



1 **Application of Gamma functions to the determination of** 2 **unit hydrographs**

3 Hongyan Li ^[1,*], Yangzong Cidan ^[2]

4 [1] *Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University,*
5 *Changchun, 130021, China; lihongyan@jlu.edu.cn;*

6 [2] *Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University,*
7 *Changchun, 130021, China; aerlin321@163.com*

8 * Correspondence: lihongyan@jlu.edu.cn

9
10 **Abstract:** There are many methods for calculating unit hydrograph, such as analysis method, trial algorithm and
11 least squares method. But these methods have certain requirements for flood data and the unit hydrograph may
12 not be optimal. Based on the theory of composition, a hydrological system was viewed as a **generalized collection**
13 in this study and Gamma functions were used to simulate the basin convergence process. At the same time, the
14 Gamma function is parameterized and the parameters of Gamma function are optimized by **genetical algorithm**,
15 **which is based on the minimum error between the calculation of confluence process and the measurement process**,
16 before deriving the unit hydrograph. The Collins iteration method was used to compute the unit hydrograph. The
17 results of actual calculated examples **showe** that this method is more precise than other methods, while it **can also**
18 illustrate the **law of runoff**.

19 **Keywords:** Theory of composition; Gamma function; Genetic algorithm; Unit hydrograph; Collins iteration
20 method

21 **0. Introduction**

22 The unit hydrograph is an important method for simulating the flow concentration **of a**
23 conceptual hydrological model; it was proposed by Sherman (1932). In actual applications, the
24 derivation of the unit hydrograph is still an important component when forecasting the basin
25 rainfall and runoff.

26 A unit hydrograph (Viessman, 1989; Raghunath, 2006) refers to the unit net constant
27 rainfall uniformly distributed over a watershed of unit surface and for unit duration. Periods of 1, 3,
28 6, and 12 h can be selected and the unit rainfall (runoff depth) is generally 10 mm. The actual net
29 rainfall often does not equal 1 unit for these time periods, so it is necessary to make two basic
30 assumptions when calculating the watershed flow **concentration process**.

31 (a) Assumption of linear hydrological system: if the unit period rainfall is k units, the
32 formation of the flow process is k times the unit hydrograph ordinate.

33 (b) Superposition assumption: if the rainfall lasts m periods, the formation of the flow process
34 is the superposition of each rainfall period.

35 Based on the above assumptions, the basin outlet section discharge hydrograph can be
36 expressed as:



$$Q_i = \sum_{j=1}^m h_j q_j \begin{cases} i = 1, 2, \dots, l \\ j = 1, 2, \dots, m \end{cases} \quad (1)$$

$$\left. \begin{matrix} i - j + 1 = 1, 2, \dots, n \end{matrix} \right\}$$

where Q_i is the basin outlet section of each period discharge in m³/s;

h_j is the rainfall in each period in mm;

q_{i-j+1} is the ordinate of the unit hydrograph in each period in m³/s;

i is the number of periods for the basin outlet section flow hydrograph;

j is the net number of rainfall periods; and

In essence, the unit hydrograph is the characteristic of watershed concentration in the form of discharge hydrograph, i.e. concentration curve (d.johnstone, 1949). The calculation method of unit hydrograph confluence takes the basin as a whole and assumes that the net rain is uniformly distributed over the whole basin, without considering the inhomogeneity within the system; the basin confluence system is a linear time-invariant system, and at the same time, it is viewed that the net rainfall and the formation of the flow process are in agreement to superposition relationship. Therefore, the essential characteristics of the unit hydrograph are lumped resistance, linearity, and time invariance.

Conceptually, the unit hydrograph is a linear time-invariant basin system with a convergent flow curve. However, the physical mechanism of the watershed conflux is not considered by the method used to derive the unit hydrograph (Ramirez, 2000). The principle used to calculate the unit hydrograph is based on the system input (rainfall), which is converted using the unit hydrograph to determine the system response output (outlet section flow process), where the error is minimized. The traditional methods of derivation are as follows.

Analytical approach: the basin outlet section of the surface runoff is Q_1, Q_2, \dots, Q_l , the rainfall process is h_1, h_2, \dots, h_m , where Eqn. (1) comprises q_1, q_2, \dots, q_n unknown linear algebraic equations. The solutions of the equations can be obtained using a unit hydrograph.

$$Q_i = \sum_{j=1}^m \frac{h_j q_j}{10^{i-j+1}} \begin{cases} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{cases} \quad (2)$$

where n is the number of periods of the unit hydrograph and $n = l - m + 1$.

In theory, there are no errors with the analytical approach when using the rainfall runoff measurements. This approach can obtain the correct answer if the watershed conflux conforms to a



64 linear time-invariant system (when the convergence of watershed system meets multiple
65 proportions and superposition assumption, for a linear system). However, (1) the data
66 measurements have errors and, (2) rainfall-runoff system is not a linear time-invariant system. Due
67 to the accumulation of errors, unreasonable solutions often occur. For example, the unit
68 hydrograph may be irregular or appear to be negative.

69 (b) Trial and error: the unit hydrograph is assumed to be q'_i . The flow Q' determined by
70 the unit hydrograph is then compared with the measured flow Q . The unit hydrograph q'_i is
71 produced when the error between Q' and Q satisfies certain error.

72 The trial and error method proposed by Collins (Collins, 1939) uses an iterative strategy. For
73 a period of uneven net rainfall, the computational convergence is fast if the time period is large.
74 However, the Collins iterative method has the following disadvantages: (1) the iterations are based
75 on the initial unit hydrograph, but there is no strict method for selecting the initial unit hydrograph;
76 and (2) trial-error approach does not necessarily converge to a solution.

77 (c) Least squares method: the measured surface runoff is assumed to be Q and the error is
78 as follows: $\varepsilon = Q - Q'$. If $\sum \varepsilon^2 = \sum (Q - Q')^2 \rightarrow \min$, we try to convert Eqn. (1) into a
79 normal equations system where the number of equations is equal to the variables. The optimal
80 estimation of q'_i can be solved using the least squares method. The theory of the least squares
81 method is better, but the results are sometimes fluctuating or negative.

82 There are many methods for determining unit hydrographs, e.g., the Z transform method and
83 the harmonic analysis method (Dodge, 1973). A previous study (Hanson and Johnson 1964)
84 classified and compared the usual unit hydrograph calculation methods.

85 In recent decades, the use of probability distribution functions (pdfs) to develop synthetic unit
86 hydrographs (SUH) has received much attention because of its similar properties to unit
87 hydrographs. First, the type of unit hydrograph needs to be subjected to mathematical analysis.
88 Typical functions, such as a parabola P-III (Yuan, et al., 1991) (Zhai and Li, 2004) curve, can be
89 used to describe the unit line and a mathematical model of the unit hydrograph can be established.
90 A previous study (Bhunya, et al., 2007) explored the potential of using four popular pdfs, i.e.,
91 two-parameter Gamma, three-parameter Beta, two-parameter Weibull, and one-parameter
92 Chi-square distribution, for deriving a SUH. The Gamma functions are the most widely used
93 functions (Singh, 2005, 2009; Bhunya, et al., 2003). This approach aims to determine the
94 relationship between each variable in a unit hydrograph, which facilitates a more in-depth analysis
95 of a unit hydrograph.

96 In the present study, we used Gamma functions to describe a unit hydrograph and determined
97 why a unit hydrograph may follow this distribution. The Gamma function parameters were



98 optimized using a genetic algorithm. Finally, the unit hydrograph was obtained using the Collins
99 iterative method.

100 For a specific basin, the confluence time of a flood is relatively stable and can be determined
101 according to the flood datas. Therefore, the use of gamma function to derive unit hydrograph is
102 only in the trial calculation of parameters β and k . The unit hydrograph expressed by gamma
103 function based on the combination of these parameters is optimal, which is compared with the use
104 of P-III function to adjust its statistical parameters. The principle of hydrological frequency
105 calculation by line fitness of numbers (mean, variation coefficient C_v and skewness coefficient C_s)
106 is very similar. Meanwhile, the time interval t , parameters and values of gamma function can
107 parameterize the regional characteristics of river basin confluence, which has the important
108 significance for this study.

109 1 Mathematical model and method

110 1.1 Mathematical model

111 In spatial or temporal physical entropy-based modeling of hydrology and water resources, the
112 cumulative distribution function (CDF) of a design variable (e.g., a flux or a discharge) is
113 analyzed in terms of its concentration (e.g., stage of flow) (Cui, et al., 2012). The theory of
114 composition proposed by Zhang (2003) provides a model and a uniform calculation method for
115 studying the composition of things. This theory considers the analysis of three concepts, i.e., the
116 general set, the distribution function, and the degree of complexity. This theory is also considered
117 the most highly approved principle followed by random systems, i.e., the entropy principle.

118 The variable x is continuous and random, and can be viewed as a general set of flag
119 variables. If the pdf $f(x)$ of x agrees with the following function:

$$120 \quad f(x) = \frac{\beta^k}{(k-1)!} x^{k-1} e^{-\beta x}, x > 0 \quad (3)$$

121 Then the pdf follows a Gamma distribution, where β and k are shape and scale parameters.
122 This is one of the famous Pearson pdfs, which is known as a Pearson type III distribution. The
123 curve has a peak with a left-right asymmetry. In nature, many phenomena follow this distribution.
124 In China, hydrological studies often use the Pearson type III distribution to simulate hydrological
125 data series, because it has a greater than or equal to zero lower bound on the variable requirements
126 and its elasticity is greater than the normal distribution (Ye and Xia, 2002). This choice is based
127 on experience, but it lacks a theoretical justification.

128 Using entropy theory, a previous study (Zhang, 2003) described the physical form of this
129 distribution. By analyzing the structure of Eqn. (3), is not difficult to show that it has a negative
130 exponential distribution, which is a part of the exponential function, and it also has the
131 characteristics of a power function in a Pareto-family distribution. The exponential distribution



132 corresponds to the constraints on the invariant algebraic average of the flag variables, while the
 133 power function corresponds to the constraints on the invariant geometric mean. It may be
 134 speculated that the constraints on the Gamma distribution are the fixed algebraic average and
 135 geometric means of the variables.

136 In this study, $f(x)$ is the pdf of a positively defined random variable, i.e.,

$$137 \int_0^{\infty} f(x) \quad (4)$$

138 u represents the algebraic average of variables, thus

$$139 u = \int_0^{\infty} xf(x)dx \quad (5)$$

140 while v is the geometric mean of the random variable x , which can be expressed as the
 141 algebraic average of the logarithm, i.e.,

$$142 v = \int_0^{\infty} \ln xf(x)dx \quad (6)$$

143 The entropy of the random variable x can be written as

$$144 H = - \int_0^{\infty} f(x) \ln xf(x)dx \quad (7)$$

145 Given the constraints in Eqns. (4), (5), and (6), the Lagrange method can be used to estimate
 146 the distribution function F based on the maximum entropy to determine the distribution
 147 function. Thus, F is defined as follows:

$$148 F = - \int_0^{\infty} f \ln f dx + C_1 \left(\int_0^{\infty} f dx - 1 \right) + C_2 \left(\int_0^{\infty} xf dx - u \right) + C_3 \left(\int_0^{\infty} \ln xf dx - v \right) \quad (8)$$

149 Where, C_1 , C_2 , and C_3 are undetermined constants. The entropy principle demands that the
 150 value of F is maximal. The partial derivative of $f(\bullet)$, i.e., the partial derivative is 0, can be
 151 obtained using Eqn (8). The results are as follows.

$$152 f(x) = \exp(C_1 - 1 + C_2 x + C_3 \ln x) \quad (9)$$

153 This formula can be used to obtain the distribution function. It is the product of the power
 154 function and the exponential function, and its form is identical to a Gamma function.

155 Hydrological data are random variables that exceed zero. If the hydrological processes are
 156 stationary, then the algebraic average and geometric mean of the hydrological characteristics
 157 variable can be approximated as a fixed value. For example, a mean basin annual runoff is
 158 basically stable (the algebraic average is constant), so the probability of a major flood occurring is

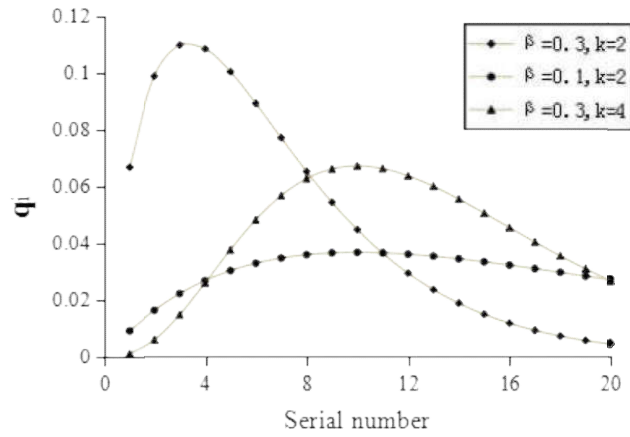


159 small, and most floods are close to the normal value of the accumulated years (the geometric mean
 160 is constant). However, the uncertainty of a different type of flood occurring each time is maximal.
 161 Thus, the complexity of the outcome is maximized. This is consistent with the following: “In a
 162 generalized (objective, system, sampling experiment), if the algebraic average and geometric
 163 mean of the variables (values of statistical indicators) are constant and the complexity is maximal,
 164 then we can conclude that the probability (the percentage) of the flag value (all the values of the
 165 variables) for each individual must obey a Gamma distribution (Pearson type III distribution)
 166 (Zhang, 2003).”

167 At the same time, a lot of practical experiences showed that Gamma function can reflect the
 168 characteristics of flood probability. In the view of this, the unit line q_i is defined as Gamma
 169 function.

$$q_i = \frac{\beta^k i^{k-1} e^{-\beta i}}{(k-1)!}, i = 1, 2, \dots, n \quad (10)$$

171 When the parameters of β and k value were different types, the line type of q_i was
 172 different, as shown in Fig. 1.



173

174

Fig. 1 The change of function under different parameters

175 **1.2 Method**

176 **1.2.1 Genetic algorithms**

177 Genetic Algorithms (GAs) is an effective global search method, which simulates natural
 178 selection and genetic mechanism. This method of searching the optimal solution of the problem
 179 through natural evolutionary process has been applied in many problems such as function



180 optimization and combinatorial optimization. Genetic algorithm can automatically acquire and
181 accumulate the knowledge of search space in the search process, and adaptively control the search
182 process to find the best solution(Davis,1991;Michalewicz,1996). The genetic algorithm regards a
183 family of randomly generated feasible solutions as the parent population, takes fitness function
184 (objective function or one of its transformation forms) as the measurement of the ability of the
185 parent individual to adapt to the environment, generates the offspring individual through selection
186 and hybridization, and then mutates the latter, eliminating the fittest and the fittest, so that the
187 individual adapts to the environment through repeated evolutionary iterations. With the continuous
188 improvement of ability, excellent individuals keep approaching the optimum point(Yuan,2002).
189 After several generations, the algorithm converges to the best individual. The best individual in a
190 group is likely to be the optimal or approximate optimal solution of the problem.

191 As a new random search and optimization method to simulate biological evolution, genetic
192 algorithm has been widely used in the field of optimization(Chen,1996;Li,2009). The parameter
193 optimization of many empirical formulas of hydrological models is essentially based on the global
194 optimization ability of genetic algorithm.

195 1.2.2 Collins iteration method

196 Iterative method is a mathematical process to solve the problem by finding approximate
197 solutions that meet the restrictive conditions from an initial value. Iterative algorithm is also a
198 basic method to solve problems by computer. It makes use of the characteristic of fast computing
199 speed and suitable for repetitive operation, so that the computer can repeat a set of instructions (or
200 steps). When the instructions (or steps) are executed, a new value of the variable will be derived
201 from the original value of the variable. Assuming that we want to derive an approximate solution,
202 we should determine an initial value, an iteration function and a restriction condition according to
203 the actual situation and data firstly, until the absolute value of the initial value and the calculated
204 approximation value is less than a certain value. That is to say, we find the exact desired value.

205 2 Approach used to determined the unit hydrograph

206 The overall calculation process is divided into two parts: (1) the parameters of the unit
207 hydrograph are optimized using the genetic algorithm, so the initial unit hydrograph can be
208 calculated; and (2) the final unit hydrograph is calculated using the Collins iterative method.

209 2.1 Calculation of the initial unit hydrograph using a genetic algorithm

210 (a) Parametrization of Gamma Function

211 For simplicity, Eqn. (10) is transformed as follows

$$212 \begin{cases} q_i = \frac{1}{\Gamma(x)} \cdot x^{x-1} \cdot e^{-x} & i=2, \dots, n-1 \\ q_1 = q_n = 0 \end{cases} \quad (11)$$



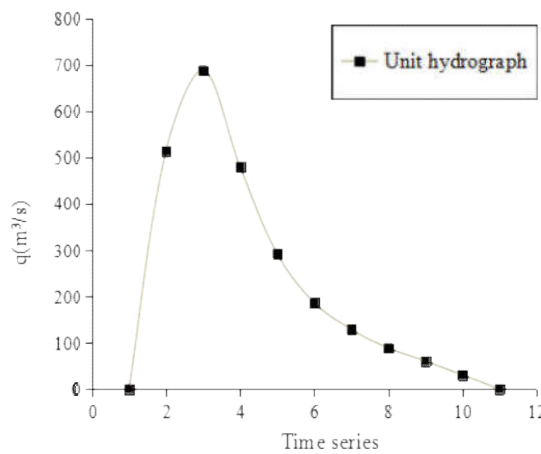
213 Where, x_1 , x_2 , and x_3 are the constants of the unit hydrograph q_i , which is the argument
 214 of the problem for the genetic algorithm. i is the argument of the unit hydrograph, which is a
 215 period number. i is the known number, which is processed by the genetic algorithm. n is the
 216 time period of the unit hydrograph, which is defined according to the actual engineering problem.

217 The variables x_1 , x_2 , and x_3 are in a range of [0, 5]. The chromosome is coded as a floating
 218 point value.

219 If one chromosome is $v^i = [1.7450, 8.7014, 1.5042]$ and $n = 11$ (n is the time period of
 220 the unit hydrograph), the unit hydrograph obtained using the constant given above is as follows.

$$\begin{cases} q_i = \frac{1.7450^{8.7014}}{(8.7014-1) \cdot (i+1.5042)} \cdot e^{(8.7014-1) \cdot (-1.7450/(i+1.5042))} & i = 2, 3, \dots, 10 \\ q_1 = q_{11} = 0 \end{cases} \quad (12)$$

222 Using Eqn. (12), the results obtained for the unit hydrograph are [0, 570, 688, 563, 356, 187, 223
 86, 35, 13, 5, 0], as shown in Fig. 2.



224

225 **Fig. 2** Unit hydrograph q_i

226 (b) Determination of objective function

227 According to the basic principle used to derive the unit hydrograph, the objective function of
 228 the genetic algorithm can be expressed as follows:

$$\max : \varphi(q(x_1, x_2, x_3)) = \frac{1}{(Q'(q_i) - Q)^2} \quad (13)$$

229



230 Where $Q_i(q)$ is the converging flow obtained by the unit hydrograph $q_i(x_1, x_2, x_3)$ in a

231 basin, and Q is the measured flow in the basin outlet section.

232 Physical interpretation of the objective function Eqn. (13)

233 A set of parameters, x_1^* , x_2^* , and x_3^* , are searched based on the following conditions. The

234 inverse square of the difference between $Q_i(q)$ and Q is maximal. To avoid computations if

235 the objective function value is too small, the expansion coefficient M is introduced. The value

236 of M is determined according to the specific situation. Eqn. (13) is converted into the following

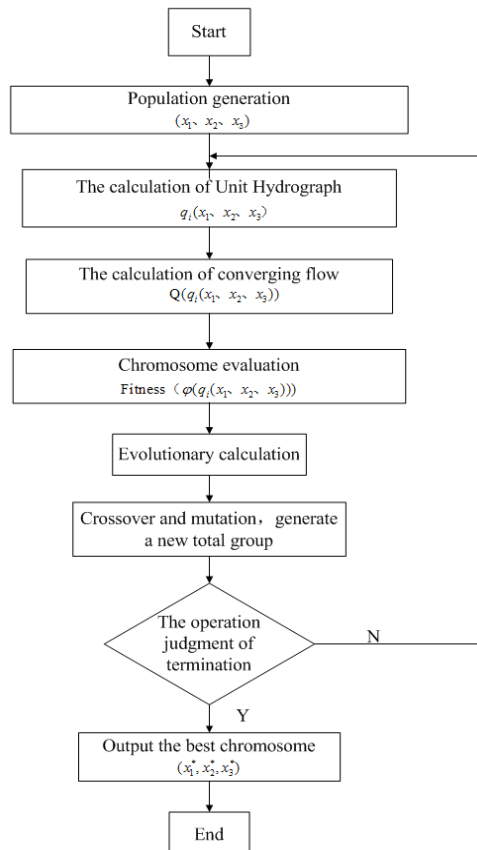
237 form.

$$\max : \varphi(q_i(x_1, x_2, x_3)) = \frac{M}{(Q_i(q) - Q)^2} \quad (14)$$

238

239 (c) Optimization parameters

240 We use the steps shown in Figure 3 of genetic algorithm to optimize the unit lineparameters:



241



242 **Fig. 3** The processes map that was used the genetic algorithm to optimize the parameters

243 The optimum parameters (x_1^*, x_2^*, x_3^*) are obtained by the above steps, and then the unit line
 244 is calculated by the optimum parameters, as follows:

$$245 \begin{cases} q_i = \frac{x_1^* x_2^*}{(x_2^* x - 1)^2} \cdot (i + x_3^*)^{\frac{x_1^* - 1}{x_2^*}} \cdot e^{-x_3^* \cdot (i + x_3^*)}, i = 2, \dots, n - 1 \\ q_1 = q_n = 0 \end{cases} \quad (15)$$

246 **2.2 Calculation of the final unit hydrograph using the iterative method**

247 Collins iteration method is used to calculate the final unit hydrograph. Firstly, each period of
 248 the net rainfall runoff process is calculated by unit hydrograph q_i^* . Meanwhile, the maximum net

249 rainfall h_{max} and its runoff process $Q(q_i^*, h_{max})$ are determined, and overall net rainfall total

250 runoff $\sum Q(q_i^*, h_{max})$ is calculated; Secondly, another unit hydrograph

251 $q_i' = \frac{Q - \sum Q(q_i^*, h_{max})}{h_{max}}$ is deduced, and new q_i' is deduced continuously by q_i^* according to

252 the restriction condition of $\varepsilon = |q_i^* - q_i'| \leq ErrorExcepted$ until the error between the two

253 units meets the requirement, then the final unit hydrograph $q_i = q_i'$.

254 **3 Examples**

255 **Example 1**

256 In Table 1, the data were taken from a previous study (Zhuang and Lin, 1986).

257 **Table 1** The calculations of example 1

258 R: unit hydrograph; Q: measured discharge; Q' , q' : discharge and unit hydrograph of trial and error method;
 259 discharge and unit hydrograph of GACIM

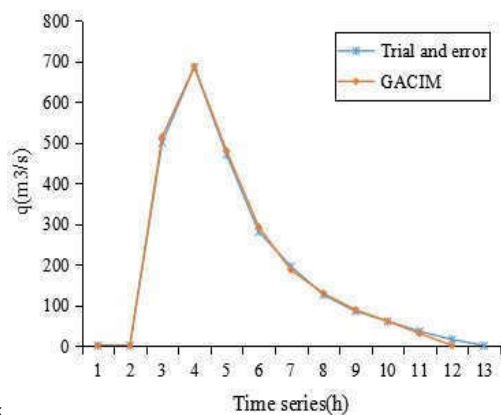
Time series(h)	R(mm)	The measured runoff Q(m ³ /s)	The trial and error method		GACIM	
			Q' (m ³ /s)	q' (m ³ /s)	Q'' (m ³ /s)	q'' (m ³ /s)
①	②	③	④	⑤	⑥	⑦
0		0	0	0	0	0
6	3.8	0	0	0	0	0
12	3.9	50	190	500	195	514
18	0	252	455	685	461	687
24	27.3	662	446	470	450	480
30	2.9	1700	1650	280	1700	292



36	2210	2200	195	2210	186
42	1630	1610	125	1630	129
48	1020	981	85	1020	88
54	650	669	60	650	60
60	440	433	35	440	30
66	290	288	15	290	0
72	190	195	0	190	
78	100	113		100	
84	40	51		9	
90	0	4		0	

260

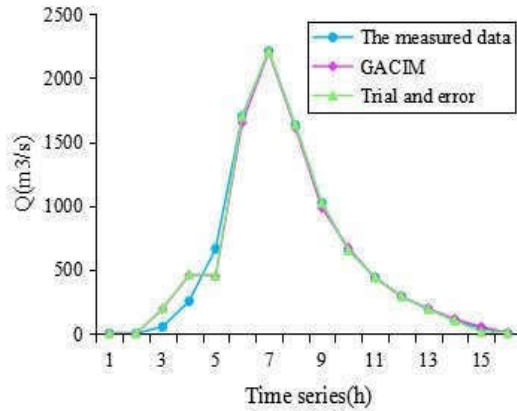
261 The unit hydrographs determined using the two methods are shown in columns (5) and (7) in
 262 Table 1. A comparison of the unit hydrographs is shown in Figure 4. The flow processes
 263 calculated using the unit hydrographs are shown in columns (4) and (6) in Table 1. A comparison
 264 between the calculated flow process and the measured flow is shown in Figure 5.



265

266

Fig. 4 Example1: Unit hydrographs ascertained by two methods



267

268

Fig. 5 Example1: The flow process of outletsection

269

270

271

272

273

274

275

276

The actual hydrological data in Table 1 show that the period number for the runoff was 5 and the period number for the flow process in the outlet section was 15. According to the theory of the unit hydrograph, the period number for the unit hydrograph should be $15 - 5 + 1 = 11$. Using the GACIM (genetic algorithm and the Collins iterative method), the period number for the unit hydrograph was 11. Using the trial and error method, the period number for the unit hydrograph was 12. To further consider the performance of the two methods, we compared the flow process in the outlet section and the results obtained using the two unit hydrographs (measured value and calculated value). A statistical analysis of the results is shown in Table 2.

277

Table 2 The error statistics of example 1

Project Method	GACIM	The trial and error
The error of flood peak(m³/s)	0	10
The maximum error of discharge(m³/s)	212	216
The average absolute error of discharge(m³/s)	37.31	46.19
The total error of flood peak discharge(m³/s.h)	-111	-51
The relative error of flood peak discharge (%)	-1.20	-0.55

278

279 Example 2

280

In Table 3, the data were taken from a previous study (Li and Zheng, 1982).

281

Table 3 The calculations of example 2

Time series(h)	R(mm)	The measured runoff Q(m³/s)	The trial and error method	GACIM
----------------	-------	-----------------------------	----------------------------	-------



	Q' (m ³ /s)	q' (m ³ /s)	Q'' (m ³ /s)	q'' (m ³ /s)		
①	②	③	④	⑤	⑥	⑦
0		0	0	0	0	0
6	15.3	97	96	63	97	49
12	7.4	214	215	110	214	119
18	5.8	304	308	124	304	149
24		371	374	143	371	128
30		294	294	76	294	85
36		190	202	41	190	47
42		123	120	30	123	23
48		80	80	22	80	10
54		49	52	12	49	4
60		30	22	0	19	0
66		15	7		4	
72		0	0		0	

282

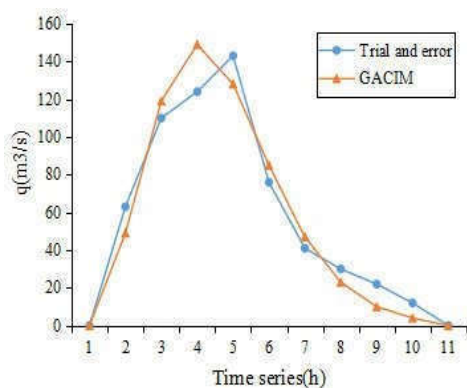
283

284

285

286

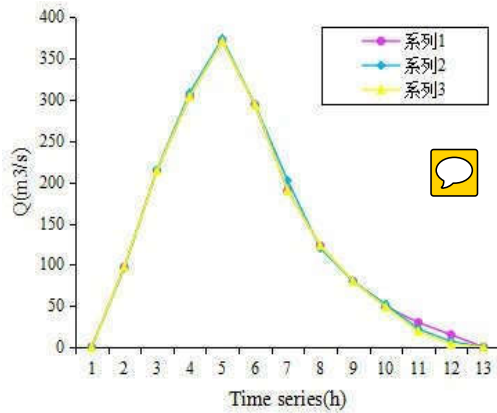
The unit hydrographs determined using the two methods are shown in columns (5) and (7) in Table 3. A comparison of the unit hydrograph is shown in Figure 6. The flow processes calculated using the unit hydrographs are shown in columns (4) and (6) in Table 3. A comparison of the calculated flow process and the measured flow is shown in Figure 7.



287

288

Fig. 6 Example2: Unit hydrographs ascertained by two methods



289

290

Fig. 7 Example2: The flow process of outletsection

291

Figure 4 shows that GACIM was significantly better than the trial and error method in terms of the shape of the curve. We also compared the flow process in the outlet section and the data obtained using the two unit hydrographs. A statistical analysis of the results is shown in Table4.

292

293

294

Table 4 The error statistics of example2

Project Method	GACIM	The trial and error
The error of flood peak(m3/s)	0	-3
The maximum error of discharge(m3/s)	11	-12
The average absolute error of discharge(m3/s)	1.69	-0.23
The total error of flood peak discharge(m3/s.h)	22	-3
The relative error of flood peak discharge (%)	1.25	-0.17

295

296

From Table 2 and Table 4, the calculation accuracy of BGACM is obviously better than that of trial-and-error method in most projects. Although the total error of flood volume is larger than that of trial-and-error method, the relative error of flood volume is only 1.2% and 1.25%, so it does not affect the application of actual projects.

297

298

299

Figure 6 and 7 show that the flow processes of the two unit hydrographs were similar. However, a comparison of the shapes of the unit hydrograph showed that the continuity and smoothness of GACIM were better than the trial and error method. The GACIM method conformed better with the features of a time-invariant system.

300

301

302

303

It can be seen that BGACM method is better at simulating river basin confluence process, which depends on the physical mechanism of the algorithm, while trial-and-error method pays more attention to the balance of total flood volume. This is the respective characteristics and advantages of the two algorithms exactly.

304

305

306

307



308 4 Discussion

309 The present study used a combination of a genetic algorithm and the Collins iterative method
310 (GACIM) for determining a unit hydrograph. The method and implementation steps were
311 described, while examples and analyses were used to demonstrate the scientificity, reliability and
312 practicability of this method. The outcomes of this study are discussed below.

313 (a) Difference between GACIM and other methods

314 In principle, GACIM is based on composition theory and it describes the physical mechanism
315 and process of flood confluence using mathematical equations. Using the basic concept of the unit
316 hydrograph and a genetic algorithm as a mathematical tool, this method can be used to simulate
317 the flood confluence process.

318 Therefore, the simulation of the convergence process is more accurate with GACIM.

319 Other methods for calculating unit hydrographs include the analysis method, least squares
320 method, and the trial and error method. These methods are more focused on the unit hydrograph as
321 an outlet flow process and they fit the measured flow precisely, but they ignore the composition
322 and structure of the unit hydrograph itself. Example 2 shows that GACIM performed better at
323 simulating the basin confluence process, whereas other methods paid more attention to the balance
324 of the total flood volume.

325 (b) Genetic operator design issues

326 A genetic algorithm is a very useful optimization tool. Its biggest advantage is that it has
327 wide adaptability and unlimited problem space, so it can handle many different constraints. This
328 strategy uses a penalty factor. This is because the genetic algorithm method delivers exhaustive
329 engineering accuracy if the population is sufficiently large.

330 There are two types of genetic algorithm, i.e., the standard genetic algorithm (crossover and
331 mutation) and evolutionary computing (selection). A genetic algorithm simulates the
332 recombination of genes to create new offspring in each generation, whereas evolutionary
333 computation is a population process that updates each generation.

334 In this study, a genetic algorithm was used to optimize the parameters of the Gamma function
335 and the unit hydrograph was calculated according to the law of basin confluence. Thus, the
336 parameters were generated by a genetic algorithm. Therefore, the design of the genetic operators is
337 related directly to whether reasonable generation parameters could be obtained.

338 A genetic algorithm has two components: crossover and mutation. Crossover is the main
339 genetic operation that generates new individuals, but it also maintains the relative stability of the
340 population at the same time. However, the variation is a basic calculation and the main effect is to
341 produce a new gene from the population, which provides new information for the population.

342 In general, the initial population of the genetic algorithm is generated in the value space.
343 Crossover and mutation are performed in the value space. In the present study, the value of the
344 Gamma function was in a certain range. Initially, we could not define a reasonable space. If the
345 value space is too large, a bigger population must be used to meet the needs of the individual



346 distribution density. However, this greatly reduces the computational speed. If the value space is
 347 too small, it might not meet the parameter combination required for the engineering precision. To
 348 solve this problem, we observed the following principles during the design of the genetic operator.

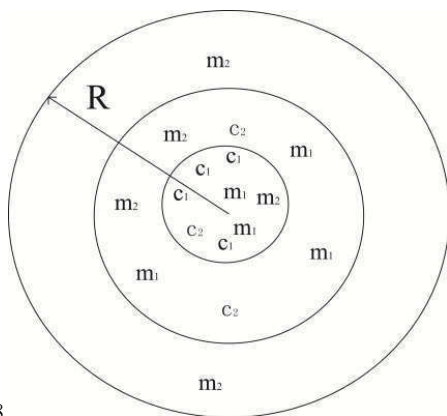
349 (i) The crossover operator was determined where a new individual was generated at random
 350 in the space $[0, a]$ ($a > 0$).

351 (ii) The random expansion value space of the mutation operator was $[0, a]$ and its amplitude
 352 was random. For the floating point coding mutation operator, the following program was
 353 implemented in MATLAB:

- 354 1) numMut=round(size(parent,1)*Ops/2); to calculate the number of variations
- 355 2) numPop=size(parent,1); to calculate the size of the population
- 356 3) numPara=size(parent,2); to calculate the number of parameters
- 357 4) for j=1:numPara
- 358 5) for i=1:numMut
- 359 6) a=round(rand*(numPop-2)+2); select a male parent
- 360 7) parent(a,numPara) = parent(a,numPara)*(1+rand/gen); generate a new generation
- 361 8) end
- 362 9) end.

363 The parameters of the offspring chromosome were calculated during step 7) of this program,
 364 where the variation in the amplitude was related to the number of evolutionary passages. The
 365 variation in the amplitude declined gradually with increasing passage numbers.

366 Figure 8 illustrates the crossover operator and mutation operator with the passage of
 367 evolutionary time.



368
 369 **Fig. 8** The sketch map of the continuation of parameters valuespace
 370

371 In the first generation, the filial generation caused by the crossover operator was still in the
 372 initial parameter space, which corresponds to the inner loop in Figure 8. However, the filial
 373 generation caused by the mutation operator was beyond this range and it expanded to the second



374 ring. In the second generation, the filial generation of the crossover operator was extended to the
375 second ring while the filial generation of the mutation operator was extended to the third ring. The
376 expanding amplitude of the adjacent ring decreased with the passage of evolutionary time. This
377 method was repeated until the predetermined evolutionary algebra was completed.

378 In Examples 1 and 2, the initial values of the parameters were [0 5]. The two sets of
379 optimized parameters were as follows.

380 Example 1: [1.7450, 8.7014, 1.5042]

381 Example 2: [1.6052, 7.9209, 0.30007]

382 These two examples demonstrate the design rationality and the validity of the genetic
383 operator.

384 (c) Research methods for hydrological analysis and calculation

385 The factors that affect hydrological phenomena are very complex. There is still no accurate
386 understanding of the causal relationships among hydrologic phenomena. It is considered that
387 hydrological phenomena involve certainty and randomness, which form the basis of hydrological
388 research. Therefore, causal analysis and probabilistic statistics are the main methods used for
389 hydrological analysis and calculation. In practical applications, causal analysis is confined mostly
390 to qualitative analysis. Quantitative problems demand empirical statistical relationships based on
391 actual observational data.

392 Based on the theory of composition, the distribution function in statistical physics has been
393 extended to hydrology as a non-physics field. Thus, hydrological systems can be viewed as a
394 generalized collection. The regularities of hydrological phenomena have been simulated using
395 distribution functions. Distribution functions and functional relationships have been determined
396 using observation data, which generally means that objective laws are formalized.

397 The present study was a preliminary attempt to investigate the quantitative relationships
398 among hydrological phenomena based on the theory of composition and its distribution function.
399 The author believes that this theory could be a new approach to exploring hydrological rules.

400 **Author Contributions:** Hongyan Li conceived and designed the paper; Cidan Yangzong analyzed the
401 data; Hongyan Li and Cidan Yangzong wrote the paper.

402 **Foundation item:** The author hereby would like to express deep gratitude to the key special project of
403 “Efficient 399 Development and Utilization of Water Resources” (2017YFC0406005), China-ROK
404 cooperation project (51711540299), 400 Natural Science Foundation of Jilin Province (20180101078JC)
405 and other projects which have given support to the 401 research of this paper.

406 **Acknowledgments:** The authors hereby would like to express deep gratitude to the key special project
407 of “Efficient Development and Utilization of Water Resources” (2017YFC0406005), China-ROK
408 cooperation project (51711540299), Natural Science Foundation of Jilin Province (20180101078JC)
409 and other projects which have given support to the research of this paper.

410 **Data Availability Statement:** The data sources and code used are provided in the article



411 References

- 412 1. Bhunya P K, Mishra S K, Berndtsson R. Simplified two-parameter gamma distribution
413 for derivation of synthetic unit hydrograph[J]. Journal of Hydrologic Engineering, 2003,
414 8(4):226-230.
- 415 2. Bhunya P K, Berndtsson R, Ojha C S P, et al. Suitability of Gamma, Chi-square, Weibull,
416 and Beta distributions as synthetic unit hydrographs[J]. Journal of Hydrology, 2007,334(1):
417 28-38.
- 418 3. Bobbitt P. Constitutional fate: Theory of the Composition[M]. New York: Oxford
419 University Press, 1982.
- 420 4. Collins W T. Runoff distribution graphs from precipitation occurring in more than one
421 time unit[J]. Civ. Eng.(NY), 1939, 9(9): 559-561.
- 422 5. Cui H, Singh V P. On the cumulative distribution function for entropy-based hydrologic
423 modeling[J]. Transactions of the ASABE, 2012, 55(2):429-438.
- 424 6. Dooge J. Linear theory of hydrologic systems[M]. Agricultural Research Service, US
425 Department of Agriculture, 1973.
- 426 7. Johnstone D, Cross W P. Elements of applied hydrology[M]. Ronald Press Company,
427 1949.
- 428 8. Li, H.L., Zheng, Z.Q.(1982). Hydrological forecast and calculation[J]. Beijing:China.
429 Water Resources and Electric Power Press,(in Chinese)
- 430 9. Raghunath H M. Hydrology: principles, analysis and design[M]. New Age International,
431 2006.
- 432 10. Ramirez J A. Prediction and modeling of flood hydrology and hydraulics[J]. Inland
433 Flood hazards: Human, riparian and aquatic communities, 2000: 293-333.
- 434 11. Sherman L R K. Streamflow from rainfall by the unit-graph method[J]. Eng. News
435 Record, 1932, 108: 501-505.
- 436 12. Singh S K. Clark's and Espey's unit hydrographs vs the gamma unit hydrograph/Les
437 hydrogrammes unitaires de Clark et de Espey vs l'hydrogramme unitaire de forme loi
438 gamma[J]. Hydrological sciences journal, 2005, 50(6).
- 439 13. Singh S K. Time Base as an Invertible Function of the Parameters of Gamma Unit
440 Hydrograph[J]. Journal of Irrigation and Drainage Engineering, 2009, 135(6): 802-805.
- 441 14. Hanson T.L and Johnson H.P. Unit Hydrograph Methods Compared[J] Transactions of
442 The ASAE, 1964.448-451.
- 443 15. Viessman Jr W. The dynamics of water policy[J]. Water Management, 1989.41-47.
- 444 16. Ye, S.J., Xia, J. Century's retrospect and looking into the future of hydrological
445 science[J] Advance in Water Science, 2002, 13(1): 93-104 (in Chinese).
- 446 17. Yuan, J., Cheng, Y.B., Zhao, Y.F. The parabolic unit hydrograph model of flood
447 forecasting of river course in Gezhou Dam[J]. Yangtze River. 1991, 6-7. (in Chinese)
- 448 18. Zhai, G.J., Li, Y.Q. Math-model and Its Parameter Recognition of the Finite-period
449 Unit Line[J]. Journal of Engineering and Technical Higher Specialty Institutions,
450 2004, 1:001. (in Chinese).
- 451 19. Zhang, W.X. The theory of composition [M]. Hefei: China University of Science and
452 Technology of China press, 2003 (in Chinese).
- 453 20. Davis, L. Handbook of genetic algorithms[M]. New York :Van Nostrand Reinhold, 1991.
- 454 21. Michalewicz & Zbigniew. Genetic algorithms + data structures = evolution program.
455 Berlin:Springer-Verlag(1996).
- 456 22. Yuan ximin, Li hongyan, Liu shukun, etc. Application of Neural Network and Genetic
457 Algorithms in Water Science[M]. Beijing:China Water Resources and Hydropower



- 458 Publishing House, 2002(in Chinese).
459 23. Chen guoliang, Wang xufa, Zhuang zhenquan, etc. Genetic Algorithms and Their
460 Applications[M]. Beijing: People's Posts and Telecommunications Publishing House,
461 1996(in Chinese).
462 24. LI hongyan , Liu xiaowei ,Li shiming . Application of the optimum empirical formula
463 By GAS to the calculation of thesediment concentration process in the Lower Yellow
464 River[J].Journal of Sediment Research, 2009, 30(3):30–36.
465 25. Zhuang,Y.L., Lin, S.Y.(1986). Hydrological forecast[J]. Beijing: China Water
466 Resources And Electric Power Press, (in Chinese).