see also Singh (1988) for further details.

Mays and Coles (1980) presented a linear programming (LP) model for the determination of composite UH. This model uses the $f$-index method for the estimation of infiltration losses. Prasad et al. (1999) applied an LP model to estimate the optimal loss-rate parameters and UH by considering the inherent characteristics of infiltration and UH. Mays and Taur (1982) developed a nonlinear programming (NLP) model to determine the optimal UH. This method does not require losses to be specified a priori. Unver and Mays (1984) extended the method of Mays and Taur (1982) by incorporating an infiltration equation to estimate the optimal loss-rate parameters and UH.

Although these methods have been shown to perform well for certain situations, their main disadvantage 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, since the number of unknowns is generally large (Bhattacharjya 2004).

Unit hydrographs have common characteristics with probability distribution functions, such as positive ordinates and unit area. As a result, probability distribution functions have recently gained enormous interest in deriving UH. In this approach, the number of unknowns is less and equal to the number of probability distribution parameters. Bardsley (2003) used the inverse Gaussian distribution as an alternative to the gamma distribution as a two-parameter descriptor of the IUH. The inverse Gaussian

distribution was capable of deriving some hydrographs where the gamma would fail. Bhattacharjya (2004) used gamma and log-normal probability distributions to represent the UH for developing two nonlinear optimization models and solved them using binary-coded genetic algorithms. The gamma and log-normal distribution estimated the time to peak correctly. Log-normal distribution predicted peak discharge more or less properly; whereas gamma distribution did not satisfactorily estimate the peak discharge. Moreover, the results showed fairly similar performance of the distributions and the linear optimization model. Bhunya et al. (2007) explored the potential of four popular probability distribution functions (Gamma, Chi square, Weibull, and Beta) to derive synthetic unit hydrograph (SUH) using field data. The results showed that the Beta and Weibull distributions are more flexible in hydrograph prediction. Nadarajah (2007) provided simple Maple programs for determining SUH from eleven of the most flexible probability distributions and derived expressions for the unknown parameters in terms of the time to peak, the peak discharge, and the time base. Rai et al. (2010) derived the UH using the Nakagami-m distribution and compared its results with those of seven other distribution functions over 13 watersheds. The Nakagami-m distribution yielded UHs and direct runoff hydrographs successfully. Singh (2011) employed the entropy theory to derive a general IUH equation on two small agricultural experimental watersheds. This equation was specialized into some distributions, such as the gamma distribution, Lienhard distribution, and Nakagami-m distribution. The results indicated that surface runoff hydrographs computed using the derived IUH equation were in satisfactory agreement with the observed hydrographs.

In the present study, a nonlinear unconstrained optimization model is presented to transmute UHs into probability distribution functions. Six probability distribution functions are considered: two-parameter gamma, two-parameter Gumbel, two-parameter log-normal, two-parameter normal, three-parameter Pearson, and two-parameter Weibull distribution. The nonlinear least squares optimization formulation is solved by (1) programming in Mathematica and (2) by applying genetic algorithm. The potential of these six probability distribution functions is tested on data from the Lighvan catchment in the northwest of Iran. The nonlinear optimization method is compared with the traditional linear least squares method. One particular novelty of this study is the use of Mathematica for solving the nonlinear optimization formulation problem involved in deriving UH. Since Mathematica has extensive symbolic and numerical capabilities, it enables the calculations in a simpler, faster, and more accurate manner. It also has several statistical distributions already built-in.

The rest of this paper is organized as follows the next section presents a brief description of the six probability distribution functions, nonlinear least squares optimization method, and formulation to transmute UH into probability distribution, genetic algorithm, and traditional least squares methods. After describing the case study area, the results of calibration and validation of the methods are discussed. Finally, the conclusions are drawn.

## Materials and methods

### Probability distribution functions

In this study, six popular probability distribution functions are considered: gamma, Gumbel, log-normal, normal, Pearson, and Weibull. A brief description of these functions can be found in Table 9.

### Nonlinear least squares optimization method

In this method, a formula is presented to transmute UH into probability distributions. The objective function is to minimize the sum of the squares of deviation between the actual and the estimated direct runoff hydrographs. This can be written as

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

where $e_n$ is the deviation between the $n$th ordinates of the estimated and actual direct runoff hydrographs, given by

$$e_n = \sum_{m=1}^{n \leq M} P_m U_{n-m+1} - Q_n', \tag{3}$$

where $Q_n'$ is the $n$th ordinate of the actual direct runoff hydrograph, $U_{n-m+1} = f(x)$, where $f(x)$ is a probability distribution function and $x = (n - m + 1) \times \Delta t$.

Two constraints must be considered for this objective function: (1) the area under the UH must be unity; and (2) the UH ordinates must be positive. These are given by

$$1 - \Delta t \sum_{\substack{r=1 \\ U_r \geq 0}}^{N-M+1} U_r = 0 \quad r = 1, 2, 3, \ldots, N - M + 1. \tag{4}$$

In this method, the number of unknowns is equal to the parameters of the probability distribution. In this study, this method is performed by programming in Mathematica and by applying genetic algorithm which is briefly described in next sub-section 2.3.

### Genetic algorithm

The genetic algorithm (GA) is a search technique based on the concept of natural selection inherent in the natural genetics, and combines an artificial survival of the fittest with genetic operators abstracted from nature (Holland 1975). The major difference between GA and the classical optimization search techniques is that the GA works with a population of possible solutions, whereas the classical optimization techniques work with a single solution. An individual solution in a population of solutions is equivalent to a natural chromosome. Like a natural chromosome completely specifies the genetic characteristics of a human being, an artificial chromosome in GA completely specifies the values of various decision variables representing a decision or a solution. For most GAs, the candidate solutions are represented by chromosomes coded with either a binary number system or a real decimal number system. These chromosomes are evaluated based on their performance with respect to the objective function. The GA that employs binary strings as its chromosomes is called the binary-coded GA; whereas the GA that employs real-valued strings as its chromosomes is called the real-coded GA. The real-coded GAs offer certain advantages over the binary-coded GAs as they overcome some of the limitations of the binary-coded GAs (Deb and Agarwal 1995; Deb 2000). Regardless of the coding method used, the GA consists of three basic operations: reproduction, crossover or mating, and mutation. Reproduction is a process in which individual strings are copied according to their fitness (Goldberg 1989). Crossover is considered as 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 (Bhattacharjya 2004). Genetic algorithms start with randomly generating an initial population ($p$) of possible solutions. The population is then operated by the three basic operators in order to produce better offspring for the next generation. This process would repeat till the individual is better enough to suit the objective function.

### Linear least squares method

The least squares method minimizes the objective function which is the sum of squares of deviations of the actual and predicted direct runoff hydrographs. According to Eq. (1), the matrix form of the convolution equation can be written as

$$[Q] = [P][U]. \tag{5}$$

Then, the unit hydrograph is derived using Eq. (6):

$$[U] = \left[[P]^T[P]\right]^{-1}[P]^T[Q], \tag{6}$$

where $T$ and $-1$ indicate the transpose and inverse of the matrices, respectively. Further details about this method can be found in Singh Singh (1988). In this study, all the calculations of this method are performed in Mathematica.

## Study area and data

In this study, the potential of the six probability distribution functions for UH is investigated using data from the Lighvan River in northwest Iran. The Lighvan River watershed is located in East Azarbaijan in the northwest part of Iran (see Fig. 1), between 46°20′30″ and 46°27′30″ east latitude and 37°45′55″ to 37°49′30″ north longitude. This watershed is an important part of the catchment of Talkheh River watershed and has a drainage area of 76.19 km². The maximum and minimum elevations of the watershed are about 3500 and 2000 m, respectively. The length of longest stream is 17 km. The average stream slope is 11 %. The Lighvan River drains into Talkheh River and Urmia Lake. For this watershed, data availability is generally scarce. For the present analysis, data of rainfall and runoff corresponding to four different storms (Storm A, Storm B, Storm C, and Storm D) are considered for calibration of the models. Data corresponding to two other storms (Storm E

and Storm F) are used for validation of the models. Details of these datasets are presented in Table 1.

It is relevant to note that the effective rainfall rates are computed using the $\Phi$-index for each rainfall hyetograph, and the direct runoff hydrographs are obtained by separating base flow from flow hydrographs using the constant-discharge method.

## Results and discussion

We use six probability distribution functions for deriving unit hydrographs for the above datasets: two-parameter gamma, two-parameter Gumbel, two-parameter log-normal, two-parameter normal, three-parameter Pearson distribution, and two-parameter Weibull. The probability distribution parameters are determined using the nonlinear least squares optimization method by programming in Mathematica and by applying the genetic algorithm. The results are also compared with those obtained using the traditional linear least squares method.

### Nonlinear optimization by programming in Mathematica

In the present analysis, the storm data are used to derive a 1-hour unit hydrograph. All the models used involve an inverse problem that optimizes the probability distribution function parameters by minimizing the difference between the actual and predicted direct runoff hydrographs. The



**Fig. 1** Geographical location of Lighvan watershed, Iran
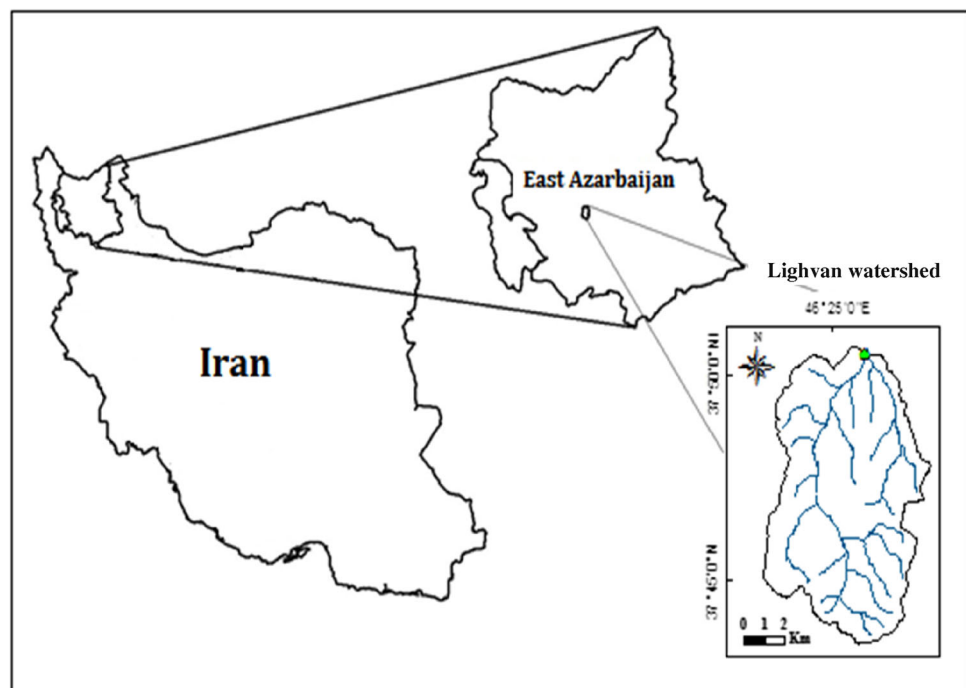
**Table 1** Storm data for Lighvan watershed, Iran

| Time (hr) | Storm A May 23, 2003 | | Storm B June 15, 2003 | | Storm C May 15, 2005 | | Storm D May 16, 2005 | | Storm E (test) May 24, 2003 | | Storm F (test) March 6, 2004 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P (mm) | Q (mm/hr) | P (mm) | Q (mm/hr) | P (mm) | Q (mm/hr) | P (mm) | Q (mm/hr) | P (mm) | Q (mm/hr) | P (mm) | Q (mm/hr) |
| 1 | 0.04 | 0.003828 | 0.44 | 0.04742 | 0.17 | 0.029789 | 0.4 | 0.015604 | 0.43 | 0.004512 | 0.43 | 0.011756 |
| 2 | | 0.008759 | | 0.061746 | | 0.02629 | | 0.054092 | | 0.063483 | 0.92 | 0.024154 |
| 3 | | 0.011794 | | 0.039133 | | 0.019434 | | 0.050357 | | 0.038323 | | 0.0879 |
| 4 | | 0.011794 | | 0.041869 | | 0.016745 | | 0.022845 | | 0.03496 | | 0.089464 |
| 5 | | 0.005781 | | 0.033 736 | | 0.016077 | | 0.024186 | | 0.026276 | | 0.155487 |
| 6 | | 0.000947 | | 0.016908 | | 0.010153 | | 0.024186 | | 0.020032 | | 0.120408 |
| 7 | | | | 0.013179 | | 0.008213 | | 0.024186 | | 0.020032 | | 0.089464 |
| 8 | | | | 0.013179 | | 0.008213 | | 0.021511 | | 0.018002 | | 0.080174 |
| 9 | | | | 0.015659 | | 0.006289 | | 0.018867 | | 0.018002 | | 0.068148 |
| 10 | | | | 0.019426 | | 0.005016 | | 0.018211 | | 0.019014 | | 0.057962 |
| 11 | | | | 0.020695 | | 0.005016 | | 0.018211 | | 0.020032 | | 0.052283 |
| 12 | | | | 0.020695 | | 0.004382 | | 0.018211 | | 0.020032 | | 0.049482 |
| 13 | | | | 0.016908 | | 0.003121 | | 0.018211 | | 0.019014 | | 0.049482 |
| 14 | | | | 0.014415 | | 0.001867 | | 0.018211 | | 0.019014 | | 0.041233 |
| 15 | | | | 0.011949 | | 0.001243 | | 0.015604 | | 0.015998 | | 0.038535 |
| 16 | | | | 0.010725 | | 0.003121 | | 0.012389 | | 0.015998 | | 0.037195 |
| 17 | | | | 0.010725 | | | | 0.009221 | | 0.014019 | | 0.037195 |
| 18 | | | | 0.009508 | | | | 0.007967 | | 0.01304 | | 0.037195 |
| 19 | | | | 0.008297 | | | | 0.007343 | | 0.010139 | | 0.034536 |
| 20 | | | | 0.007092 | | | | 0.0061 | | 0.005434 | | 0.030595 |
| 21 | | | | 0.002338 | | | | | | 0.004512 | | 0.02543 |
| 22 | | | | | | | | | | 0.004512 | | 0.024154 |
| 23 | | | | | | | | | | 0.004512 | | 0.020367 |
| 24 | | | | | | | | | | 0.004512 | | 0.015408 |
| 25 | | | | | | | | | | | | 0.014184 |
| 26 | | | | | | | | | | | | 0.012967 |
| 27 | | | | | | | | | | | | 0.011756 |
| 28 | | | | | | | | | | | | 0.009353 |
| 29 | | | | | | | | | | | | 0.006976 |
| 30 | | | | | | | | | | | | 0.005798 |
| 31 | | | | | | | | | | | | 0.003459 |
| 32 | | | | | | | | | | | | 0.0023 |
| 33 | | | | | | | | | | | | 0.001147 |

probability distribution parameters are obtained using least squares optimization method.

## Calibration of the models

The parameters of probability distributions obtained for the storms (A–D) are shown in Table 2. The 1-hour unit hydrographs for the four datasets are presented in Fig. 2a–d, and the resulting direct runoff hydrographs are indicated in Fig. 3a–d, respectively. Figure 2a–d indicate that none of the models have tail oscillations. The oscillations of the

UH determined by the least squares method for B, C and D storms may be caused by errors in data measurements, the rainfall abstractions, base flow separation, and non-uniform temporal and spatial distribution of rainfall. All the distribution functions predict the peak discharge, the time to peak, and the shape of the UH successfully for storm A. For storm B, all the distributions estimate the time to peak correctly. The peak discharge estimated by the Weibul and log-normal functions is closer to the actual value. The performance of all the models except the normal and Gumbel is satisfactory in predicting the peak discharge,

**Table 2** Parameters of probability distribution functions calibrated by the nonlinear mathematical optimization method for Lighvan watershed

| Storm | Gamma | | Gumbel | | Log-normal | | Normal | | Pearson | | Weibull | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | α | β | α | β | α | β | α | β | α | β | α | β | γ |
| A | 0.6774 | 5.2076 | 3.6075 | 1.2123 | 1.2293 | 0.3992 | 3.2550 | 1.3035 | 6.9597 | 0.5211 | 3.7207 | 2.9095 | −0.1073 |
| B | 7.0135 | 1.0402 | 2.4152 | 3.3580 | 1.7216 | 1.1307 | 2.1155 | 4.0066 | 1.0400 | 7.0156 | 7.3400 | 1.0124 | 0 |
| C | 3.9788 | 1.1616 | 2.7072 | 2.8155 | 1.3491 | 1.0494 | 2.1211 | 2.9119 | 1.1617 | 3.9788 | 4.6761 | 1.0884 | 0 |
| D | 5.6905 | 1.4903 | 6.9192 | 6.0451 | 1.8902 | 0.9611 | 5.0665 | 5.8073 | 1.4902 | 5.6911 | 8.9275 | 1.2677 | 0 |



**Fig. 2** Comparison of UHs derived using the linear least squares method and distribution functions calibrated by the nonlinear mathematical optimization method for Lighvan watershed: **a** Storm A; **b** Storm B; **c** Storm C; and **d** Storm D
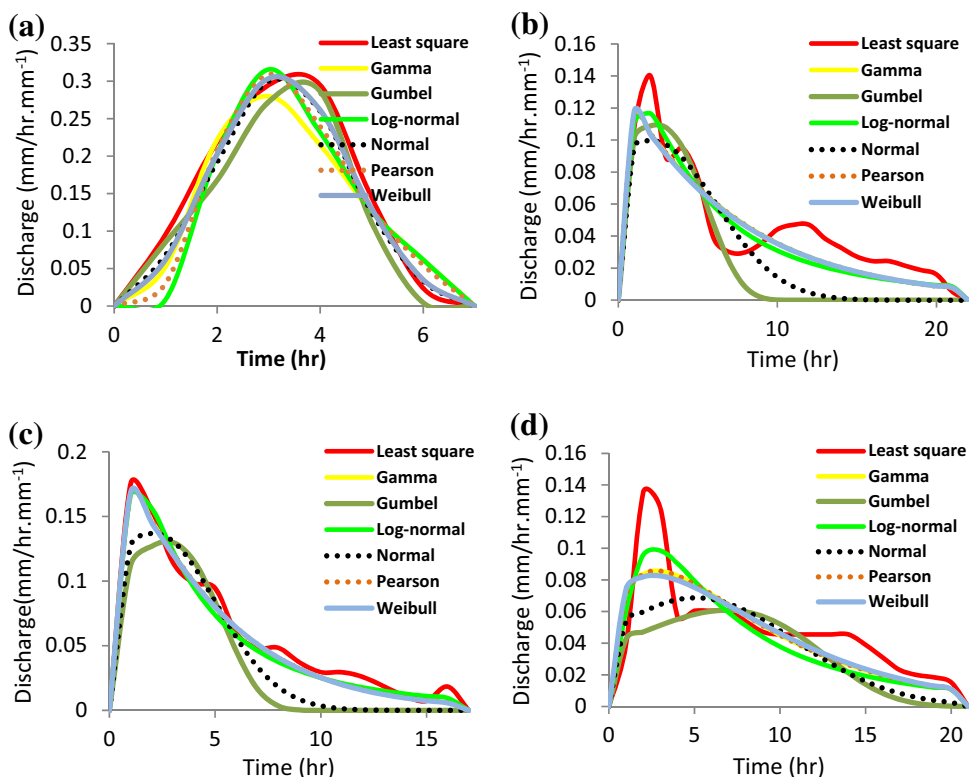
time to peak, and the UH shape for storm C. The normal and Gumbel distributions are also not successful in predicting the time to peak and the rising limb of the UH for storm D. The peak discharge is estimated with less error by the log-normal model. Similar results can be obtained from Fig. 3a–d. Table 3 shows the objective function values for different models. As can be seen from this table, the Gumbel and normal distribution functions have high objective function values for all the storms except storm A. The objective function values of gamma and Pearson models are almost the same for all the storm data. For storm A, the Weibull and normal distributions outperform the other distribution, because these distributions showed a high ability in predicting the rising and recession limbs as seen from Fig. 2a. For storms B and D, the lowest objective

function values are for the log-normal distribution, whereas for storm C the gamma and Pearson show the lowest value of the objective function. If average value of the objective functions is considered for all four storms, then the log-normal distribution gives the lowest objective function value (0.000473). The objective function value of the linear least squares method is very low which indicates the high ability of this method than the nonlinear optimization method for deriving the UH.

Generally, based on the visual comparison at the calibration stage using the nonlinear optimization method, it was observed that the log-normal distribution estimates the time to peak and peak flow properly for all storms. This distribution along with the gamma, Pearson, and Weibull predicts the rising and recession limbs of the unit

**Fig. 3** Comparison of observed
and estimated DRHs related to
the linear least squares method
and distribution functions
calibrated by the nonlinear
mathematical optimization
method for Lighvan watershed:
**a** Storm A; **b** Storm B; **c** Storm
C; and **d** Storm D
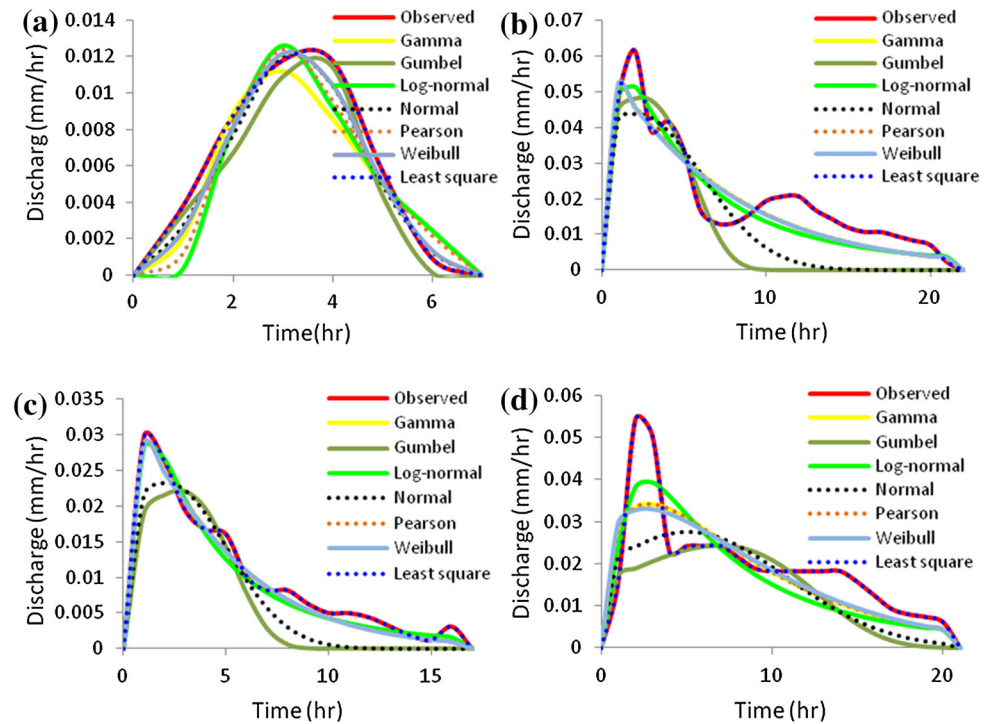


**Table 3** Objective function values for six distribution functions calibrated by the linear and nonlinear mathematical optimization method for Lighvan watershed

| Model | Storm A | Storm B | Storm C | Storm D |
|---|---|---|---|---|
| Gamma distribution | 0.000016 | 0.000911 | 0.000026 | 0.001224 |
| Gumbel distribution | 0.000008 | 0.002922 | 0.000382 | 0.002687 |
| Log-normal distribution | 0.000022 | 0.000788 | 0.000030 | 0.001051 |
| Normal distribution | 0.000005 | 0.002399 | 0.000234 | 0.002055 |
| Pearson distribution | 0.000014 | 0.000911 | 0.000026 | 0.001224 |
| Weibull distribution | 0.000005 | 0.000914 | 0.000027 | 0.001270 |
| Least squares | 2.416E−19 | 1.9184E−18 | 1.1971E−18 | 2.0000E−18 |

hydrographs more or less perfectly. Moreover, the log-normal distribution was recognized as the most successful model based on the average value of the objective function.

*Validation of the models*

In order to validate the models, average values of the parameters of distribution functions obtained for the four storms were calculated and 1-hour unit hydrographs were

derived using the distribution functions with the known parameters. The direct runoff hydrographs were obtained from these unit hydrographs by convoluting them with effective rainfall rates for storms E and F. Figure 4a, b illustrate the derived unit hydrographs for storms E and F, and the resulting direct runoff hydrographs are shown in Fig. 5a, b, respectively. As can be seen from Figs. 4, 5, the log-normal distribution predicts the peak flow for both storms and the time-to-peak for storm E with less error. The Weibull and Pearson distributions perform well in estimating the peak discharge and the time-to-peak for storm F, respectively. Furthermore, none of the distributions predict the rising and recession limbs properly. However, the Gamma and Pearson models estimate the limbs fairly well. Note that the tail end of the hydrographs for storm E is also properly predicted by the Gamma and Pearson distributions. Since storm F is the only one which occurred in winter season when the watershed is covered by snow, one can expect to not see good performance of the models for this storm.

Besides the visual comparison, the model performance is also evaluated using following three statistical measures:

1. Root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(Q_{e_i} - Q_{o_i}\right)^2}{n}} \qquad (7)$$

2. Mean absolute error (MAE):

Fig. 4 Comparison of UHs derived using the linear least squares method and distribution functions calibrated by the nonlinear mathematical optimization method for Lighvan watershed: **a** Storm E; and **b** Storm F
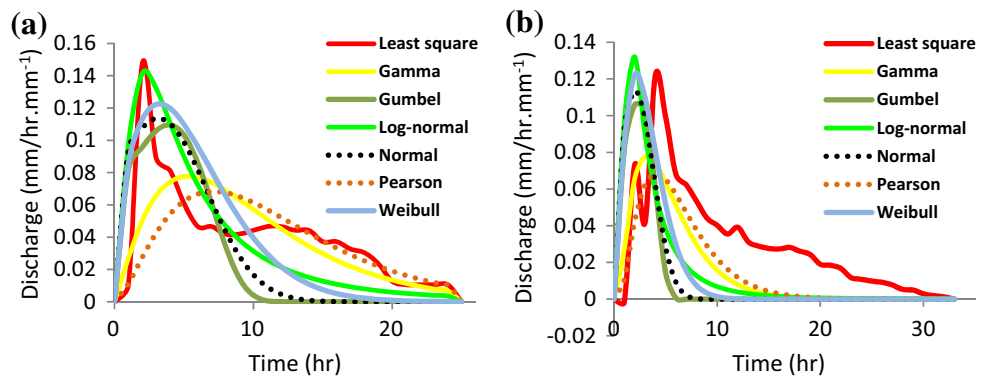


Fig. 5 Comparison of observed and estimated DRHs related to the linear least squares method and distribution functions calibrated by the nonlinear mathematical optimization method for Lighvan watershed: **a** Storm E; and **b** Storm F
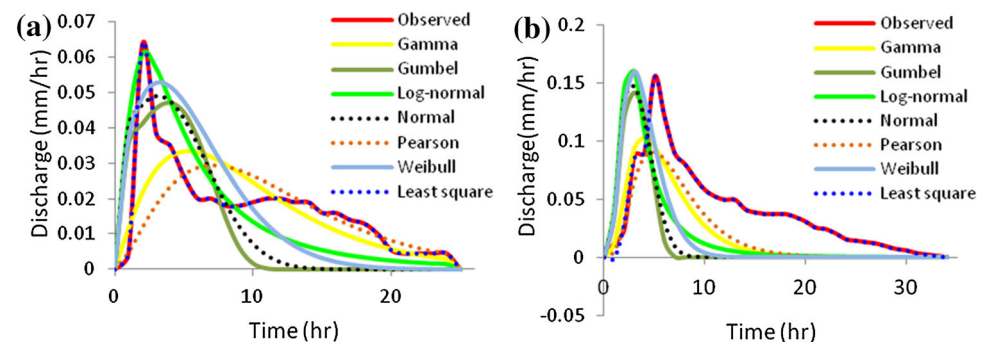


$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |Q_{e_i} - Q_{o_i}| \qquad (8)$$

3. Correlation coefficient (CC):

$$\text{CC} = \frac{\sum_{i=1}^{n}(Q_{o_i} - \bar{Q}_o)(Q_{e_i} - \bar{Q}_e)}{\sqrt{\sum_{i=1}^{n}(Q_{o_i} - \bar{Q}_o)^2}\sqrt{\sum_{i=1}^{n}(Q_{e_i} - \bar{Q}_e)^2}}, \qquad (9)$$

where $Q_{o_i}$ and $Q_{e_i}$ are the $i$th observed and estimated DRH ordinates, respectively; $\bar{Q}_o$ and $\bar{Q}_e$ represents the average discharge of the observed and estimated DRH, respectively, and $n$ is the number of ordinates.

Table 4 presents the values of performance criteria. According to this table, the performance criteria values of the gamma and Pearson distributions were close to each other. The gamma distribution with the lowest value of RMSE (0.010) and MAE (0.006 mm/h),the Pearson distribution with the lowest value of RMSE (0.025), MAE (0.021 mm/h), and the highest value of CC (0.929) show successful performances for storms E and F, respectively. The performance of the log-normal model with the highest value of CC (0.776) and the low value of RMSE and MAE (0.012 and 0.009 mm/h, respectively) is successful for storm E. The Gumbel distribution may not be a suitable model for estimating the UH because of its high RMSE and MAE values for both storm data. Generally, the performance of almost all the models is more accurate for storm E than for F. The least squares method shows satisfactory results, considering its statistical measure values for both

events. For storm F, this method indicates more error than storm E, because it generated a negative value for the first ordinate of the UH which is the main disadvantage of this method. According to the results of the study done by Singh (1976), the derived unit hydrographs using the least squares method may not have a unit volume and some unit hydrographs ordinates may be negative.

In general, the results of the validation stage showed that the lognormal distribution performance is satisfactory in predicting peak flow and time to peak. The gamma and Pearson models showed acceptable performances in simulating both limbs of the unit hydrographs. Hence, according to the values of statistical measures, these distributions outperformed the others.

## Nonlinear optimization by applying genetic algorithm

In this study, the real-coded genetic algorithm in MATLAB software was applied to determine optimal probability distribution parameters. The genetic algorithm parameters, such as crossover and mutation probability applied in this study are given in Table 5.

### Calibration of the models

The optimal probability distribution parameters are shown in Table 6. Figure 6a–d illustrate the UHs obtained for storms A, B, C, and D, respectively, and Fig. 7a–d present

**Table 4** Performance criteria values for six distribution functions calibrated by the linear and nonlinear mathematical optimization method for Lighvan watershed

| Model | RMSE (mm/hr) | | MAE (mm/hr) | | CC | |
|---|---|---|---|---|---|---|
| | Storm E | Storm F | Storm E | Storm F | Storm E | Storm F |
| Gamma distribution | 0.010 | 0.027 | 0.006 | 0.023 | 0.619 | 0.863 |
| Gumbel distribution | 0.015 | 0.048 | 0.013 | 0.038 | 0.642 | 0.397 |
| Log-normal distribution | 0.012 | 0.043 | 0.009 | 0.034 | 0.776 | 0.495 |
| Normal distribution | 0.014 | 0.047 | 0.012 | 0.038 | 0.670 | 0.428 |
| Pearson distribution | 0.012 | 0.025 | 0.006 | 0.021 | 0.402 | 0.929 |
| Weibull distribution | 0.013 | 0.042 | 0.012 | 0.035 | 0.710 | 0.571 |
| Least squares | 2.9E−10 | 0.002 | 2.45E−10 | 0.001 | 1.000 | 0.998 |

the corresponding DRHs. From Figs. 6, 7, it can be seen that for storm A, all the models estimate the time to peak properly, but the peak discharge is estimated more correctly by the Pearson and Weibull distributions. The performance of the normal model in estimating the rising limb of the unit hydrograph is noticeable. Figure 6b shows all the models estimate the time-to-peak properly. However, the accuracy of the Pearson and log-normal distributions is high in predicting the peak flow. Almost all the models predict the rising limb of the unit hydrograph well. For storm C, all the models estimate the time to peak perfectly. The gamma, lognormal, Pearson, and Weibull distributions predict the peak flow and the rising and recession limbs of the UH with less error. For storm D, the gamma, Pearson and log-normal distribution models estimate the time-to-peak and the UH limbs satisfactorily. The peak discharge is estimated properly also by the log-normal model. Table 7 illustrates the objective function values of the distributions. According to this table, the Weibull distribution for storms A and D, and the log-normal and gamma functions for storms B and C give minimum values of the objective function, respectively. Based on the average value of the objective function, the Weibull distribution outperforms the other models for all storms. According to Tables 3 and 7, using the genetic algorithm caused an increase in the objective function values of the models rather than applying the nonlinear mathematical optimization. In other words, the nonlinear mathematical optimization method outperforms the genetic algorithm at the calibration stage.

Generally, at the calibration stage using the genetic algorithm method, the lognormal, Pearson, and gamma models predicted the time to peak more or less properly for all storms. These models along with the Weibull distribution were also successful in simulating the rising and falling limbs of the UHs for all storms except A. The lognormal distribution showed high ability in estimating the peak value for storms B, C, and D. However, the Pearson model can compute well the peak discharge for storms A,

**Table 5** Genetic algorithm parameters

| | |
|---|---|
| Population size ($p$) | 15* (number of variables) |
| Crossover probability ($p_c$) | 1.00 |
| Mutation probability ($p_m$) | 0.01 |
| Generation ($g$) | 200* (number of variables) |

B, and C. The Weibull distribution was distinguished as the most successful model based on the average value of the objective functions because of the excellent ability in preserving the UH shape of storm A.

### Validation of the models

Figure 8a, b show the estimated one-hour unit hydrographs using the average values of the obtained distributions parameters and effective rainfall data for storms E and F, and Fig. 9a, b indicate the corresponding direct runoff hydrographs, respectively. According to Figs. 8, 9, for storm E, the Gumbel, log-normal, and normal models estimate the time to peak perfectly. The log-normal distribution shows high potential in predicting the peak flow. The models did not have a high ability in estimating the recession limbs. For storm F, only the gamma and Weibull distributions estimate the time to peak and peak discharge precisely, respectively. All the models except gamma and Pearson show poor performance in predicting the rising and recession limbs of the UH. Table 8 gives the values of the three statistical measures. The Table illustrates that the gamma distribution with the lowest value of RMSE (0.010), MAE (0.007) and fairly high value of CC (0.697) may be the best model for storm E. This distribution also shows the lowest value of RMSE (0.016), MAE (0.012), and the highest value of CC (0.922) for storm F as the most suitable model. The Pearson model shows almost similar results with the gamma distribution for both storms. The log-normal model gives the highest CC value (0.738) and low RMSE (0.013) and MAE (0.009) values for storm E.

**Table 6** Parameters of probability distribution functions calibrated by genetic algorithm for Lighvan watershed

| Storm | Gamma | | Gumbel | | Log-normal | | Normal | | Pearson | | Weibull | | |
|-------|-------|-------|--------|-------|------------|-------|--------|-------|---------|-------|---------|-------|-------|
| | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\gamma$ |
| A | 0.8700 | 4.4860 | 3.5430 | 1.4040 | 1.3580 | 0.6170 | 2.9960 | 1.4440 | 6.9700 | 0.5030 | 3.7380 | 2.7820 | −0.0180 |
| B | 7.0100 | 1.0000 | 2.0085 | 3.3390 | 1.6695 | 1.1663 | 1.3436 | 3.9870 | 1.0848 | 5.9851 | 7.6640 | 1.0000 | −0.5284 |
| C | 3.9986 | 1.1843 | 1.8231 | 2.9389 | 1.2484 | 1.0642 | 2.0452 | 2.6178 | 1.2912 | 3.0973 | 3.9400 | 0.9990 | −0.6781 |
| D | 4.3955 | 1.5072 | 2.0252 | 6.7924 | 1.6042 | 0.8630 | 3.9806 | 5.9895 | 1.7393 | 4.0860 | 8.9220 | 1.3910 | −0.3195 |

**Fig. 6** Comparison of UHs derived using the linear least squares method and distribution functions calibrated by the genetic algorithm for Lighvan watershed: **a** Storm A; **b** Storm B; **c** Storm C; and **d** Storm D
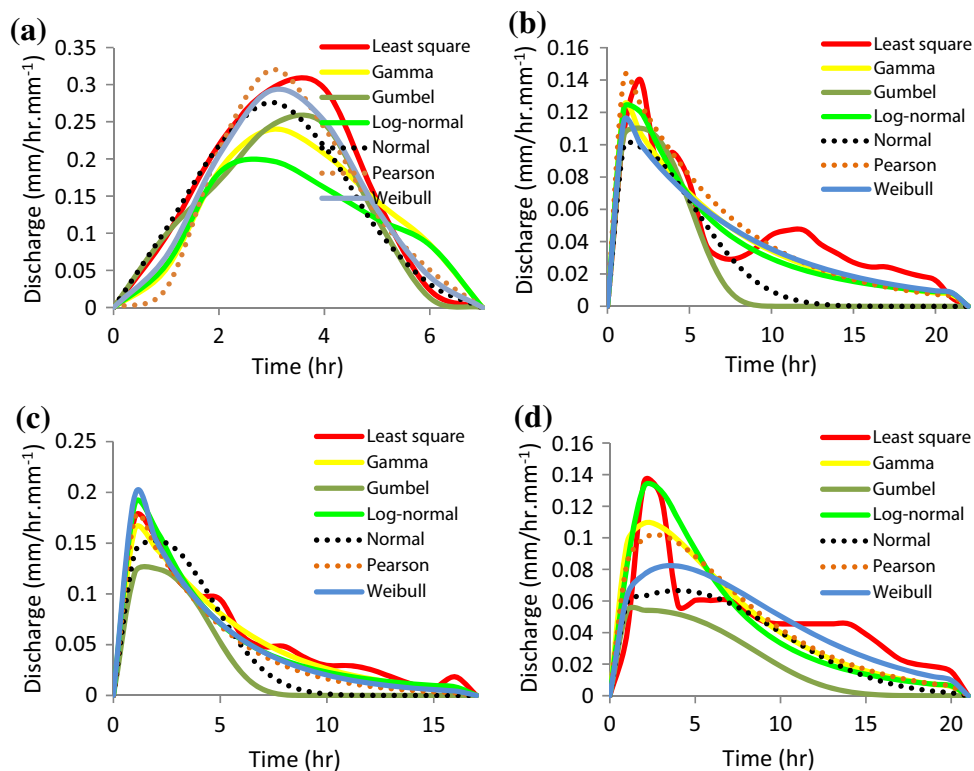


The Gumbel model performance according to the statistical measures is poor for both storms. Similar to the previous validation stage, the models performance for storm E is better than for storm F. Using the genetic algorithm has improved the models capability just for storm F compared with the nonlinear mathematical optimization method.

Generally, at the validation stage, the log-normal distribution showed good performance in predicting the time to peak and peak flow of the UH for storm E. The gamma and Pearson distributions were able to preserve the UH shape. Hence, the gamma distribution with the lowest value of RMSE and MAE, and the highest value of CC is the best model for both storms. The Pearson model indicated similar results with the gamma distribution.

## Conclusions

In this study, a nonlinear model was developed to transmute a unit hydrograph into a probability distribution function. The gamma, Gumbel, log-normal, normal, Pearson, and Weibull probability distribution functions were used to derive 1-hour unit hydrographs. The main advantage of this model is that the number of parameters to be determined is equal to the number of probability distribution
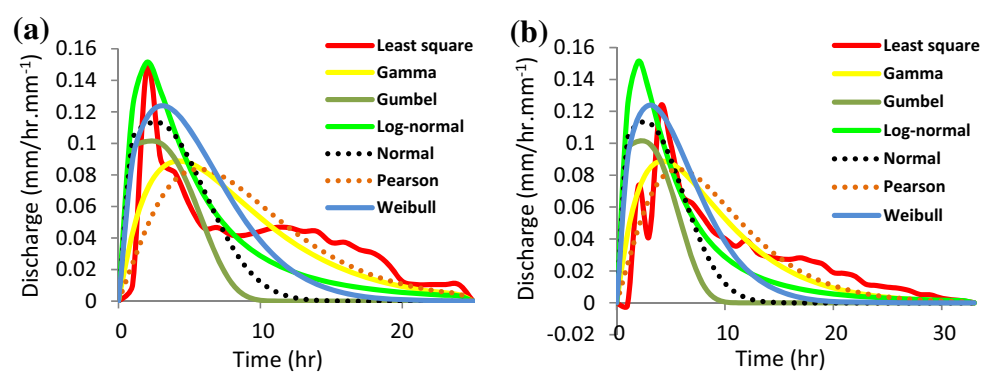
**Fig. 7** Comparison of observed and estimated DRHs related to the linear least squares method and distribution functions calibrated by the genetic algorithm for Lighvan watershed: **a** Storm A; **b** Storm B; **c** Storm C; and **d** Storm D
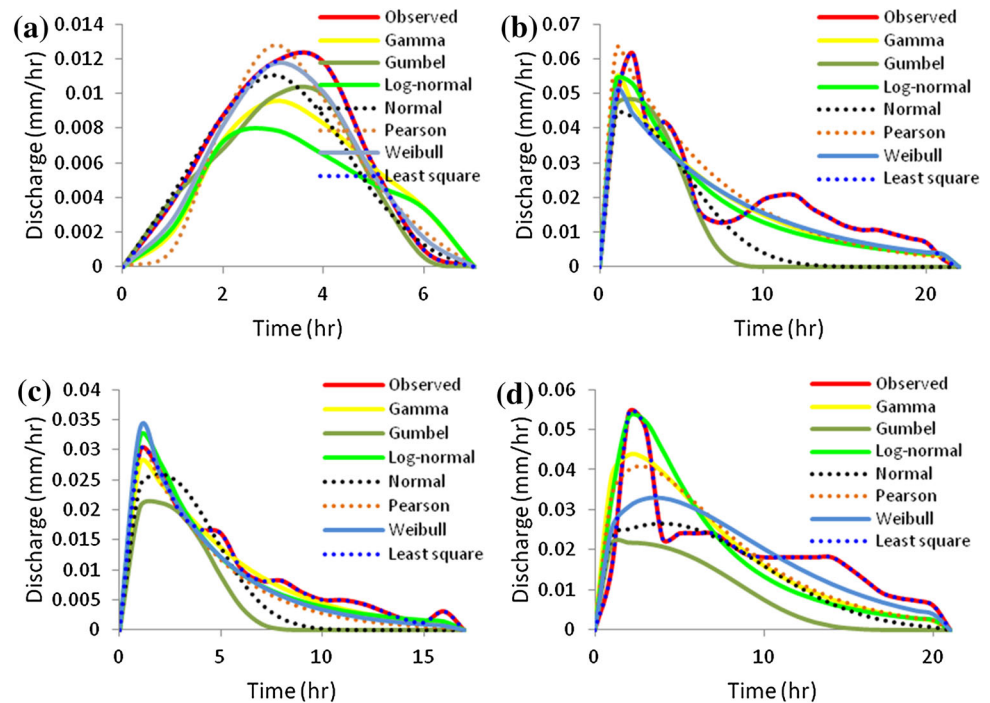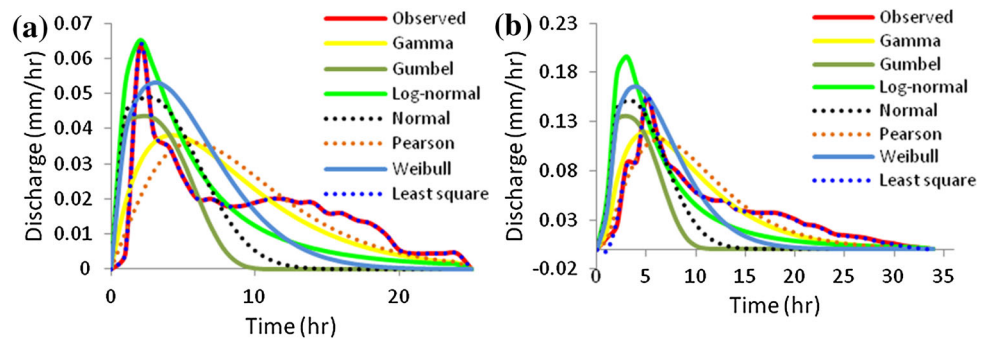


**Table 7** Objective function values for six distribution functions calibrated by the genetic algorithm for Lighvan watershed

| Model | Storm A | Storm B | Storm C | Storm D |
|---|---|---|---|---|
| Gamma distribution | 0.000028 | 0.000926 | 0.000027 | 0.001624 |
| Gumbel distribution | 0.000013 | 0.002987 | 0.000457 | 0.003910 |
| Log-normal distribution | 0.000055 | 0.000826 | 0.000047 | 0.001644 |
| Normal distribution | 0.000013 | 0.002531 | 0.000251 | 0.002149 |
| Pearson distribution | 0.000015 | 0.001150 | 0.000080 | 0.001492 |
| Weibull distribution | 0.000005 | 0.000923 | 0.000065 | 0.001342 |

**Fig. 8** Comparison of UHs derived using the linear least squares method and distribution functions calibrated by the genetic algorithm for Lighvan watershed: **a** Storm E; and **b** Storm F

**Fig. 9** Comparison of observed and estimated DRHs related to the linear least squares method and distribution functions calibrated by the genetic algorithm for Lighvan watershed: **a** Storm E; and **b** Storm F



**Table 8** Performance criteria values for six distribution functions calibrated by the genetic algorithm for Lighvan watershed

| Model | RMSE (mm/hr) | | MAE (mm/hr) | | CC | |
|---|---|---|---|---|---|---|
| | Storm E | Storm F | Storm E | Storm F | Storm E | Storm F |
| Gamma distribution | 0.010 | 0.016 | 0.007 | 0.012 | 0.697 | 0.922 |
| Gumbel distribution | 0.015 | 0.038 | 0.012 | 0.032 | 0.655 | 0.654 |
| Log-normal distribution | 0.013 | 0.039 | 0.009 | 0.024 | 0.738 | 0.702 |
| Normal distribution | 0.014 | 0.036 | 0.012 | 0.028 | 0.669 | 0.736 |
| Pearson distribution | 0.011 | 0.016 | 0.007 | 0.013 | 0.576 | 0.906 |
| Weibull distribution | 0.013 | 0.032 | 0.012 | 0.023 | 0.711 | 0.839 |

parameters. In this case, six different storm datasets from the Lighvan catchment were provided. Four storm datasets were used for models calibration and two for validation. The calibration of models was performed using the nonlinear least squares optimization methods, by programming in Mathematica and by applying the genetic Algorithm, and using the traditional linear least squares method.

In general, the following conclusions may be drawn:

1. The log-normal distribution function has a high potential in predicting the peak flow and the time to peak of the UH.

2. The gamma and Pearson distributions are more able in preserving the rising and recession limbs of the UH.

3. The log-normal, gamma, and Pearson distribution functions can be applied for quick and approximate estimation of unit hydrographs for the Lighvan catchment.

4. The genetic algorithm did not improve the models performance significantly compared with the nonlinear mathematical optimization.

5. The nonlinear optimization methods are not superior to the linear least squares method when there is only one excess rainfall pulse, but are comparable. The main disadvantage of the traditional least squares method is that it may generate negative unit hydrograph ordinates especially when the number of excess rainfall pulses is bigger than one.

# Appendix 1

See Appendix Table 9.

**Table 9** The probability distribution functions (pdf) used in this study

| Pdf | Formula | Parameters |
| --- | --- | --- |
| Gamma distribution (Bhattacharjya 2004) | $f(x) = \frac{x^{\beta-1}}{\alpha^\beta \Gamma(\beta)} \exp\left(-\frac{x}{\alpha}\right)$ <br> for $x\rangle 0,\ \alpha\rangle 0,\ \beta\rangle 0$ | $\alpha\beta = \frac{1}{n}\sum_{i=1}^{n} x_i$ <br><br> $\alpha^2\beta^2 + \beta\alpha^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ |
| Gumbel distribution (Gumbel 1960) | $f(x) = \frac{1}{\beta}\exp\left[-\exp\left(+\frac{x-\alpha}{\beta}\right) + \frac{x-\alpha}{\beta}\right]$ | $\alpha = \frac{1}{n}\sum_{i=1}^{n} x_i - 0.5772157\beta$ <br><br> $\beta = 0.7797\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$ where $\bar{x}$ is the mean of $x_i$'s. |
| Log-normal distribution (Nadarajah 2007) | $f(x) = \left(1/\left(x\beta\sqrt{2\pi}\right)\right)\exp\left[-(\ln x - \alpha)^2 \big/ (2\beta^2)\right]$ <br> for $x\rangle 0,\ -\infty\langle\alpha\langle\infty,\ \beta\rangle 0$ | $\alpha = \frac{1}{n}\sum_{i=1}^{n}\ln x_i$ <br><br> $\beta = \sqrt{\frac{\sum_{i=1}^{n}(\ln x_i - \alpha)^2}{n}}$ |
| Normal distribution (Rao and Hamed 2000) | $f(x) = \left(1/\left(\beta\sqrt{2\pi}\right)\right)\exp\left[-(x-\alpha)^2 \big/ (2\beta^2)\right]$ | $\alpha = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ <br><br> $\beta = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\alpha)^2}{n}}$ |
| Pearson distribution (Rao and Hamed 2000) | $f(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)}$ <br> for $x\rangle\gamma$ | $\alpha = \frac{4\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^3}{n\left(\sum_{i=1}^{n}(x_i-\bar{x})^3\right)^2}$ <br><br> $\beta = \sqrt{\frac{1}{n\alpha}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ <br><br> $\gamma = \bar{x} - \alpha\beta$ |
| Weibull distribution (Bhunya et al. 2007) | $f(x) = \frac{\beta x^{\beta-1}}{\alpha^\beta}\exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right]$ <br> for $x\rangle 0,\ \alpha\rangle 0,\ \beta\rangle 0$ | $\alpha\Gamma\left(\frac{1}{\beta}+1\right) = \frac{1}{n}\sum_{i=1}^{n} x_i$ <br><br> $\alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ |

## References

Bardsley WE (2003) An alternative distribution for describing the instantaneous unit hydrograph. J Hydrol 62(1–4):375–378

Bender DL, Roberson JA (1961) The use of a dimensionless unit hydrograph to derive unit hydrographs for some Pacific basins. J Geogr Res 66:521–527

Bhattacharjya RK (2004) Optimal design of unit hydrographs using probability distribution and genetic algorithms. Sadhana 29(5):499–508

Bhunya PK, Berndtsson R, Ojha CSP, Mishra SK (2007) Suitability of Gamma, Chi square, Weibull, and Beta distributions as synthetic unit hydrographs. J Hydrol 334:28–38

Bruen M, Dooge JCI (1984) An efficient and robust method for estimating unit hydrograph ordinates. J Hydrol 70:1–24

Chow VT, Maidment DR, Mays LR (1988) Applied hydrology. McGraw-Hill International Editions, Singapore

Collins WT (1939) Runoff distribution graphs from precipitation occurring in more than one time unit. Civ Eng 9:559–561

Deb K (2000) An efficient constraint handling method for genetic algorithms. Comput Methods Appl Mech Eng 186:311–338

Deb K, Agarwal RB (1995) Simulated binary crossover for continuous search space. Complex Syst 9:115–148

Deininger RA (1969) Linear program for hydrologic analysis. Water Resour Res 5:1105–1109

Dooge JCI, Bruen M (1989) Unit hydrograph stability and linear algebra. J Hydrol 111(1–4):377–390

Dooge JCI, Garvey BJ (1978) The use of Meixner function in the identification of heavily damped systems. Proc R Irish Acad Sec A 78(18):157–179

Goldberg DE (1989) Genetic algorithms in search, optimization, and in machine learning. Addison Wiley, USA

Gumbel EJ (1960) Multivariate extreme distributions. Bull Inter Statist Inst 39(2):471–475

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor 183

Levi E, Valdes E (1964) A method for direct analysis of hydrographs. J Hydrol 2:182–190

Mays LW, Coles L (1980) Optimization of unit hydrograph determination. J Hydraul Div ASCE 106(HY1):85–97

Mays LW, Taur CK (1982) Unit hydrograph via nonlinear programming. Water Resour Res 18(4):744–752

Nadarajah S (2007) Probability models for unit hydrograph derivation. J Hydrol 344:185–189

Newton DW, Vinyard JW (1976) Computer-determined unit hydrograph from floods. J Hydraul Div 93(5): 219–235

O'Donnell T (1960) Instantaneous unit hydrograph by harmonic analysis. IASH Pub 51:546–557

Prasad TD, Gupta R, Prakash S (1999) Determination of optimal loss rate parameters and unit hydrograph. J Hydrol Eng 4:83–87

Raghavendran R, Reddy PJ (1975) Synthesis of basin response with inadequate data. Nord Hydrol 6:14–27

Rai RK, Sarkar S, Upadhyay A, Singh VP (2010) Efficacy of Nakagami-m distribution function for deriving unit hydrograph. Water Resour Manag 24:563–575

Rao AR, Hamed KH (2000) Flood Frequency Analysis. Print ISBN: 978-0-8493-0083-7

Sherman LK (1932) Stream flow from rainfall by the unit hydrograph method. Eng News Rec 108:501–505

Singh KP (1976) Unit hydrographs: a comparative study. Water Resour Bull 12(2):381–392

Singh VP (1988) Hydrologic systems, vol 1. Prentice Hall, Englewood Cliffs

Singh VP (2011) An IUH equation based on entropy theory. Trans ASABE 54(1):131–140

Unver O, Mays LW (1984) Optimal determination of loss rate functions and unit hydrographs. Water Resour Res 20(2):203–214