Research Article

Application of Harmony Search to Design Storm Estimation from Probability Distribution Models

Sukmin Yoon,[1] Changsam Jeong,[2] and Taesam Lee[1]

[1] *Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju 660-701, Republic of Korea*
[2] *Department of Civil & Environmental Engineering, Induk University, Wolgye 2-dong, Nowon-gu, Seoul 139-052, Republic of Korea*

Correspondence should be addressed to Taesam Lee; tae3lee@gnu.ac.kr

The precision of design storm estimation depends on the selection of an appropriate probability distribution model (PDM) and parameter estimation techniques. Generally, estimated parameters for PDMs are provided based on the method of moments, probability weighted moments, and maximum likelihood (ML). The results using ML are more reliable than the other methods. However, the ML is more laborious than the other methods because an iterative numerical solution must be used. In the meantime, metaheuristic approaches have been developed to solve various engineering problems. A number of studies focus on using metaheuristic approaches for estimation of hydrometeorological variables. Applied metaheuristic approaches offer reliable solutions but use more computation time than derivative-based methods. Therefore, the purpose of the current study is to enhance parameter estimation of PDMs for design storms using a recently developed metaheuristic approach known as a harmony search (HS). The HS is compared to the genetic algorithm (GA) and ML via simulation and case study. The results of this study suggested that the performance of the GA and HS was similar and showed more accurate results than that of the ML. Furthermore, the HS required less computation time than the GA.

1. Introduction

Floods that occurred due to severe rain or major storm generally result in damage to properties and negative impacts on human activity. Flood estimation is the process of minimizing property damage and reducing the threat to human activity. Design storm based on precipitation frequency analysis is one of the main procesess for flood estimation as well as a statistical representation of a precipitation event.

The primary purpose of precipitation frequency analysis in hydrometeorology is to estimate the magnitude of a storm event with a given frequency of occurrence. It can also be used to estimate the frequency of occurrence of a storm event with a given magnitude [1]. The precision of precipitation frequency analysis depends on the selection of an appropriate probability distribution model (PDM) and parameter estimation techniques. A number of PDMs have been developed to describe the probability distribution of the hydrometeorological variables. In practice, it is often assumed that the correct PDM is a member of the developed PDMs.

However, the selection of an appropriate PDM is still one of the major problems in precipitation frequency analysis [2].

For each of the developed PDMs, estimated parameters are provided based on alternative estimation techniques, such as the method of moments (MOM), probability weighted moments (PWM), linear function of ranked observations (L-moments), and maximum likelihood (ML) [1–5]. The MOM is one of the most simple and commonly used methods for estimating the parameters. PWM and L-moments are discussed by Stedinger et al. [6]. Estimates for PWM can be obtained from order statics. Additionally, Stedinger et al. [6] recommended a parameter estimator for PWM in regionalization studies. L-moments tend to produce less variable estimates for higher moments, when an unusually large or small observation happens to be present in a sample [1]. The moment estimators, including MOM, PWM, and L-moments, are a simpler way to obtain the PDM parameters but are less accurate than the ML estimators. The results using ML are generally more reliable than the other methods. However, the ML is more laborious than the other methods

because an iterative numerical solution, such as the Newton-Raphson method [7], must be used for estimating the parameters.

Metaheuristic approaches have been developed to solve various engineering optimization problems (e.g., linear and stochastic, dynamic, and nonlinear). These approaches include genetic algorithms [8, 9], ant colony optimization [10], simulated annealing [11], tabu searches [12], and evolutionary computation methods. Metaheuristic approaches use a stochastic random search instead of a gradient search so that intricate derivative information is unnecessary [13]. Therefore, the metaheuristic approaches have been shown to be a useful strategy to solve optimization problems in hydrology [5, 14–21].

Karahan et al. [22] used the genetic algorithm (GA) to predict rainfall intensities for a given set of return periods. The results showed that the proposed GA could be used to develop rainfall intensity-duration-frequency relationships with the lowest mean-squared error between the observed and predicted intensities. Karahan et al. [22] concluded that predicted intensities were in good agreement with the analyzed return period. Hassanzadeh et al. [23] studied the most suitable PDM for annual maximum discharge in East-Azerbaijan, Iran. Hassanzadeh et al. [23] employed the GA and ant colony optimization (ACO) techniques to find the parameters of PDMs. The performance of these algorithms was evaluated by comparison with conventional methods, such as MOM, ML, and PWM. The results showed that the GA and ACO were effective optimization tools compared to other methods for the parameter estimation of PDMs. The GA and ACO techniques could also be used for systems that are more complex and involve nonlinear optimization problems. However, the GA and ACO techniques need more computation time than the MOM, ML, and PWM methods to find the parameters of PDMs. Although the GA and ACO techniques are reasonable alternatives to solve hydrological problems, large computation time is an obvious disadvantage.

Therefore, the purpose of this study is to enhance the design storm estimation with improving parameter estimation of PDMs using a recently developed metaheuristic approach, a harmony search (HS) by Geem et al. [24]. The performance of the HS approach is compared with the GA and conventional methods (i.e., ML).

This paper is organized in the following manner. Section 2 introduces the methodology for the parameters of PDMs with statistical test criteria. The results of the simulation are shown in Section 3, and a case study is reported in Section 4. Finally, the summary and conclusions are presented in Section 5.

2. Methodology

2.1. Probability Distribution Models. To test the performance of the proposed parameter estimation approach, the two-parameter lognormal (LN2), two-parameter gamma (GAM2), generalized extreme value (GEV), and Gumbel (GUM) distribution models are used. The general properties of the PDMs are given in Appendices.
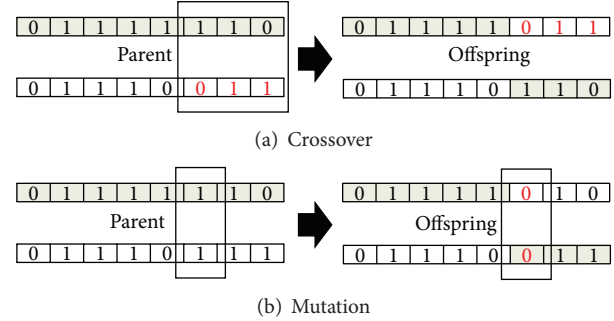

(a) Crossover

(b) Mutation

Figure 1: Illustration of crossover and mutation.

2.2. Metaheuristic Approaches. In the current study, two metaheuristic approaches (i.e., GA and HS) were used to estimate the parameters of PDMs.

2.2.1. Genetic Algorithms. The GA is a stochastic search algorithm based on natural evolution and mechanisms of population genetics [8, 9]. The simple ideas of GA are based on the biological processes of survival and adaption. Only the best of the population is allowed to survive and propagate to successive generations. Therefore, the GA does not require derivatives of the objective function to solve complex and discontinuous optimization problems. A number of GAs are introduced, but the following general description encompasses most of the important features.

The analogy with nature is established by the creation of a set of solutions, called a population. The initial population of solutions is usually chosen at random and allowed to evolve over a number of generations. Each individual in a population is represented by a set of parameter values that completely describe a solution and undergo constant change by means of genetic operations of reproduction, crossover, and mutation. At each generation, the fitness of individuals with respect to the objective function is calculated for reproduction and propagated to the next generation. Based on the fitness, individuals with relatively high fitness (called parent) are selected for reproduction of the next generation. For example, as shown in Figure 1(a), there are two parents selected with binary cording. The strings, including the last three digits of two parents (i.e., 110 and 011), are recombined to produce offspring that will comprise the next generation. Then, the parents are replaced in the population by the offspring to keep a stable population size. The recombination operation is usually called the crossover. The offsprings (new generations) have a higher average fitness than their parents (previous generations). Occasionally, mutation is introduced into the population to prevent the convergence to a local optimum and help generate unexpected directions in the parameter space, as shown in Figure 1(b). A fixed fitness of generations has been created when a generation has reached the highest fitness and there is no further improvement with repeated iteration [23].

The GA method has been widely applied in a variety of engineering optimization problems. It has been established that the GA approach is an attractive alternative to solve

optimization problems with nondifferentiable, nonlinear, and multimodal objective functions [9, 25, 26].

2.2.2. Harmony Search.

The HS developed by Geem et al. [24] is a phenomenon-mimicking algorithm inspired by the improvisational processes of musicians. The algorithm is based on natural musical performance processes in which a musician searches for a better state of harmony. Assume that the optimization problem is

$$\text{Minimize} \quad O_b(a),$$
$$\text{Subject to} \quad a_i \in \{\text{Low}_i, \text{Up}_i\}, \quad i = 1, \dots, N, \tag{1}$$

where $O_b(a)$ is an objective function; a is the set of each decision variable a_i; $i = 1, \dots, N$ (representing probability parameter); N is the number of decision variables (the number of parameters for a probability distribution); and Low_i and Up_i are the upper and lower limits, respectively, of the decision variable a_i.

To solve the optimization problem, the procedure of the HS is as follows.

(1) Generate the harmony memory (HM) randomly up to the harmony memory size (HMS) from a uniform distribution, denoted as

$$a_i^j \sim U\left[\text{Low}_i, \text{Up}_i\right], \tag{2}$$

where $i = 1, \dots, N$; $j = 1, \dots, \text{HMS}$; and $U[a, b]$ represents the uniform distribution ranging from a to b. The generated HM is

$$\text{HM} = \begin{bmatrix} a_1^1 & a_2^{11} & \cdots & a_N^1 \\ a_1^2 & a_2^2 & \cdots & a_N^2 \\ & & \vdots & \\ a_1^{\text{HMS}} & a_2^{\text{HMS}} & \cdots & a_N^{\text{HMS}} \end{bmatrix} = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^{\text{HMS}} \end{bmatrix}. \tag{3}$$

(2) Improvise a new set $(\tilde{a}_i)$ by

$$\tilde{a}_i = \begin{cases} \tilde{a}_i \in \{a_i^1, a_i^2, \dots, a_i^{\text{HMS}}\} & \text{w.p. HMCR} \\ \tilde{a}_i \sim U\left[\text{Low}_i, \text{Up}_i\right] & \text{otherwise,} \end{cases} \tag{4}$$

where HMCR is the harmony memory considering rate for all the variables $\tilde{a}_i$ and $i = 1, \dots, N$. Equation (4) implies that each decision variable of a new harmony set is sampled from the same variable of HM in (3) for the probability of HMCR. Otherwise, generate the harmony set from the uniform distribution as in (2).

(3) Adjust each variable of the improvised new set $(\tilde{a}_i)$ with the probability of the pitch adjusting rate (PAR) as

$$\tilde{a}_i^* = \begin{cases} \tilde{a}_i + \varepsilon & \text{w.p. PAR} \\ \tilde{a}_i & \text{otherwise,} \end{cases} \tag{5}$$

where ε is generated from $U[-h, h]$. Here, h is the arbitrary distance bandwidth. If h is large, the variability of the adjusted value $\tilde{a}_i^*$ is also large.

(4) Update the HM by replacing the worst harmony corresponding to the worst objective function with the improvised and adjusted new set.

(5) Repeat steps (2) to (4) until the termination criterion has been met.

Following the empirically based HS parameter range is recommended to produce a sufficient solution of 0.7–0.95 for HMCR, 0.2–0.5 for PAR, and 10–50 for HMS [24].

2.2.3. Optimization Function.

The derivative-based method usually uses the integral of the square of the error (ISE) as an optimization function to find a global solution because a derivative of ISE can be obtained relatively well. The metaheuristic approaches use a stochastic random search instead of a derivative search to find global optimization, so that various forms of the optimization function based on the integral of the absolute magnitude of the error (IAE), the integral of the time-absolute error (ITAE), and ISE are applied.

In this study, we use two metaheuristic approaches to estimate the parameters of PDMs. To attain the optimum result, an equivalent objective function (or target function) is used with two metaheuristic approaches. The objective function in this study is constructed using ISE with observed and estimated values and is expressed by

$$\text{Minimize} \quad \left(O_b = \sum_{i=1}^{n}\left(\frac{x_i - X_i}{x_i}\right)^2\right), \tag{6}$$

where O_b is an objective function; x_i is the ith ordered observation value; and X_i is the estimated value from the corresponding ith empirical cumulative probability of the selected PDM.

2.3. Statistical Test Criteria.

Three test criteria are used to assess the adequacy of the proposed parameter estimation method: the correlation coefficient (CC), coefficient of efficiency (CE), and root mean square error (RMSE) [27, 28]. The mathematical forms of these criteria are as follows:

$$\text{CC} = \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})(X_i - \overline{X})\right]}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}}, \tag{7}$$

where $\bar{x}$ and $\overline{X}$ are the mean of the observed data and the mean computed data from the fitted PDM, respectively. The range of CC is $-1 \le \text{CC} \le 1$. Here, the CC is also equivalent to the Filliben Q-Q correlation test, which was proposed by Filliben [29] as a test of normality. Vogel [30] extended this approach to lognormal and Gumbel distributions for the goodness-of-fit test and GEV [31]:

$$\text{CE} = 1 - \frac{\sum_{i=1}^{n}(x_i - X_i)^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}, \quad -\infty \le \text{CE} \le 1, \tag{8}$$

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - X_i)^2\right]^{1/2}, \quad 0 \le \text{RMSE} \le \infty.$$

TABLE 1: Strategy for the GA and HS.

Metaheuristic approach	Strategy	Value
Genetic algorithm	Population size	500
	Generations	1000
	Stall generation	100
	Selection	Stochastic uniform
	Crossover function	Scattered
	Mutation function	Gaussian
Harmony search	HMS	500
	Maximum iteration	100000
	HMCR	0.8
	PAR	0.4
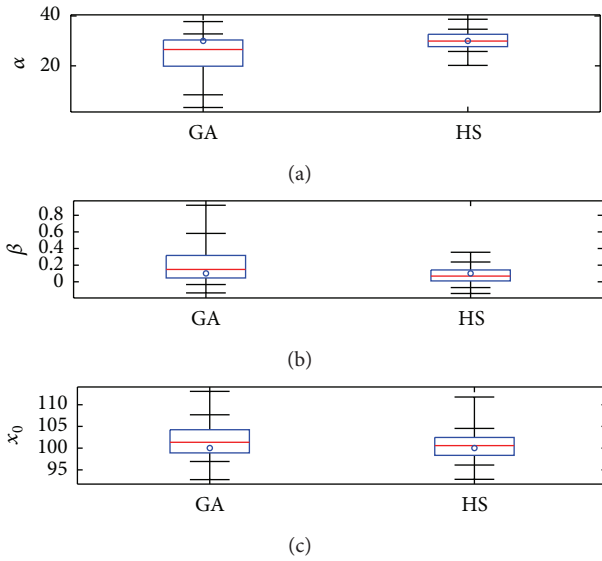
FIGURE 2: Boxplots of estimated GEV parameters using 100 simulations based on GA and HS; (a) scale parameter, α; (b) shape parameter, β; and (c) location parameter, x_0.

Note that higher CC and CE values and lower RMSE values represent better performance. Estimated statistics are described with boxplots. Boxes indicate the interquartile range (IQR), and whiskers extend to 1.5 IQR. The horizontal lines inside the boxes depict the median of the data.

3. Simulation Study

In the simulation study, we employ the GEV distribution model and estimate the parameters (i.e., scale (α), shape (β), and location (x_0)) using the GA and HS techniques. For the simulation study, we produce 100 data sets using a GEV random number. An individual data set with a record length of 100 has the same parameters, which are $\alpha = 30$, $\beta = 0.1$, and $x_0 = 100$. The strategies for the GA and HS are shown in Table 1.

One hundred series are simulated for the parameter estimation of the GEV distribution model. The results of the
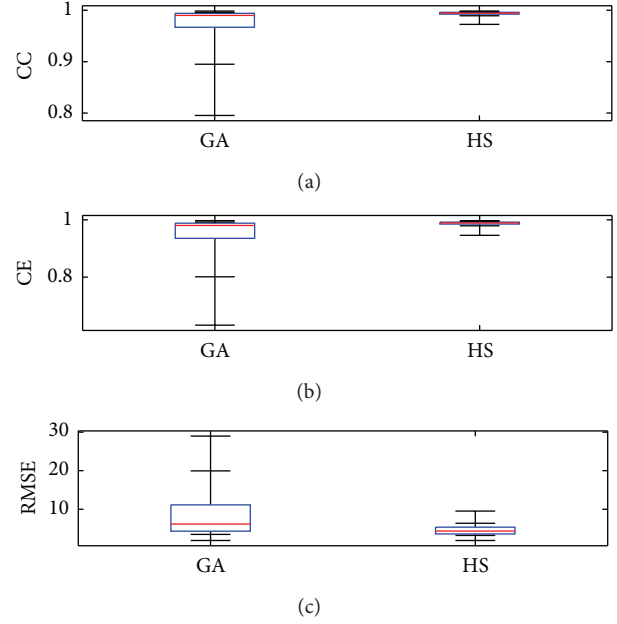

FIGURE 3: Boxplots of test criteria for estimated GEV parameters using 100 simulations based on GA and HS; (a) correlation coefficient; (b) coefficient of efficiency; and (c) root mean square error.

parameter estimation based on the GA and HS methods are compared using boxplots, as shown in Figure 2. As shown in Figures 2(a) and 2(b), the α and β derived from the HS indicate a better performance than those of the GA. In addition, the variability of α and β derived from the GA is greater than that of the HS. However, as shown in Figure 2(c), the x_0 derived from the HS and GA shows similar results.

The statistical test criteria in (7) and (8) were computed using the estimated parameters. In Figure 3, the test criteria for the proposed metaheuristic approaches are shown. Recall that higher CC and CE values and lower RMSE values represent better performance. As shown in Figure 3(a), the CCs for the HS are higher, while the variability is narrower than the CCs for the GA. The CE shows a similar pattern to the CC, as shown in Figure 3(b). The RMSE for the HS shows lower values than that of the GA, as shown in Figure 3(c). The RMSE indicates that the HS results have more accurate values than the GA results.

Although the metaheuristic approaches (e.g., the GA and ACO) are reasonable alternatives to solve hydrological problems, significant computation time is an obvious disadvantage [23]. Therefore, we investigate the computation time of two metaheuristic approaches. The computation time for GEV parameter estimation using 100 simulations based on the GA and HS is compared using boxplots, and the results are shown in Figure 4. The computation time for the GA is in the range of 10 to 20 minutes. However, the computation time for the HS is very short and is less than 1 minute. To determine computation time by varying conditions, the population size for the GA is changed to 100, 500, and 1000, while the harmony size for HS is also changed to 100, 500, and 1000. As shown in Figure 5, the computation time for the GA

TABLE 2: Target values for the parameters of employed PDMs and estimated parameters by the GA and HS.

Distribution	Parameters	Value		
		Target	GA	HS
LN2	Location	4.7	4.60~4.81* (4.70)**	4.60~4.81 (4.70)
	Scale	0.4	0.32~0.50 (0.40)	0.32~0.50 (0.40)
GAM2	Shape	7	4.15~11.38 (7.17)	4.03~11.38 (7.16)
	Scale	17	10.29~30.49 (17.16)	10.29~31.27 (17.18)
GUM	Location	140	125.08~147.08 (136.93)	125.06~147.04 (136.93)
	Scale	30	24.18~37.73 (29.92)	24.16~37.80 (29.92)

*The values are presented in the range of min to max.
**() shows mean of values.

TABLE 3: Comparison of test criteria for the three PDMs.

Distribution	Criteria	Metaheuristic approach	
		GA	HS
LN2		0.96~1.00* (0.99)**	0.96~1.00 (0.99)
GAM2	CC	0.97~1.00 (0.99)	0.97~1.00 (0.99)
GUM		0.73~0.94 (0.89)	0.73~0.94 (0.89)
LN2		0.92~1.00 (0.98)	0.92~1.00 (0.98)
GAM2	CE	0.93~1.00 (0.99)	0.93~1.00 (0.99)
GUM		0.54~0.89 (0.79)	0.54~0.89 (0.79)
LN2		2.92~16.67 (6.18)	2.92~16.67 (6.18)
GAM2	RMSE	2.43~13.33 (4.95)	2.43~13.29 (4.96)
GUM		11.32~44.12 (19.62)	11.32~44.12 (19.62)
LN2		23.76~52.25 (33.86)	6.19~6.32 (6.22)
GAM2	Computation time (sec)	406.49~1531.10 (793.46)	77.92~83.70 (80.16)
GUM		21.99~30.07 (25.28)	5.33~5.50 (5.40)

*The values are presented in the range of min. to max.
**() shows mean of values.

rapidly increases as the population size increases. Meanwhile, the fitness values of the objective function are only slightly improved. In contrast, the computation time for the HS rarely increases despite increasing harmony size, and the fitness values of the objective function remain consistent. Note that the fitness values of the objective function for HS with a small harmony size are better than the fitness values of the GA. The HS approach for estimation of the GEV parameter produces more reliable results than the GA approach and uses less computation time.

In addition, three PDMs (i.e., LN2, GAM2, and GUM) are used in a simulation study. The same procedure as the simulation study using GEV is employed for the parameter estimation of each PDM. The target values for the parameters of the applied PDMs and the estimated parameters by the GA and HS methods are reported in Table 2. In addition, the results of the three test criteria and the computation time for each PDM are summarized in Table 3. The results of additional simulation studies show that the two metaheuristic approaches are appropriate methods for the parameter estimation of PDMs. Furthermore, the difference between the GA and HS methods is that the HS method requires less computation time than the GA method while still providing reliable results.

4. Case Study

In the current case study, we carried out precipitation frequency analysis for design storm with annual maximum hourly rainfall data recorded at 74 rainfall gauges in South Korea. The annual maximum hourly rainfall data were extracted from the Korea Meteorological Administration website, http://www.kma.go.kr/. The record lengths of extracted rainfall data range from 20 to 100 years (average record length is approximately 40 years), and the locations of the 74 rainfall gauges are presented in Figure 6.

Four PDMs (i.e., LN2, GAM2, GEV, and GUM) were applied to the annual maximum hourly rainfall data. Parameters corresponding to the four PDMs were estimated using the three parameter estimation approaches: ML, GA, and HS. To compare the three parameter estimation approaches, the
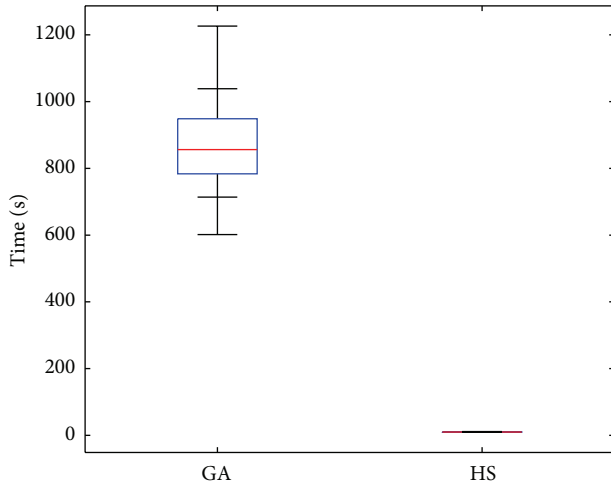
FIGURE 4: Boxplot of computation time for GEV parameter estimation using 100 simulations based on GA and HS techniques.
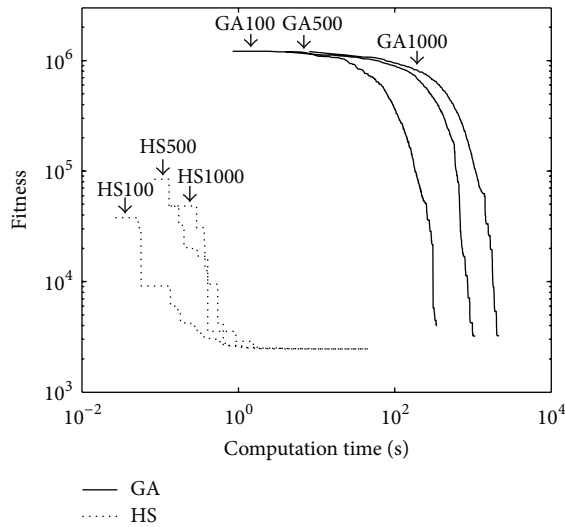

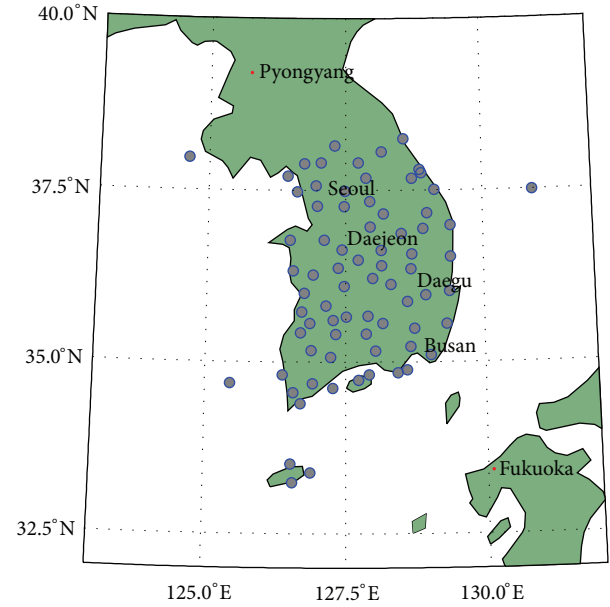FIGURE 5: Comparison of computation time for GEV parameter estimation based on GA and HS methods.

FIGURE 6: The locations of the employed 74 rain gauges in Republic of Korea.

The RMSE indicates that the two metaheuristic approaches for the parameter estimation of PDMs have a more accurate performance than ML.

In addition, the results of parameter estimation for Seoul, Daejeon, Daegu, and Busan, which are four major cities of South Korea, are summarized in Table 5. Furthermore, values of the quantiles for the four cities are estimated, and the results are summarized in Table 6.

5. Summary and Conclusion

The HS was developed by Geem et al. [24] and has been applied to solve a variety of engineering problems in a number of previous studies. Because the HS adopts a stochastic random search to find the optimized best solution, initial value settings of decision variables and derivative information to reach global optimum are not required. Furthermore, after considering all of the existing vectors based on the HMCR and PAR, the HS generates a new vector, whereas the GA only considers two vectors. These features increase the flexibility of the HS and produce a better solution. Therefore, we applied the HS method to enhance design storm estimation in the current study.

The results of the simulation study based on the four PDMs (i.e., LN2, GAM2, GEV, and GUM) showed that the HS approach for the parameter estimation of PDMs produced reliable results when measuring test criteria (i.e., CC, CE, and RMSE). Furthermore, the fitness values of the objective function for HS were better than the fitness values of the GA with a small harmony size. Additionally, the computation time for the HS approach was less than that of the GA method.

Precipitation frequency analysis for design storm was conducted to assess the performance of the proposed method

quantiles for different return periods ($T = 10, 25, 50,$ and 100 years) were estimated at the 74 rain gauges in South Korea. The variability of the quantiles for each PDM is shown in Figure 7. Note that quantiles imply the statistically estimated design storm at a given return period.

As shown in Figure 7, the results of frequency analysis for the employed rainfall data show that the quantiles estimated by the two metaheuristic approaches present similar distributions, except for the quantiles of $T = 50, 100$ for LN2, as shown in Figure 7(a). However, the quantiles based on the ML method show a slightly different distribution when compared with the two metaheuristic approaches.

Table 4 summarizes the results of test criteria derived from the three parameter estimation approaches. The CCs for ML, GA, and HS represent similar values, and the CE represents a similar pattern to that of CC. However, RMSE for GA and HS shows lower values than the results of ML.
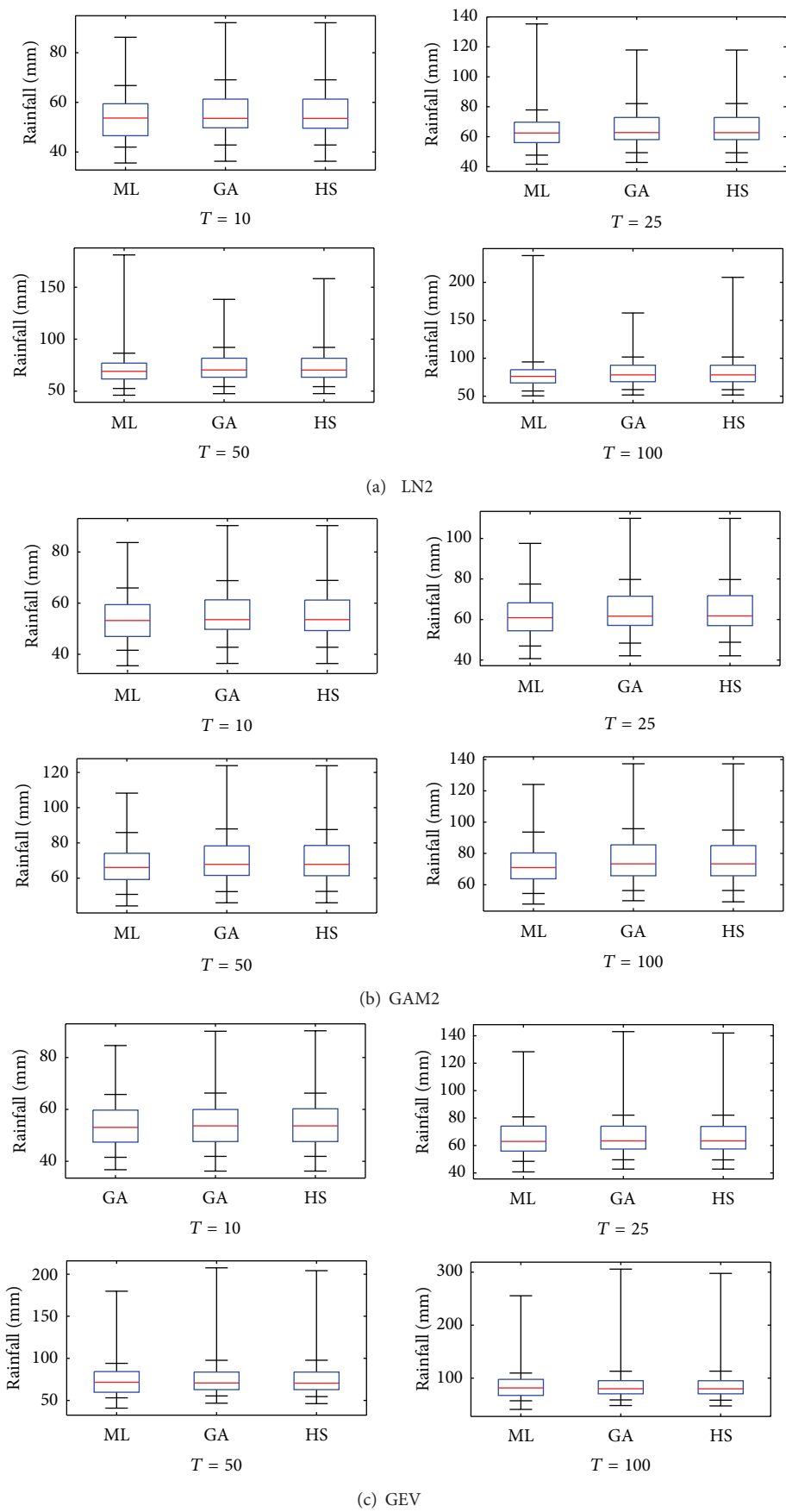
(a) LN2

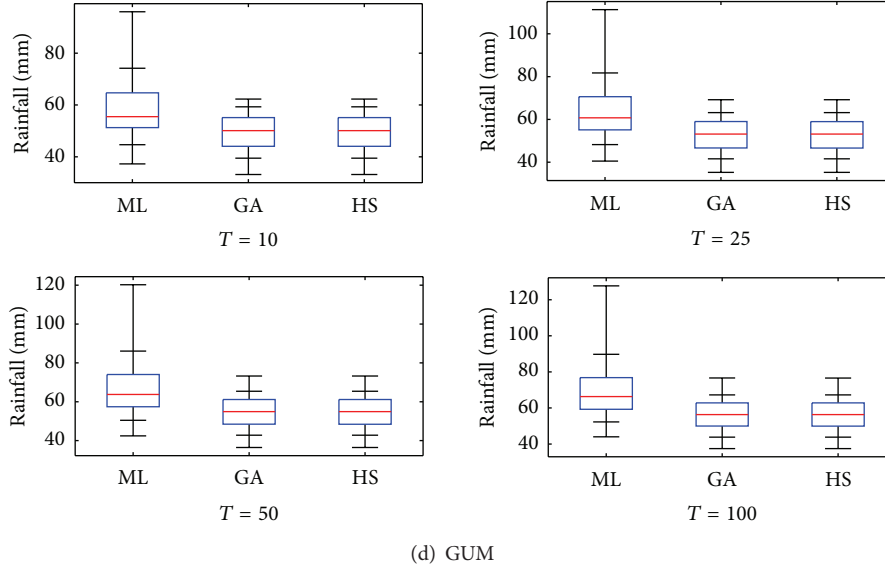(b) GAM2

(c) GEV

Figure 7: Continued.

(d) GUM

Figure 7: Quantiles using four PDMs at 74 stations in Republic of Korea. (a) LN2, (b) GAM2, (c) GEV, and (d) GUM.

with the four PDMs (i.e., LN2, GAM2, GEV, and GUM) and the annual maximum hourly rainfall data recorded at 74 rainfall gauges in South Korea. The results of precipitation frequency analysis were compared with those of the ML and GA methods. The results in this study suggested that performances of the GA and HS were similar and presented more accurate results than the ML in precipitation frequency analysis for annual maximum hourly rainfall data. Accordingly, we conclude that the proposed parameter estimation method based on the HS approach is a useful alternative for the parameter estimation of PDMs, particularly when conventional methods cannot be applied to estimate the parameters of PDMs.

Appendices

Probability Distribution Models

A. Two-Parameter Lognormal (LN2) Distribution

The transformation of random variable X is

$$Y = \log_a (X), \tag{A.1}$$

where a is the base of the logarithm.

Assuming that the mean and variance of Y are μ_y and σ_y^2, respectively, then

$$f(x) = \frac{k}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\log_a (x) - \mu_y}{\sigma_y} \right)^2 \right] \tag{A.2}$$

is the probability density function (PDF, i.e., $f_X(x) = dF_X(x)/dx$) of the LN2. The cumulative distribution function (CDF, i.e., $F(x) = P(X \le x)$) of the LN2 is

$$F_X (x) = \int_{-\infty}^{y} \frac{k}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\log_a (x) - \mu_y}{\sigma_y} \right)^2 \right] dx. \tag{A.3}$$

Note that $k = 1$ if $a = e$ (base-e logarithm) and $k = \log_{10}(e)$ if $a = 10$ (base-10 logarithm). In relation to the random variable X, μ_y controls the scale and is called the scale parameter, while σ_y controls the skewness and is regarded as a shape parameter. Additionally, note that in relation to the variable Y, μ_y is the location parameter, and σ_y is the scale parameter.

The quantile function for LN2 corresponding to the nonexceedance probability q is

$$x_T = \exp_a \left(\mu_y + z\sigma_y \right), \tag{A.4}$$

where x_T is a quantile of return period T for LN2 and z is the standard normal variate corresponding to the non-exceedance probability q.

B. Two-Parameter Gamma (GAM2) Distribution

The PDF of the GAM2 is

$$f(x) = \frac{1}{|\alpha|\,\Gamma(\beta)} \left(\frac{x}{\alpha} \right)^{\beta-1} \exp\left(-\frac{x}{\alpha} \right),$$

$$\Gamma(\beta) = \int_0^{\infty} z^{\beta-1} e^{-z} dz, \tag{B.1}$$

TABLE 4: Comparison of test criteria for the annual maximum rainfall data for the 74 rain gauges stations in Republic of Korea.

Criteria	Distribution type	Mean			Std.		
		ML	GA	HS	ML	GA	HS
CC	LN2	0.98	0.98	0.98	0.023	0.017	0.018
	GAM2	0.97	0.98	0.97	0.027	0.024	0.024
	GEV	0.98	0.99	0.99	0.019	0.008	0.008
	GUM	0.88	0.88	0.88	0.059	0.059	0.059
CE	LN2	0.95	0.96	0.96	0.054	0.034	0.044
	GAM2	0.94	0.95	0.95	0.054	0.046	0.046
	GEV	0.96	0.98	0.98	0.054	0.016	0.016
	GUM	0.85	0.77	0.77	0.523	0.114	0.114
RMSE	LN2	3.01	2.60	2.65	2.147	1.623	1.815
	GAM2	3.12	2.89	2.90	2.133	1.941	1.937
	GEV	2.52	1.94	1.94	1.550	0.833	0.838
	GUM	10.55	6.48	6.48	6.368	3.302	3.302

TABLE 5: The results of the parameter estimation for the four major cities in Republic of Korea.

Rain gauge	Distribution	ML			GA			HS		
		Shape	Scale	Location	Shape	Scale	Location	Shape	Scale	Location
Seoul	LN2	—	0.431	3.684	—	0.445	3.680	—	0.445	3.680
	GAM2	5.550	7.871	—	4.638	9.401	—	4.635	9.406	—
	GEV	0.056	14.178	34.684	0.089	13.943	34.450	0.088	13.964	34.453
	GUM	—	25.898	54.769	—	13.574	51.443	—	13.579	51.438
Daejeon	LN2	—	0.290	3.617	—	0.281	3.622	—	0.281	3.622
	GAM2	12.348	3.140	—	11.990	3.239	—	12.141	3.205	—
	GEV	−0.162	9.959	34.307	−0.166	10.309	34.314	−0.170	10.337	34.335
	GUM	—	11.084	44.388	—	8.362	43.506	—	8.356	43.503
Daegu	LN2	—	0.374	3.419	—	0.402	3.410	—	0.402	3.410
	GAM2	7.065	4.647	—	5.751	5.693	—	5.803	5.653	—
	GEV	0.117	9.034	26.461	0.086	9.441	26.593	0.083	9.486	26.599
	GUM	—	17.510	40.272	—	9.075	38.022	—	9.079	38.022
Busan	LN2	—	0.483	3.615	—	0.406	3.641	—	0.404	3.642
	GAM2	5.112	8.040	—	5.419	7.605	—	5.416	7.614	—
	GEV	−0.085	14.902	33.590	−0.090	15.273	33.591	−0.093	15.317	33.608
	GUM	—	19.020	50.318	—	12.878	48.438	—	12.873	48.436

where $0 \leq x < \infty$ for $\alpha > 0$ and $-\infty < x \leq 0$ for $\alpha < 0$. The parameters α and β are the scale and shape parameters. The shape parameter β is restricted to $\beta > 0$, while α may be positive or negative. $\Gamma(\beta)$ is the complete gamma function, which is the integral function. From (B.1), the CDF of GAM2 is derived to be

$$F(x) = P(\beta, z) = \frac{1}{\alpha\Gamma(\beta)}$$
$$\times \int_0^x \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x}{\alpha}\right) dz, \quad \text{for } \alpha > 0,$$
$$F(x) = P(\beta, z) = 1 - \frac{1}{\alpha\Gamma(\beta)}$$
$$\times \int_0^x \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x}{\alpha}\right) dz, \quad \text{for } \alpha < 0.$$
(B.2)

The function $P(\beta, z)$ in (B.2) is the incomplete gamma function. Therefore, the quantile for GAM2 must be estimated numerically. The quantile x_T of return period T for GAM2 corresponding to the nonexceedance probability q is alternatively estimated using the following:

$$x_T = \mu_x + K_T\sigma_x,$$
(B.3)

where μ_x and σ_x are the mean and standard deviation, respectively. K_T is the frequency factor, which is a function of return period T and the skewness coefficient [2].

TABLE 6: Quantiles for different return periods (T = 10, 25, 50, and 100 years) of the four major cities in Republic of Korea.

Rain gauge	Distribution	$T = 10$			$T = 25$			$T = 50$			$T = 100$		
		ML	GA	HS	ML	GA	HS	ML	GA	HS	ML	GA	HS
Seoul	LN2	69.2	70.1	70.1	84.7	86.3	86.4	96.5	98.8	98.8	108.6	111.5	111.6
	GAM2	68.5	70.7	70.7	80.9	84.6	84.6	89.6	94.4	94.4	97.9	103.9	103.9
	GEV	68.7	69.2	69.2	84.3	86	86	96.5	99.5	99.4	109	113.7	113.6
	GUM	76.4	62.8	62.8	85	67.3	67.3	85	67.3	67.3	94.3	72.2	72.2
Daejeon	LN2	53.9	53.6	53.6	61.8	61.1	61.1	67.4	66.6	66.6	73	71.8	71.8
	GAM2	53.4	53.7	53.7	60.1	60.5	60.5	64.6	65.2	65.1	68.9	69.6	69.5
	GEV	53.1	53.7	53.7	59.2	59.9	59.8	63.1	63.9	63.8	66.6	67.5	67.3
	GUM	53.6	50.5	50.5	57.3	53.3	53.3	57.3	53.3	53.3	61.3	56.3	56.3
Daegu	LN2	49.3	50.7	50.7	58.8	61.2	61.2	65.8	69.1	69.2	72.9	77.2	77.2
	GAM2	49.3	51.0	51.0	57.3	60.1	60.0	62.9	66.4	66.4	68.1	72.5	72.4
	GEV	49.7	50.0	50.1	61.5	61.4	61.3	71.2	70.4	70.3	81.5	79.9	79.7
	GUM	54.9	45.6	45.6	60.7	48.6	48.6	64.2	50.4	50.4	67.0	51.9	51.9
Busan	LN2	69.0	64.1	64.1	86.6	77.6	77.4	100.2	87.7	87.5	114.3	98.0	97.7
	GAM2	65.4	64.9	64.9	77.7	76.8	76.8	86.4	85.1	85.2	94.7	93.1	93.2
	GEV	64.1	64.7	64.7	75.3	76.0	76.0	83.0	83.8	83.7	90.3	91.1	90.9
	GUM	66.2	59.2	59.2	72.6	63.5	63.5	72.6	63.5	63.5	79.4	68.1	68.1

C. Generalized Extreme Value (GEV) and Gumbel (GUM) Distribution

The PDF and CDF of the GEV for a random variable x are expressed as follows:

$$f(x) = \frac{1}{\alpha}\left[1 - \beta\left(\frac{x - x_0}{\alpha}\right)\right]^{(1/\beta)-1} F(x)$$
$$F(x) = \exp\left\{-\left[1 - \beta\left(\frac{x - x_0}{\alpha}\right)\right]^{1/\beta}\right\}, \tag{C.1}$$

where α, β, and x_0 are the scale, shape, and location parameter, respectively. β plays an important role such that if $\beta = 0$, the distribution tends to resemble a type-1 or Gumbel distribution; if $\beta < 0$, the resulting distribution is type-2 or Log-Gumbel distribution; and if $\beta > 0$, it is a type-3 distribution or Weibull distribution. The quantile functions for GEV and Gumbel corresponding to the non-exceedance probability q are given in the following:

$$x_T = x_0 + \frac{\alpha}{\beta}\left\{1 - \left[-ln\left(1 - \frac{1}{T}\right)\right]^{\beta}\right\} \quad \text{for GEV,}$$
$$x_T = x_0 - \alpha \ln\left[-\ln\left(1 - \frac{1}{T}\right)\right] \quad \text{for Gumbel,} \tag{C.2}$$

where x_T is a quantile of return period T for GEV and Gumbel.

Acknowledgment

References

[1] C. T. Haan, *Statistical Methods in Hydrology*, Iowa State Press, Ames, Iowa, USA, 2002.

[2] J. D. Salas, R. A. Smith, G. Q. Tabios, and J. H. Heo, in *Statistical Computing Techniques in Water Resources and Environmental Engineering*, Lecture Note, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, Colo, USA, 2008.

[3] J. R. M. Hosking, "The theory of probability weighted moments," Research Report RC12210, IBM Thomas J. Watson Research Center, New York, NY, USA, 1986.

[4] A. R. Rao and K. H. Hamed, *Flood Frequency Analysis*, CRC Press, New York, NY, USA, 2000.

[5] N. T. Kottegoda and R. Rosso, *Applied Statistics for Civil and Environmental Engineers*, Wiley-Blackwell, West Sussex, UK, 2008.

[6] J. R. Stedinger, R. M. Vogel, and E. Foufoula-Georious, "Frequency analysis of extream events," in *Handbook of Hydrology*, chapter 18, McGraw-Hill, New York, NY, USA, 1994.

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, Mass, USA, 1986.

[8] J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York, NY, USA, 1989.

[10] M. Dorigo and T. Stützle, *Ant Colony Optimization*, MIT Press, Hong Kong, 2004.

[11] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[12] F. Glover and C. McMillan, "The general employee scheduling problem. An integration of MS and AI," *Computers and Operations Research*, vol. 13, no. 5, pp. 563–573, 1986.

[13] K. S. Lee and Z. W. Geem, "A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory

and practice," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 36–38, pp. 3902–3933, 2005.

[14] K. C. Abbaspour, R. Schulin, and M. T. van Genuchten, "Estimating unsaturated soil hydraulic parameters using ant colony optimization," *Advances in Water Resources*, vol. 24, no. 8, pp. 827–841, 2001.

[15] H. R. Maier, A. R. Simpson, W. K. Foong, K. Y. Phang, H. Y. Seah, and C. L. Tan, "Ant colony optimization for the design of water distribution systems," in *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, vol. 111, American Society of Civil Engineers, 2001.

[16] A. C. Zecchin, H. R. Maier, A. R. Simpson, M. Leonard, and J. B. Nixon, "Ant colony optimization applied to water distribution system design: comparative study of five algorithms," *Journal of Water Resources Planning and Management*, vol. 133, no. 1, pp. 87–92, 2007.

[17] S.-H. Dong, "Genetic algorithm based parameter estimation of Nash model," *Water Resources Management*, vol. 22, no. 4, pp. 525–533, 2008.

[18] J. Reca, J. Martínez, C. Gil, and R. Baños, "Application of several meta-heuristic techniques to the optimization of real looped water distribution networks," *Water Resources Management*, vol. 22, no. 10, pp. 1367–1379, 2008.

[19] V. Nourani, S. Talatahari, P. Monadjemi, and S. Shahradfar, "Application of ant colony optimization to optimal design of open channels," *Journal of Hydraulic Research*, vol. 47, no. 5, pp. 656–665, 2009.

[20] R. K. Rai, S. Sarkar, and V. P. Singh, "Evaluation of the adequacy of statistical distribution functions for deriving unit hydrograph," *Water Resources Management*, vol. 23, no. 5, pp. 899–929, 2009.

[21] M. Janga Reddy and S. Adarsh, "Chance constrained optimal design of composite channels using meta-heuristic techniques," *Water Resources Management*, vol. 24, no. 10, pp. 2221–2235, 2010.

[22] H. Karahan, H. Ceylan, and M. Tamer Ayvaz, "Predicting rainfall intensity using a genetic algorithm approach," *Hydrological Processes*, vol. 21, no. 4, pp. 470–475, 2007.

[23] Y. Hassanzadeh, A. Abdi, S. Talatahari, and V. P. Singh, "Meta-heuristic algorithms for hydrologic frequency analysis," *Water Resources Management*, vol. 25, no. 7, pp. 1855–1879, 2011.

[24] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, 2001.

[25] A. M. S. Zalzala and P. J. Fleming, *Genetic Algorithms in Engineering Systems*, The Institution of Engineering and Technology, London, UK, 1999.

[26] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Sons, New York, NY, USA, 1999.

[27] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models—part I: a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.

[28] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," *Journal of Hydrology*, vol. 374, no. 3-4, pp. 294–306, 2009.

[29] J. J. Filliben, "Probability plot correlation coefficient test for normality," *Technometrics*, vol. 17, no. 1, pp. 111–117, 1975.

[30] R. M. Vogel, "The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses," *Water Resources Research*, vol. 22, no. 4, pp. 587–590, 1986.

[31] J. U. Chowdhury, J. R. Stedinger, and L.-H. Lu, "Goodness-of-fit tests for regional generalized extreme value flood distributions," *Water Resources Research*, vol. 27, no. 7, pp. 1765–1776, 1991.