

# Acquisition of Rainfall in Ungauged Basins: a Study of Rainfall Distribution Heterogeneity Based on a New Method

**Ye Zhao**

Wuhan University

**Xiang zhang** (✉ [zhangxiang@whu.edu.cn](mailto:zhangxiang@whu.edu.cn))

Wuhan University School of Water Resources and Hydropower Engineering

**feng xiong**

Changjiang Water Resources Commission

**Shuying Liu**

Wuhan University

**yao wang**

Wuhan University

**changmei liang**

hubei water resources research institute

---

## Research Article

**Keywords:** Rainfall, Proper orthogonal decomposition, Spatiotemporal variations, Ungauged basin, Interpolation

**Posted Date:** January 11th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1237636/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

1 **Acquisition of rainfall in ungauged basins: a study of**  
2 **rainfall distribution heterogeneity based on a new**  
3 **method**

4 **Ye Zhao<sup>a,b</sup> Xiang Zhang<sup>a,b\*</sup> Feng Xiong<sup>c</sup> Shuying Liu<sup>a,b</sup> Yao Wang<sup>a,b</sup>**

5 **Changmei Liang<sup>d</sup>**

6 *a. State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan*

7 *University, 430072, Wuhan, China*

8 *b. Hubei Key Lab of Water System Science for Sponge City Construction, 430072, Wuhan, China*

9 *c. Bureau of Hydrology, Changjiang Water Resources Commission, 430010, Wuhan, China*

10 *d. Hubei Water Resources Research Institute, Wuhan, 430070, China*

11 **Abstract:** High-density precipitation data is always desired to capture the  
12 heterogeneity of precipitation to accurately describe the components of the hydrological  
13 cycle. However, equipping and maintaining a high-density rain gauge network involves  
14 high costs, and the existing rain gauges are often unable to meet the density  
15 requirements. The objective of this study is to provide a new method to analyze the  
16 spatiotemporal variability of the precipitation field and to solve the problem of  
17 insufficient site density. To this end, the Proper Orthogonal Decomposition (POD)  
18 method is proposed, which can analyze the spatial distribution characteristics of rainfall  
19 fields to solve data shortages. To demonstrate the feasibility and advantages of the

---

\* Corresponding author.

E-mail address: zhangxiang@whu.edu.cn

20 proposed methodology, four districts and counties (Hongshan District, Jianli County,  
21 Sui County, and Xuanen County) in Hubei province in China were selected as case  
22 studies. The principal results are as follows. (1) The proposed method is effective in  
23 analyzing the spatiotemporal variability of the rainfall field to reconstruct rainfall data  
24 in ungauged basins. (2) Compared with the commonly used Thiessen Polygon method,  
25 the Inverse Distance Weighting method, and the Kriging method, POD is more accurate  
26 and convenient, and the root mean squared error is reduced from 3.22, 1.83, 2.19 to  
27 2.09; the correlation coefficients are improved from 0.60, 0.85, 0.79 to 0.89,  
28 respectively. (3) The POD method performs particularly well in simulating the peak  
29 value and the peak time and can offer a meaningful reference for analyzing the spatial  
30 distribution of rainfall.

31 **Keywords:** Rainfall; Proper orthogonal decomposition; Spatiotemporal variations;  
32 Ungauged basin; Interpolation

33

## 34 **1 Introduction**

35 Rainfall process is a highly heterogeneous process covering an extensive range of  
36 scales in time and space (Marani, 2005; Nicótina, 2008). As the main input term of  
37 hydrological models, the variability of rainfall constitutes a significant source of  
38 uncertainty in hydrological simulations (Alexis et al., 2004), therefore, accurate and  
39 reliable estimates of the location and intensity of precipitation are crucial for water  
40 resources management and hydrological studies (Van de Beek et al., 2010).

41 Applying rain gauges is one way to obtain accurate rainfall data (Sohn et al., 2010;

42 Van de Beek et al., 2010), however, the limitation of this method is that it only provides  
43 point measurements and therefore lacks information on the spatial variability (Van de  
44 Beek et al., 2010) unless used in a rain gauge network with sufficient density. Lebel et  
45 al. (1987) found that if rainfall measurements are based on ground measurements only,  
46 their accuracy depends on the spatial variability of the rainfall process and the density  
47 of the rain gauge network. Faurès et al. (1995) concluded that the use of a single rain  
48 gauge can lead to large uncertainties in runoff estimations for a small-scale (4.4 ha)  
49 catchment. This literature suggests that high-density rain gauge networks are needed to  
50 capture rainfall heterogeneity to accurately describe the components of the hydrologic  
51 cycle. However, equipping and maintaining high-density rain gauge networks involves  
52 high costs (Wilco et al., 2018). In most areas, the sparse and nonuniform distribution of  
53 rain gauges does not meet the requirements of hydrologic simulations (Terink et al.,  
54 2018).

55 With the development of remote sensing technology, satellite rainfall monitoring  
56 has become a viable option to complement rain gauge observations. Several high-  
57 resolution satellite rainfall estimates are increasingly available at a quasi-global scale  
58 (Km A et al., 2021). These products have the potential to address the temporal and  
59 spatial sampling limitations of rainfall stations. However, remotely sensed rainfall  
60 products are subject to systematic bias and random error (Kidd, 2001), and their  
61 performance is highly affected by the topography and climatic conditions of the  
62 investigated regions (Romilly and Gebremichael, 2011, Worqlul et al., 2014, Baez-  
63 Villanueva et al., 2018, Mekonnen et al., 2021). In addition, to represent the full spatial

64 distribution of rainfall, massive satellite data is required, however, due to satellite orbit  
65 limitations, many satellite measurements are not continuous, and long records of  
66 satellite data are extremely limited (Segond et al., 2006).

67 The fusion of measured and satellite rainfall data is a general trend in quantitative  
68 rainfall analysis, and the consistent spatial and temporal resolution of multi-source data  
69 is a prerequisite for data fusion. In addition, different engineering projects have  
70 different demands for rainfall data with different spatial and temporal resolutions.  
71 Therefore, the interpolation method of precipitation, i.e., the spatial variation analysis  
72 of precipitation, has been widely studied (Dey and Mujumdar, 2019). There are various  
73 spatial interpolation methods for rainfall data, such as the Thiessen Polygon method,  
74 the Inverse Distance Weighting method (IDW) method, and the Kriging method, which  
75 are commonly used (Jin and Heap, 2011). The Thiessen Polygon method proceeds as  
76 follows (Croley and Hartmann, 1985): the interpolated points are connected two by two  
77 and the vertical bisectors of the connecting lines are made. The vertical bisectors  
78 intersect to form several polygons, thus dividing the large watershed into several sub-  
79 regions, each with a rainfall station, and the rainfall value in each region is represented  
80 by the measured value of the rainfall station within the region. The Thiessen Polygon  
81 method assumes that rainfall varies abruptly across the boundary of the polygon and  
82 that rainfall is uniformly distributed within the sub-region, which is unphysical. The  
83 IDW method assumes that the interpolated points are more influenced by the closer  
84 stations than the far ones, and thus uses the reciprocal of the squared distance as the  
85 weighting factor (Chen et al., 2015). The disadvantage of this method is that the

86 interpolation error is large when the data is significantly different from the neighboring  
87 points. The Kriging method (Armstrong and Matheron, 1986) considers not only  
88 geographical features but also spatial correlation and variability of rainfall data and uses  
89 a covariance function to make the variables unbiased and optimal. However, when there  
90 are few rainfall stations, it is quite difficult to obtain the covariance function necessary  
91 for the kriging method, which will also affect the interpolation effect to some extent  
92 (Gilbert and Simpson, 1985). In addition, these methods are interpolated for a single  
93 time point, and the acquisition of rainfall processes for long time series requires iterative  
94 operations, which is very time-consuming.

95 To accurately and efficiently simulate the rainfall processes with high temporal  
96 resolution in large basins, it is crucial to fully consider the heterogeneity of rainfall  
97 fields. Many factors may contribute to the variability of precipitation, including tropical  
98 disturbances, El-Niño Southern Oscillation, and Madden-Julian Oscillation, while the  
99 influence of spatial factors is the most significant (Pariyar et al., 2020). Variations in  
100 rainfall processes at different locations are often directly reflected in the temporal  
101 distribution of rainfall; therefore, they can be considered as “spatiotemporal variations”.  
102 For example, owing to the different locations of sites B and A, rainfall at site B might  
103 occur later than at site A, and the rainfall pattern at site B might be gentler (steeper)  
104 owing to weakening (enhancement) of the air mass (Palynchuk and Guo, 2011; Lin and  
105 Jhong, 2015). Several scholars have tried to represent rainfall in continuous space and  
106 time mathematically. For instance, Northrop (1998) developed a spatial-temporal model  
107 based on the Bartlett-Lewis process for the temporal evolution of storm cells and the

108 Neyman-Scott process for spatial arrivals. Wilks (1998) proposed an extension from a  
109 single station to multiple stations by driving single-station models at various locations  
110 in New York with temporally independent but spatially correlated random numbers.  
111 Cowpertwait et al. (2002) proposed a spatial-temporal model based on the Neyman-  
112 Scott mechanism and tested the model on nine stations in Italy. Yang et al. (2005) used  
113 generalized linear models for the generation of multisite rainfall series in Southern  
114 England. Abas et al. (2014) presented a stochastic rainfall model for the generation of  
115 hourly rainfall data in an urban area in Malaysia. However, due to the influence of  
116 topography and climate, there is a large spatial gradient in rainfall intensity, such models  
117 currently assume spatial stationarity is not credible. (Alam and Elshorbagy, 2015; Kwon  
118 et al., 2016). Improvements to include nonstationary properties of precipitation, to  
119 represent (1) the temporal characteristics of rainfall, and (2) the spatial features of the  
120 data such as the location of high values of rainfall, topographic effects, or spatial  
121 continuity, are still under development (Wheater et al., 2005).

122 Proper Orthogonal Decomposition (POD) is an effective dimensionality reduction  
123 method that uses basis functions extracted from a large number of known data to  
124 describe the original data in an optimal approximation (Aubry, 1991). Because of its  
125 ease of interpretation and broad applicability to data from both simulations and  
126 experiments (Kutz et al., 2016), the POD technique is now used widely in fields such  
127 as fluid mechanics, aerodynamics, economics, statistics, psychology to identify and  
128 analyze spatial distribution characteristics of some variables. For example, Meric et al.  
129 (2012) applied the POD method to extract information on the dynamic spatiotemporal

130 evolution of global stock markets during the 2008 economic crisis, Cheveigné and  
131 Simon (2007) applied the POD method to neuroscience to extract the evolutionary  
132 structure of neurons in the human brain, and Varraso et al. (2012) used POD to extract  
133 and analyze the main spatial transmission of epidemic characteristics. However, this  
134 method has rarely been used in rainfall analysis. The structure of a vibration field, flow  
135 field, or aerodynamic field is similar to that of a rainfall field, and these fields change  
136 constantly with time and space. The velocity in a flow field and the pressure in an  
137 aerodynamic field are similar to the rainfall intensity in a precipitation field. Therefore,  
138 the POD method also has potential for applicability in the field of rainfall simulation,  
139 where it can be used to analyze the spatial distribution characteristics of rainfall fields  
140 and as an interpolation-like method to obtain rainfall data in ungauged basins.

141 This study was conducted with the following two objectives: (1) to develop a  
142 POD-based method to extract and analyze the spatial distribution characteristics of  
143 rainfall fields. This method can be used as an interpolation-like method to reconstruct  
144 rainfall processes in ungauged basins to obtain accurate simulations of rainfall  
145 processes with high temporal resolution. and (2) to test the applicability of the proposed  
146 method through comparative analysis with the interpolated results obtained using the  
147 Thiessen Polygon method, the IDW method, and the Kriging method.

## 148 **2 Methodology**

149 The POD method, also known as the Karhunen-Loève procedure (Karhunen, 1946;  
150 Loève, 1955), is an effective dimensionality reduction method that can extract modes  
151 based on optimizing the mean square of the field variables being examined (Taira, 2017).

152 It provides an objective algorithm to decompose a set of data into a minimal number of  
153 modes, each reflecting the influence of one or some physical factor on the field being  
154 decomposed, and these modes can be used to make the optimal approximate description  
155 of the original data. In this study, POD was implemented to decompose the precipitation  
156 field ( $\mathbf{X} \in \mathbf{R}^{m \times n}$ ) into several orthogonal modes, each of which may reflect the influence  
157 of one or more physical factors on the decomposed field, and the modes reflecting the  
158 influence of spatiotemporal factors can be found through comparative analysis. By  
159 reconstructing the rainfall field with the remaining modes after removing the modes  
160 reflecting spatiotemporal characteristics, the rainfall field structure independent of  
161 spatiotemporal factors can be obtained, and the rainfall processes in the ungauged  
162 basins can be reconstructed indirectly. The detailed steps and physical meanings of the  
163 variables involved in using the POD method to obtain the spatiotemporal structure of  
164 the rainfall field and to obtain data for ungauged basins are described in the following  
165 (Fig. 1).

## 166 **2.1 Sampling and pre-processing of data**

167 To study the influence of spatiotemporal factors on the rainfall field structure, two  
168 data sampling sessions are required. The first sample matrix  $\mathbf{X}_A$  consists of a certain  
169 rainfall process of sufficient duration at multiple stations in the study area, and the other  
170 sample matrix  $\mathbf{X}_B$  is manually spatiotemporally normalized based on the previous  
171 sample. By manual alignment (Roja et al., 2021), the rainfall processes at different  
172 stations start and end simultaneously. With the above data collation, the original  
173 determinant  $\mathbf{X}_A$  has spatiotemporal variations, and the processed determinant  $\mathbf{X}_B$  can

174 be considered to be uniformly distributed in time and space.  $\mathbf{X}_A$  and  $\mathbf{X}_B$  can be denoted  
 175 as follows:

$$\begin{aligned}
 176 \quad \mathbf{X}_A &= [\mathbf{u}_A(x, t_1), \mathbf{u}_A(x, t_2), \dots, \mathbf{u}_A(x, t_n)] \in \mathbf{R}^{m \times n} \\
 177 \quad \mathbf{X}_B &= [\mathbf{u}_B(x, t_1), \mathbf{u}_B(x, t_2), \dots, \mathbf{u}_B(x, t_n)] \in \mathbf{R}^{m \times n}, \quad (1)
 \end{aligned}$$

178 where  $n$  is the length of the rainfall duration, and  $m$  reflects the number of rain gauges  
 179 in the study area. In fluid dynamics, the measured state  $\mathbf{u}(x, t_n)$  is related to the  
 180 physics, i.e., the velocity, vorticity, and stream functions at each discretized spatial grid  
 181 point. For precipitation data collected from a rainfall field, the rainfall intensity at  
 182 different rain gauges at time  $t_n$  can be denoted as  $\mathbf{u}(x, t_n)$ , which contains the mean  
 183 of  $\mathbf{u}(\bar{u}(x))$  and the fluctuating rainfall intensity  $\mathbf{v}(x, t_n)$ . It is customary to subtract  
 184  $\bar{u}(x)$  before computing the POD modes. The fluctuating rainfall intensity after  
 185 removing  $\bar{u}(x)$  at time  $n$  can be denoted as follows:

$$186 \quad \mathbf{v}(x, t_n) = \mathbf{u}(x, t_n) - \bar{u}(x), \quad (2)$$

187 the fluctuating rainfall field after removing  $\bar{u}$  at time  $t_n$  can be denoted as follows:

$$\begin{aligned}
 188 \quad \mathbf{X}_A &= [\mathbf{v}_A(x, t_1), \mathbf{v}_A(x, t_2), \dots, \mathbf{v}_A(x, t_n)] \in \mathbf{R}^{m \times n} \\
 189 \quad \mathbf{X}_B &= [\mathbf{v}_B(x, t_1), \mathbf{v}_B(x, t_2), \dots, \mathbf{v}_B(x, t_n)] \in \mathbf{R}^{m \times n}, \quad (3)
 \end{aligned}$$

190 where  $\mathbf{v}(x, t_i)$  stands for the  $i$ -th fluctuating rainfall intensity at different stations in  
 191 the study area.

## 192 2.2 POD calculation on the sample matrices

193 Various methods can be used to calculate POD, and the method selected for use in  
 194 this study was the popular SVD (singular value decomposition)-based approach (Eckart  
 195 and Young, 1936). The purpose of applying POD in this study is to find the mode(s)

196 that reflect spatiotemporal variations of the rainfall field, and to achieve this purpose,  
 197  $\mathbf{X}_A$  and  $\mathbf{X}_B$  are input to POD, respectively.

198 In matrix form, the data matrix  $\mathbf{X}_A$  and  $\mathbf{X}_B$  can be decomposed with SVD as  
 199 follows (Fig. 2):

$$\begin{aligned}
 200 \quad \mathbf{X}_A &= \mathbf{\Phi}_A \mathbf{\Sigma}_A \mathbf{\Psi}_A^T \\
 201 \quad \mathbf{X}_B &= \mathbf{\Phi}_B \mathbf{\Sigma}_B \mathbf{\Psi}_B^T, \tag{4}
 \end{aligned}$$

202 where the  $m \times n$  matrices  $\mathbf{\Phi}_A = [\phi_{A1}, \phi_{A2}, \dots, \phi_{An}]$  and  $\mathbf{\Phi}_B = [\phi_{B1}, \phi_{B2}, \dots, \phi_{Bn}]$  are  
 203 composed of the modes  $\phi_A$  and  $\phi_B$ .  $\phi_A$  reflects the spatial distribution information  
 204 of rainfall intensity containing rainfall spatial variability, and  $\phi_B$  reflects the structure  
 205 of the rainfall field after removing rainfall spatial variability. The  $n \times n$  matrix  
 206  $\mathbf{\Psi}_A = [\mathbf{a}_{A1}, \mathbf{a}_{A2}, \dots, \mathbf{a}_{An}]$  and  $\mathbf{\Psi}_B = [\mathbf{a}_{B1}, \mathbf{a}_{B2}, \dots, \mathbf{a}_{Bn}]$  contain coefficients representing  
 207 the temporal evolution of the modes  $\mathbf{\Phi}_A$  and  $\mathbf{\Phi}_B$ , respectively. The superscript T  
 208 represents a conjugate transpose matrix operation (Higham et al., 2017). Matrix  $\mathbf{\Sigma}$   
 209 holds singular values ( $\sigma_1, \sigma_2, \dots, \sigma_n$ ) along its diagonal.  $\lambda_j$  ( $\lambda_j = \sigma_j^2$ ) is the  
 210 contribution of the  $j$ -th mode to the total variance, indicating the influence of the  $j$ -th  
 211 mode on the overall rainfall field.  $\lambda$  is listed in descending order, i.e.:  
 212  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ , the larger the value of  $\lambda$ , the greater the mode contribution.  
 213 Furthermore, the percentage rainfall intensity energy contribution ( $E$ ) of each POD  
 214 mode can be obtained as follows:

$$215 \quad E_i(\%) = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \times 100. \tag{5}$$

### 216 2.3 Reconstruction of rainfall processes in ungauged basins

217 To find the modes that characterize the spatiotemporal factors of the rainfall field,  
 218 the original data and the manually normalized spatiotemporal data were subjected to  
 219 POD in the previous section. It is assumed that the energy of the first  $r$ -order modes is  
 220 close to the overall energy of the precipitation field, i.e., the modes reflecting the  
 221 spatiotemporal influence factors of the rainfall field are included in the first  $r$ -order  
 222 modes, and therefore the first  $r$ -order modes are singled out for further comparison and  
 223 analysis.

$$224 \quad \sum_{j=1}^r \lambda_j / \sum_{j=1}^n \lambda_j \approx 1. \quad (6)$$

225 By comparing the mode values, numerical distributions, and temporal evolution  
 226 coefficients of the predominant modes obtained from two POD calculations with  
 227 different inputs, those modes with significant differences are excluded because they are  
 228 considered to be related to spatiotemporal properties. Then, a homogeneous  
 229 precipitation field is reconstructed with the remaining predominant modes. The  
 230 fluctuating rainfall intensity  $\nu(x, t)$  can be expressed as a linear superposition of all  
 231 orthogonal modes ( $\phi(x)$ ) multiplied by their corresponding time factors ( $\mathbf{a}(t)$ ):

$$232 \quad \nu(x, t) = \sum_{j=1}^r \mathbf{a}_j(t) \phi_j(x), \quad (7)$$

233 where  $j$  is the subscript of the remaining modes after removing those modes related to  
 234 spatiotemporal properties. Then, the precipitation field can be reconstructed as follows:

$$235 \quad \mathbf{X}_{POD}(x, t) = \bar{u}(x) + \sum_{j=1}^r \mathbf{a}_j(t) \phi_j(x) \quad (8)$$

236 Because the temporal and spatial variations of the precipitation field are eliminated,  
 237 the errors between precipitation at close locations are greatly reduced, and the

238 precipitation processes in the ungauged basin can be represented by precipitation  
239 processes at an adjacent station (Fig. 3).

## 240 **2.4 Verification methods**

241 To assess the performance of the methods used in this study, we used three  
242 performance statistics widely used to compare the model performance such as Root  
243 Mean Square Error (RMSE), Relative Error (RE), and Correlation correlation (Corr).  
244 Based on the results, performance of these methods is presented in detail.

### 245 **(1) Root mean squared error (RMSE)**

246 The RMSE (Eq. (9)) estimates the standard deviation between the observed value  
247 and the simulated value. A small RMSE value indicates better performance whereas a  
248 higher value indicates a poor performance.

$$249 \quad RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Q_o^t - Q_m^t)^2} \quad (9)$$

250 where  $Q_o$  and  $Q_m$  refer to the observed and simulated values, respectively, and T is  
251 the total number of observations.

### 252 **(2) Relative error (RE)**

253 The RE represents the magnitude of the deviation of the simulated value from the  
254 true value and is calculated as follows:

$$255 \quad RE = \frac{1}{T} \left[ \sum_{t=1}^T \left| 1 - \frac{Q_m^t}{Q_o^t} \right| \right] \times 100\% \quad (10)$$

### 256 **(3) Correlation coefficient (Corr)**

257 The Corr is a statistical indicator to reflect the closeness of the correlation between  
258 variables and its value ranges between -1 and 1 representing a perfect negative and

259 positive correlation, respectively. The correlation coefficient is calculated by the  
260 product-difference method and is based on the deviation of the two variables from their  
261 respective means. The formula for calculating correlation can be represented as:

$$262 \quad \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}, \quad (11)$$

263 where  $\text{Cov}(X, Y)$  is the covariance of  $X$  and  $Y$ ,  $\text{Var}[X]$  is the variance of  $X$ , and  
264  $\text{Var}[Y]$  is the variance of  $Y$

### 265 **3 Study area and data**

266 Hubei Province is located in central China, between latitude 29°01'53"-33°6'47"  
267 N and longitude 108°21'42"-116°07'50" E with a total area of 185900km<sup>2</sup>. Mountains,  
268 hills, and plain lakes account for 56%, 24%, and 20% of the total area of Hubei province,  
269 respectively. Except for the alpine climate in the high mountain areas, most parts of the  
270 province belong to the subtropical monsoon humid climate, with sufficient light energy,  
271 abundant rainfall, and long frost-free periods. The average annual precipitation in Hubei  
272 province is between 800 and 1600 mm. The regional distribution of precipitation shows  
273 a decreasing trend from south to north (Fig. 4), with a maximum of 1400-1600 mm in  
274 the southwest and a minimum of 800-1000 mm in the northwest. The seasonality of  
275 precipitation distribution is obvious, with the most in summer (300-700 mm) and the  
276 least in winter (30-190 mm), respectively. The rainy season is from mid-June to mid-  
277 July, with the largest rainfall and intensity. To study the applicability of the method  
278 recommended in this paper, we selected several extreme rainstorms in Hubei Province  
279 for analysis. Hourly rainfall data at 56 stations in Hubei province are downloaded from  
280 European Centre for Medium-Range Weather Forecasts.

## 281 **4 Results**

### 282 **4.1 Reconstruction of rainfall processes at ungauged basins**

283 Because of the large spatially varying gradients of precipitation in Hongshan  
284 District, Jianli County, Sui County, and Xuanen County in Hubei province, these  
285 locations are assumed to be the ungauged basins for rainfall reconstruction (Table 1).  
286 The multiple extreme rainfall data of Hubei Province in 2020 were used as the input of  
287 POD to obtain the modes with spatiotemporal variability, and the data processed by  
288 manual spatiotemporal normalization were input to obtain the modes with  
289 spatiotemporal variability removed. Fig. 5 shows the energy contribution and the  
290 cumulative energy contribution of the modes decomposed by POD with raw data and  
291 processed data as input, respectively. The energy contributions of the modes obtained  
292 from the POD of raw data show a larger descending trend, where the cumulative energy  
293 contribution ratio of the first two modes is approximately 10% higher than that of the  
294 modes obtained by processed data, which indicates that these modes are more  
295 representative. Starting from the fifth mode, the energy contribution of processed data  
296 becomes larger than that of the raw data. And after the ninth mode, the cumulative  
297 energy contribution rates of the two compared cases are almost the same. In both cases,  
298 the cumulative energy contribution rates of the first six modes exceed 80% and thus  
299 these modes are selected as the predominant modes for subsequent analysis and  
300 reconstruction of the precipitation field.

301 Fig. 6 shows the modes and the corresponding time factors before spatiotemporal  
302 normalization, i.e., raw data is used as input to the POD. As the energy contribution of

303 the mode decreases, a higher proportion of null values is observed for modes with  
304 values close to zero, which is consistent with the physical meaning of the energy  
305 contribution. In addition, as the energy contribution rate decreases, the time factor  
306 corresponding to the mode fluctuates more frequently, which indicates that the mode  
307 becomes more unstable as the energy contribution rate decreases.

308 The mode represents the distribution structure of the rainfall parameter (intensity)  
309 in a certain physical sense, and the corresponding time factor reflects the trend of the  
310 mode over time. Because the energy contribution of the first mode exceeds 27.44%, it  
311 can be assumed that the reconstruction based on the first mode reflects the main  
312 structure of the precipitation field. The values of the first mode are mostly negative and  
313 the corresponding time factor is in the inverted triangular shape; therefore, a positive  
314 triangular rainfall intensity process can be obtained by multiplying the first mode and  
315 its corresponding time factor. The major mode values of the second mode are positive,  
316 with the corresponding time factor appearing in the shape of a triangle. A similar rainfall  
317 process that increases firstly and then decreases can be obtained by multiplying the two.  
318 This can be interpreted as the structure of this precipitation field is a single-peak rainfall  
319 process. The values of the third to sixth modes are mostly around zero, and their  
320 corresponding time factors fluctuate frequently. This indicates that they play little role  
321 in the final results of the reconstructed rainfall pattern, but there may still be modes  
322 reflecting spatiotemporal factors in them.

323 To find the modes characterizing the variations of time and space, the processed  
324 data are used as input to POD to get the modes after spatiotemporal normalization.

325 Similar to the modes obtained from the original data as input, the proportion of null  
326 values and the fluctuation frequency of the time factors increase with the decrease of  
327 the energy contribution of the modes. However, the predominant modes and their  
328 corresponding time factors change slightly due to different inputs (Fig. 7). Compared  
329 with the first mode before spatiotemporal normalization, the values of the first mode  
330 after spatiotemporal normalization are also mostly negative, decreasing from northwest  
331 to northeast and slightly higher in the southeast. The main pattern of precipitation after  
332 spatiotemporal normalization is also a single-peaked rainfall with a backward peak,  
333 which indicates that the main pattern of precipitation does not change significantly  
334 before and after the normalization of time and space. The distribution of the values of  
335 the fourth mode is almost the same before and after normalization. The values of the  
336 fifth mode show an increasing trend from west to east and reach the maximum at  
337 northeast, and the trends of time factors of the fifth modes before and after the  
338 normalization are the same. Therefore, it can be assumed that the first, fourth, and fifth  
339 modes do not reflect the influence of spatiotemporal factors on the structure of the  
340 precipitation field. However, the second, third, and sixth modes change significantly  
341 due to the normalization of time and space. Therefore, it can be assumed that the  
342 superimposed effect of these three modes reflects the spatiotemporal variations of the  
343 precipitation field. Therefore, the reconstructed precipitation field after removing these  
344 three modes can be regarded as the result of removing the spatiotemporal variations.

345 Fig. 8 and Fig. 9 show the reconstructions of precipitation fields before and after  
346 spatiotemporal normalization (removing the second, third, and sixth modes),

347 respectively. The comparison shows that the rainfall processes at different stations are  
348 more consistent and the peaks are more pronounced after the removal of spatiotemporal  
349 variations.

#### 350 **4.2 Comparative analysis of different methods**

351 Based on the results obtained in the previous section, the rainfall processes at the  
352 ungauged sites are reconstructed according to the method introduced in section 2.3. To  
353 illustrate the implementability and superiority of the proposed method, three most  
354 commonly used interpolation methods in current hydrological engineering: the  
355 Thiessen Polygon method, the IDW method, and the Kriging method, were selected for  
356 comparative study. The rainfall data obtained by different methods as well as the  
357 measured rainfall data in the ungauged stations are shown in Fig. 10. It can be seen that  
358 a 24-h rainfall varies in magnitude and time distribution in different areas of Hubei  
359 Province, and the interpolation results of different methods applied in different areas  
360 also have great differences. The Thiessen Polygon method is rough from the mechanism,  
361 and the interpolation of rainfall peak values and peak times is not accurate enough,  
362 especially in Jianli County where the rainfall gauge distribution is relatively sparse. The  
363 interpolation results of peak present time in both Hongshan District and Sui County are  
364 late, which is unfavorable to the prevention and control of rain floods. The interpolation  
365 results of the Kriging and IDW methods in the Hongshan District are small, and the  
366 peak times in both Jianli and Xuanen counties have relatively large errors. The POD  
367 method has a small simulation of the early and late rainfall processes but compared with  
368 the other three methods, it has a more accurate simulation of the peak value and peak

369 time. The median and arithmetic mean of the relative errors of the POD simulations for  
370 the peak portion of the rainfall are the smallest, which is very meaningful for  
371 hydrological design (Fig. 11).

372 Fig. 12 and Table 2 show the Corrs between the interpolated results and the  
373 measured values, as well as the RMSEs. It can be found that the results obtained by the  
374 Thiessen Polygon method have a large error, and the accuracy of the interpolation  
375 results for rain intensity above 5 mm/h is low, with a Corr of only 0.6 and an RMSE as  
376 high as 3.22mm/h. The interpolation results of Kriging method are mostly a little  
377 smaller than the measured values, which is unfavorable from the perspective of  
378 practical engineering safety. The Corr of the Kriging method is 0.79 and the RMSE is  
379 2.19mm/h. The interpolation results of the IDW method are more accurate, with an  
380 RMSE of only 1.83 mm/h and a Corr of 0.85, but the interpolation of the IDW method  
381 for rainfall exceeding 15 m/h tends to be small. Perhaps because the extracted modes  
382 contain not only the spatial and temporal characteristics of the rainfall field but also a  
383 part of the initial and final rainfall, the POD method in this case is inaccurate in  
384 simulating rainfall amounts less than 5 mm/h at the beginning and end of the rainfall  
385 period, resulting in an RMSE of 2.09 mm/h. However, the POD method is more  
386 accurate for rainfall over 10 mm/h and the Corr exceeds 0.89.

## 387 **5 Discussion and Conclusions**

388 Spatial interpolation of rainfall has been regarded as the most common method to  
389 obtain rainfall data in ungauged basins, and the non-consistency of rainfall spatial  
390 distribution is the source of error in the interpolation method. In this study, POD is

391 introduced to analyze the spatiotemporal variations of rainfall fields and is used as an  
392 interpolation-like method to obtain rainfall data in ungauged basins.

393         POD is a dimensionality reduction method that can be used to analyze the spatial  
394 distribution characteristics of different variable fields. Similar to its application to  
395 extract the flow structure in fluid mechanics (Redha et al., 2018), information on  
396 dynamic market transactions in finance (Meric et al., 2012), neuronal evolutionary  
397 structures in neuroscience (Cheveigné and Simon, 2007), and virus transmission  
398 characteristics in epidemiology (Varraso et al., 2012), our study demonstrated that POD  
399 can efficiently identify and extract the structures of the rainfall field. By comparing the  
400 results obtained from two different sets of data as input, it is possible to find the modes  
401 that reflect the characteristics of the spatial distribution of the rainfall field, and the  
402 rainfall field reconstructed by removing these modes can be regarded as spatially  
403 uniformly distributed. Based on this, the rainfall processes in the ungauged stations can  
404 be reconstructed using the measured rainfall data from adjacent stations. This is  
405 particularly advantageous for urban areas with large spatially varying rainfall gradients.

406         In the application to the case study area in Hubei Province, POD was proven an  
407 effective and reliable approach for obtaining rainfall data in ungauged basins. The first  
408 advantage of the proposed method is that compared with the traditional method, the  
409 proposed method has a significant improvement in the simulation of the peak value and  
410 peak time. Second, the use of the POD method to remove spatiotemporal correlation  
411 modes to reconstruct rainfall in ungauged basins is once and for all, and it is more  
412 efficient to obtain long series of rainfall data with the POD method than other methods

413 that require repeated interpolations. However, despite the above advantages, this  
414 method has the potential for further improvement. For example, how to separate the  
415 parts of the same mode that may contain different characteristics of the rainfall field  
416 and how to quantitatively identify the modes that characterize spatial and temporal  
417 properties remain to be discussed. It is hoped that this study will serve as a stepping  
418 stone for the readers to become familiar with POD, analyze complex issues associated  
419 with spatial heterogeneity of rainfall fields, and further advance the developments of  
420 this method in analyzing the effects of other factors (e.g., distance between stations and  
421 the ocean, slope, slope direction, wind speed) on the spatial distribution of rainfall data.

422

#### 423 **Data Availability Statement**

424 All data included in this study are available upon request by contact with the  
425 corresponding author.

#### 426 **Acknowledgments**

427 This research is supposed by the National Key Research and Development  
428 Program of China (No.2019YFC0408901) and the National Natural Science  
429 Foundation of China (No.41890823). We express our gratitude to the reviewers and  
430 editors for their comments on the manuscript.

431 **References**

- 432 1. Abas, N. et al. A comparative study of mixed exponential and Weibull distributions  
433 in a stochastic model replicating a tropical rainfall process. *Theoretical and Applied*  
434 *Climatology*. 118(3), 597-607 (2014).
- 435 2. Alam, M. S., and Elshorbagy, A. Quantification of the climate change-induced  
436 variations in intensity–duration–frequency curves in the Canadian prairies. *Journal of*  
437 *Hydrology*. 527, 990-1005 (2015).
- 438 3. Alexis, Berne, and Guy, Delrieu. Temporal and spatial resolution of rainfall  
439 measurements required for urban hydrology. *Journal of Hydrology*. 299(3-4), 166-179  
440 (2004).
- 441 4. Armstrong, M., and Matheron, G. Disjunctive kriging revisited. I, II. *Math Geol*.  
442 (8), 711-728,729-742 (1986).
- 443 5. Aubry, N. On the hidden beauty of the proper orthogonal decomposition.  
444 *Theoretical and Computational Fluid Dynamics*. 2(5), 339-352 (1991).
- 445 6. Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-  
446 Osorio, J. D., and Thinh, N. X. Temporal and spatial evaluation of satellite rainfall  
447 estimates over different regions in Latin-America. *Atmospheric Research*. 213(NOV.),  
448 34-50 (2018).
- 449 7. Chen, C., Zhao, N., Yue, T., and Guo, J. A generalization of inverse distance  
450 weighting method via kernel regression and its application to surface modeling. *Arabian*  
451 *Journal of Geosciences*. 8, 6623–6633 (2015).
- 452 8. Cheveigné AD, and Simon, J. Z. Denoising based on time-shift PCA. *Journal of*

- 453 *Neuroscience Methods*. 165(2), 297-305 (2007).
- 454 9. Cowpertwait, P., Kilsby, C. G., and O'Connell, P. E. A space-time neyman-scott  
455 model of rainfall: empirical analysis of extremes. *Water Resources Research*. 38(8), 6-  
456 1-6-14 (2002).
- 457 10. Croley, T. E., and Hartmann, H. C. Resolving thiesen polygons. *Journal of*  
458 *Hydrology*. 76(3-4), 363-379 (1985).
- 459 11. Dey, P., and Mujumdar, P. P. On the uniformity of rainfall distribution over India.  
460 *Journal of Hydrology*. 578, 124017 (2019).
- 461 12. Eckart, C., and Young, G. The approximation of one matrix by another of lower  
462 rank. *Psychometrika*. 1(3), 211-218 (1936).
- 463 13. Higham, J. E. et al. Using modal decompositions to explain the sudden expansion  
464 of the mixing layer in the wake of a groyne in a shallow flow. *Advances in Water*  
465 *Resources*. 107: 451-459 (2017).
- 466 14. Gilbert, R. O., and Simpson, J. C. Kriging for estimating spatial pattern of  
467 contaminants: potential and problems. *Environmental Monitoring and Assessment*. 5(2),  
468 113 (1985).
- 469 15. Jin, L., and Heap, A. D. A review of comparative studies of spatial interpolation  
470 methods in environmental sciences: performance and impact factors. *Ecological*  
471 *Informatics*. 6(3-4), 228-241 (2011).
- 472 16. JM Faurès, Goodrich, D. C., Woolhiser, D. A., and Sorooshian, S. Impact of small-  
473 scale spatial rainfall variability on runoff modeling. *Journal of Hydrology*. 173(1), 309-  
474 326 (1995).

- 475 17. Karhunen, K. Zur spektral theorie stochasticher prozesse. *Annales Academiae*  
476 *Scientiarum Fennicae*. 34, 13 (1946).
- 477 18. Kidd, C. Satellite rainfall climatology: a review. *International Journal of*  
478 *Climatology*. 21(9), 1041-1066 (2010).
- 479 19. Km, A., Amm, B., and Taw, A. Effect of temporal sampling mismatches between  
480 satellite rainfall estimates and rain gauge observations on modelling extreme rainfall in  
481 the upper awash basin, Ethiopia. *Journal of Hydrology*. 598(7), 126467 (2021).
- 482 20. Kutz, J. N., Brunton, S. L., Brunton, B. W., and Proctor, J. L. *Dynamic Mode*  
483 *Decomposition: Data-Driven Modeling of Complex Systems*. Society for Industrial and  
484 Applied Mathematics Philadelphia (2016).
- 485 21. Kwon, Hyun-Han, Lima, Carlos, H. R., Kim, and Jin-Young. A bayesian beta  
486 distribution model for estimating rainfall IDF curves in a changing climate. *Journal of*  
487 *Hydrology*. 540, 744-756 (2016).
- 488 22. L Nicótina, Celegon, E. A., Rinaldo, A., and Marani, M. On the impact of rainfall  
489 patterns on the hydrologic response. *Water Resources Research*. 44(12), 15-28 (2008).
- 490 23. Lebel, T., Bastin, G., Obled, C., and Creutin, J. D. On the accuracy of areal rainfall  
491 estimation: a case study. *Water Resources Research*, 23(11), 2123-2134 (1987).
- 492 24. Lin, G. F., and Jhong, B. C. A real-time forecasting model for the spatial  
493 distribution of typhoon rainfall-sciencedirect. *Journal of Hydrology*. 521, 302-313  
494 (2015).
- 495 25. Loève, M., Probability Theory, Princeton, N.J.: D Van Nostrand, (1955).
- 496 26. Marani, M. Correction to “non-power-law-scale properties of rainfall in space and

- 497 time". *Water Resources Research*. 41(8), 323-333 (2005).
- 498 27. Mekonnen, K., Melesse, A. M., and Woldesenbet, T. A. Spatial evaluation of  
499 satellite-retrieved extreme rainfall rates in the upper awash river basin, Ethiopia.  
500 *Atmospheric Research*. 249, 105297 (2020).
- 501 28. Meric, G., Lentz, C., Smeltz, W., and Meric, I. International evidence on market  
502 linkages after the 2008 stock market crash. *International Journal of Business and*  
503 *Finance Research*. 6(4): 45-57 (2012).
- 504 29. Northrop, P. A clustered spatial-temporal model of rainfall. *Proceedings of the*  
505 *Royal Society A: Mathematical, Physical and Engineering Sciences*. 454(1975), 1875-  
506 1888 (1998).
- 507 30. Palynchuk, B. A., and Guo, Y. A probabilistic description of rain storms  
508 incorporating peak intensities. *Journal of Hydrology*. 409(1-2), 71-80 (2011).
- 509 31. Pariyar, S. K., Keenlyside, N., Sorteberg, A., Spengler, T., and Ogawa, F. Factors  
510 affecting extreme rainfall events in the south pacific. *Weather and Climate Extremes*.  
511 29(6), 100262 (2020).
- 512 32. Redha, Wahidi, Semih, M., Olcmen, and James, et al. Different approaches to  
513 applying pod analysis to 3d3c data in a large measurement volume. *Fluid mechanics*  
514 *research*. 45(3), 187-201 (2018).
- 515 33. Roja, C., Hamza, V., Milind M., Ingle, S. T., Kuttippurath, J. Changes in large-scale  
516 circulation over the indo-pacific region and its association with 2018 kerala extreme  
517 rainfall event. *Atmospheric Research*. 263, 105809 (2021).
- 518 34. Romilly, T. G., and Gebremichael, M. Evaluation of satellite rainfall estimates over

519 Ethiopian river basins. *Hydrology and Earth System Sciences*. 15(5), 1505-1514 (2010).

520 35. Segond, M. L., Onof, C., and Wheeler, H. S. Spatial–temporal disaggregation of  
521 daily rainfall from a generalized linear model. *Journal of Hydrology*. 331(3-4), 674-689  
522 (2006).

523 36. Sohn, B. J., Han, H. J., and Seo, E. K. Validation of satellite-based high-resolution  
524 rainfall products over the Korean peninsula using data from a dense rain gauge network.  
525 *Journal of Applied Meteorology and Climatology*. 49(4), 367-70 (2010).

526 37. Taira, K., Brunton, S. L., Dawson, S., Rowley, C. W., Colonius, T., and Mckeon, B.  
527 J. Modal analysis of fluid flows: an overview. *AIAA Journal*. 55(12), 4013-4041 (2017).

528 38. Terink, W., Leijnse, H., van den Eertwegh, G., and Uijlenhoet, R. Spatial  
529 resolutions in areal rainfall estimation and their impact on hydrological simulations of  
530 a lowland catchment. *Journal of Hydrology*. 563, 319-335 (2018).

531 39. Van, d. B. C. Z., Leijnse, H., Torfs, P., and Uijlenhoet, R. Climatology of daily  
532 rainfall semi-variance in the netherlands. *Hydrology and Earth System Sciences*. 15(1),  
533 171-183 (2010).

534 40. Varraso, R., Garcia-Aymerich, J., Monier, F., Moual, N. L., Battle, J. D., and  
535 Miranda, G., et al. Assessment of dietary patterns in nutritional epidemiology: principal  
536 component analysis compared with confirmatory factor analysis<sup>1,2,3</sup>. *The American*  
537 *Journal of Clinical Nutrition*. 5: 1079-1092 (2012).

538 41. Wheeler, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., and Yang, C.  
539 Spatial-temporal rainfall modeling for flood risk estimation. *Stochastic Environmental*  
540 *Research and Risk Assessment*. 19(6), 403-416 (2005).

- 541 42. Wilks, D. S. Multisite generalization of a daily stochastic precipitation generation  
542 model. *Journal of Hydrology*. 210(1), 178-191 (1998).
- 543 43. Worqlul, A., W., Steenhuis, T., and S., et al. Comparison of rainfall estimations by  
544 trmm 3b42, mpeg and cfsr with ground-observed data for the lake tana basin in Ethiopia.  
545 *Hydrology and Earth System Sciences*. 18(12), 4871-4881 (2014).
- 546 44. Yang, C., Chandler, R. E., Isham, V. S., and Wheater, H. S. Spatial-temporal rainfall  
547 simulation using generalized linear models. *Water resources research*. 41(11),  
548 p.W11415.1-W11415.13 (2005).
- 549

550 **List of Tables**

551 **Table 1.** Latitude and longitude of the case studies

Ungauged basins	Latitude and longitude	Refactoring templates	Latitude and longitude
Hongshan	114.4°E, 30.5°N	Xinzhou	115.0°E, 31.0°N
Jianli	113.1°E, 30.0°N	Tianmen	113.1°E, 30.5°N
Sui	113.8°E, 31.5°N	Guangshui	113.8°E, 32.0°N
Xuanen	109.4°E, 30.0°N	Hefeng	110.0°E, 30.0°N

552

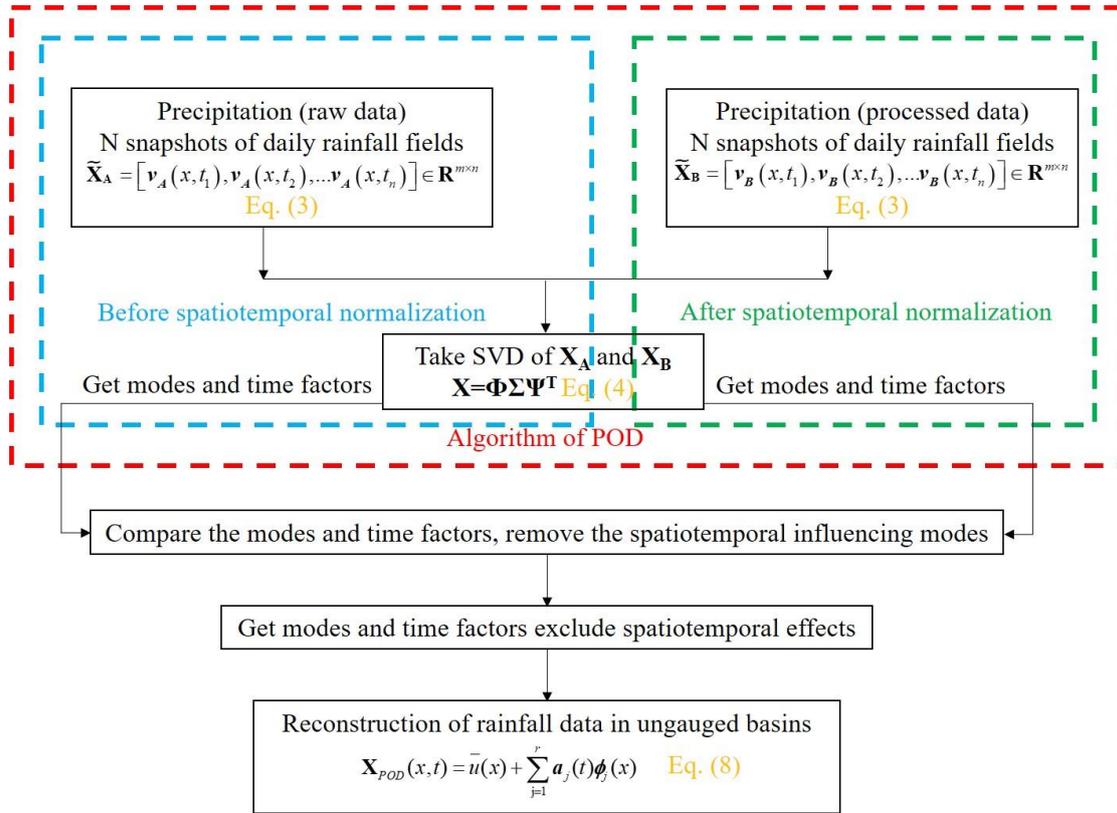
553 **Table 2.** Statistical indicators for validation of interpolation results of different

554 methods

Methods	RMSE (mm/h)	Corr	RE of rain peak (%)
Thiessen Polygon	3.22	0.60	83.56
IDW	1.83	0.85	47.57
Kriging	2.19	0.79	49.84
POD	2.09	0.89	32.40

555

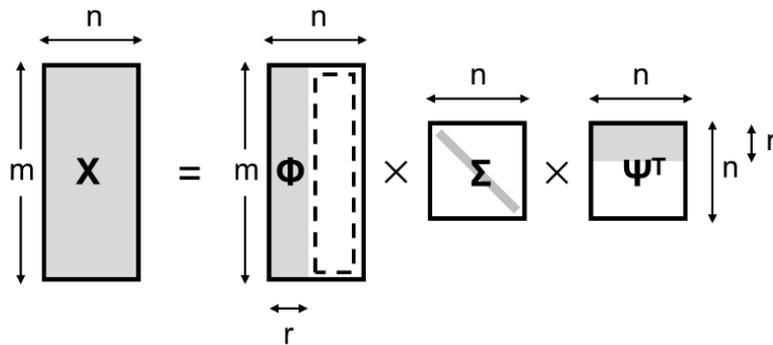
556 **List of Figures**



557

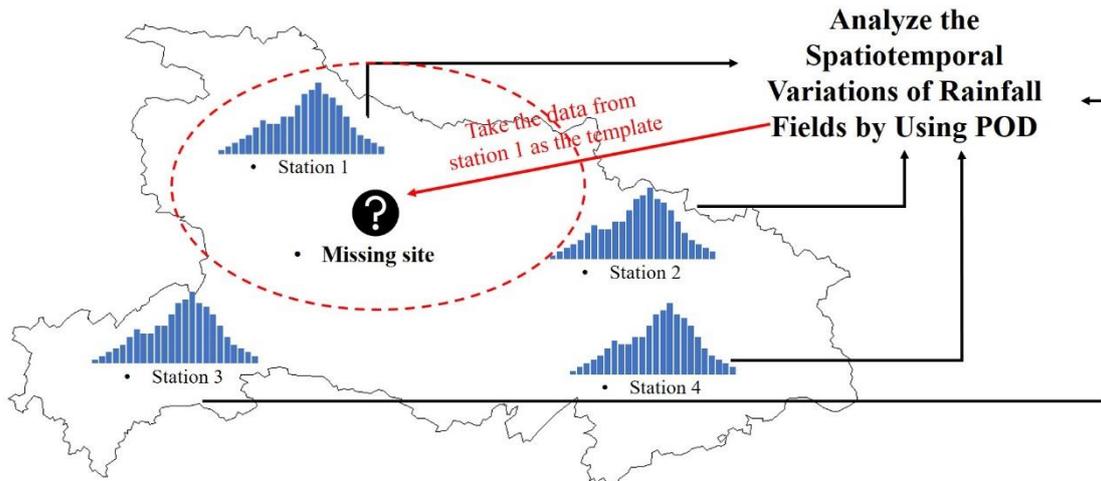
558 **Fig. 1.** Flow chart depicting the methodology proposed in this study.

559



560

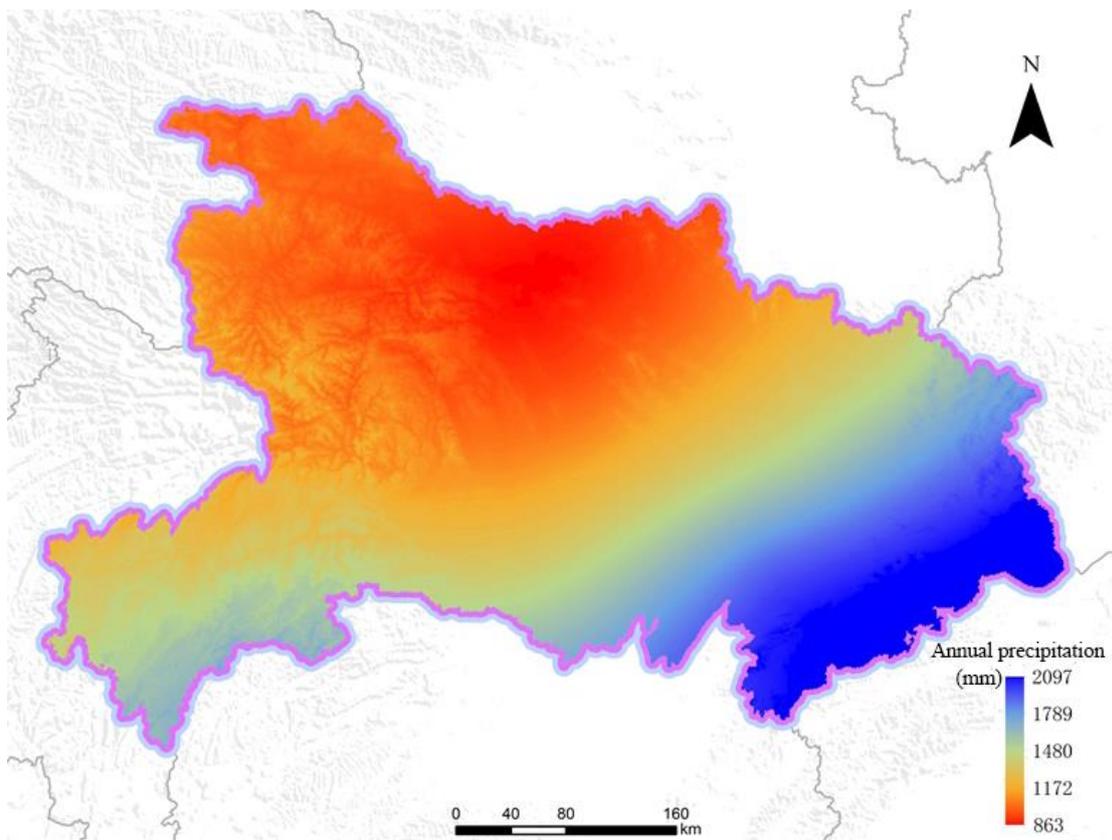
561 **Fig. 2.** Graphical representation of singular value decomposition.



562

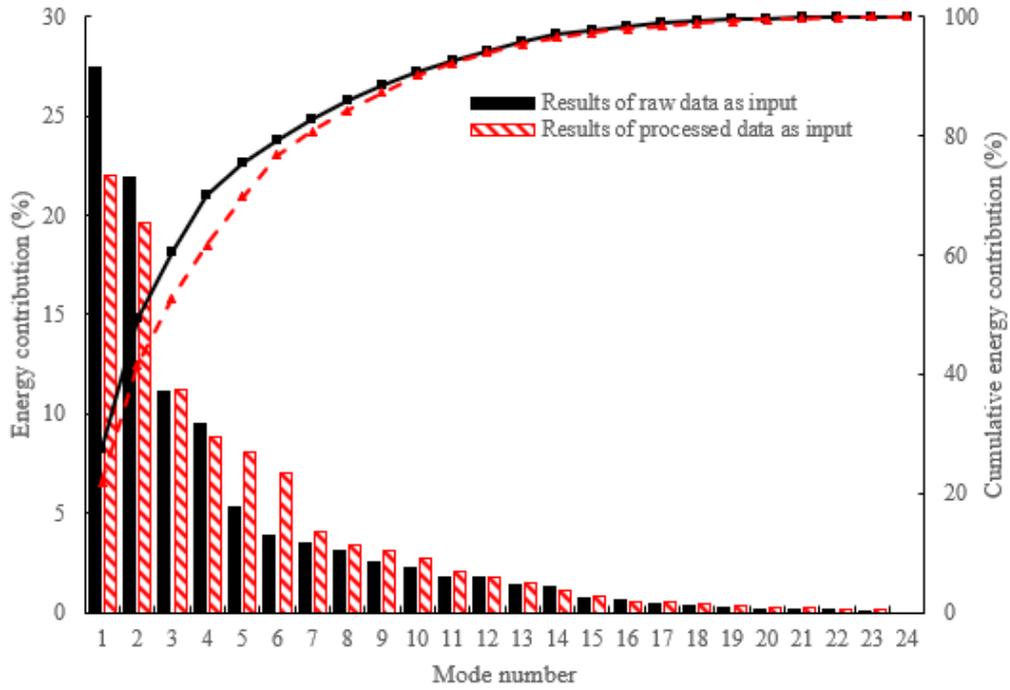
563 **Fig. 3.** Reconstruction of rainfall processes at ungauged basins.

564



565

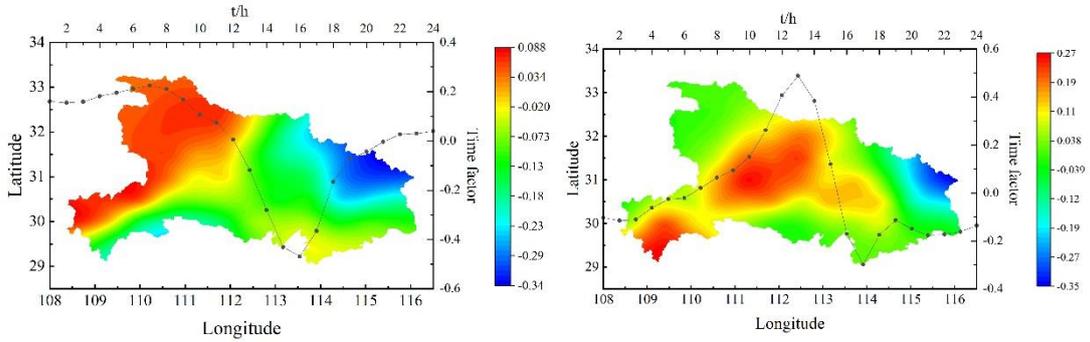
566 **Fig. 4.** Distribution of annual rainfall contours in Hubei Province.



567

568 **Fig. 5.** The energy contribution of the modes.

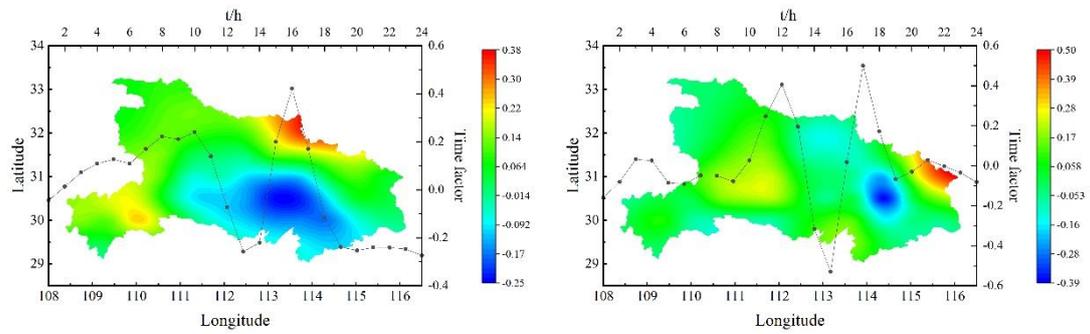
569



570

571 (a) The first mode

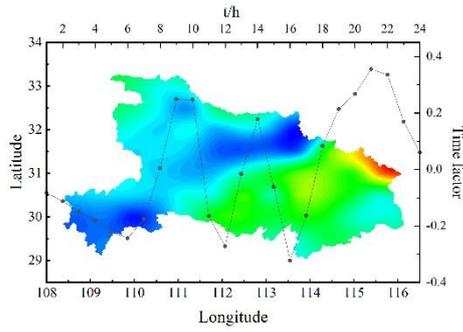
(b) The second mode



572

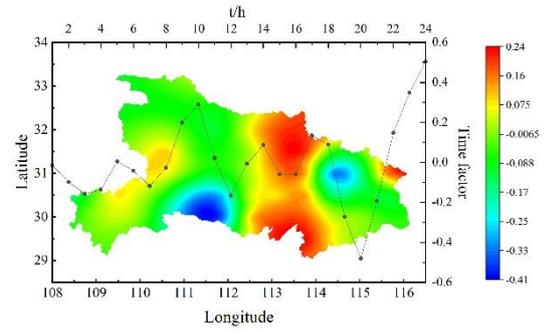
573 (c) The third mode

(d) The fourth mode



574

(e)The fifth mode

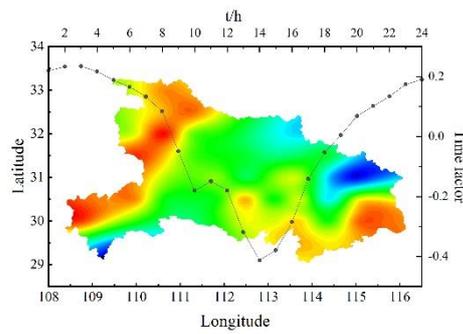


575

(f)The sixth mode

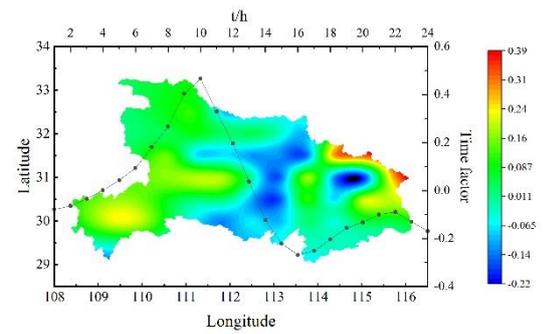
576 **Fig. 6.** Top 6 modes and corresponding time factors before spatiotemporal  
 577 normalization.

578



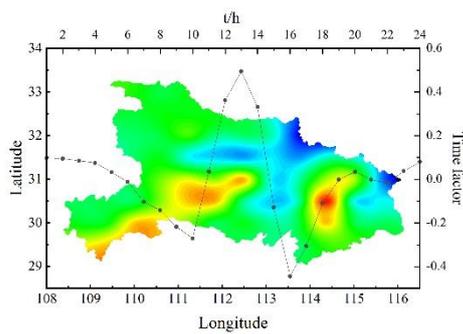
579

(a)The first mode



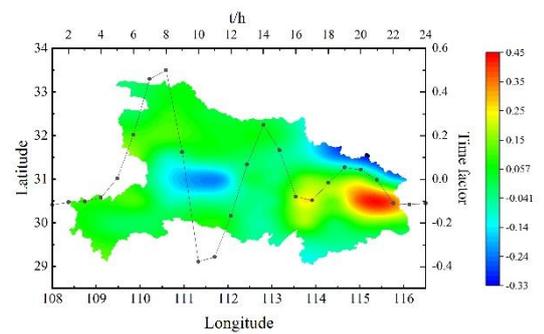
580

(b)The second mode



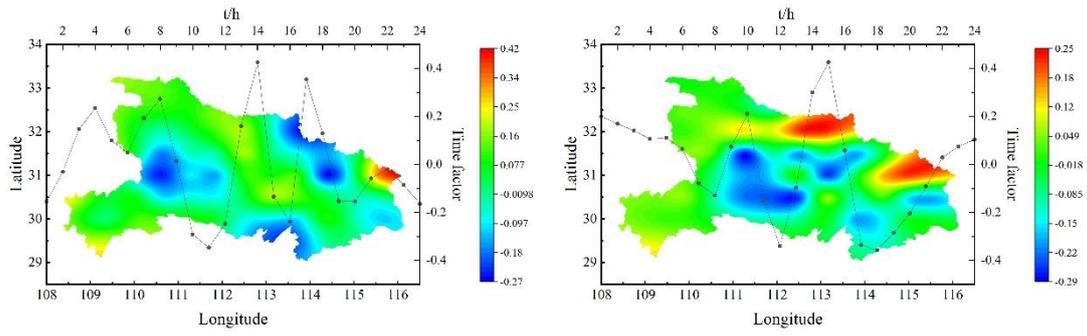
581

(c)The third mode



582

(d)The fourth mode



583

584

(e)The fifth mode

(f)The sixth mode

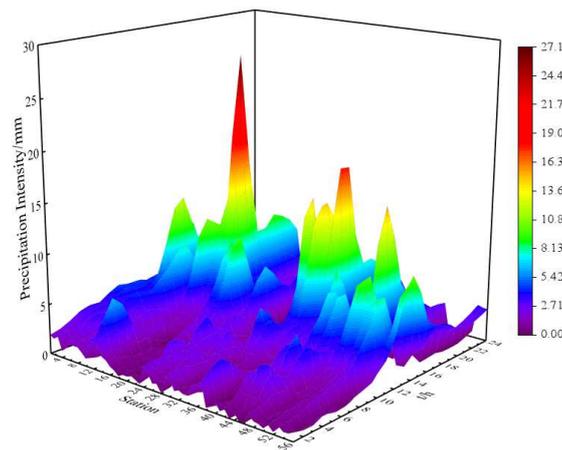
585

**Fig. 7.** Top 6 modes and corresponding time factors after spatiotemporal

586

normalization.

587

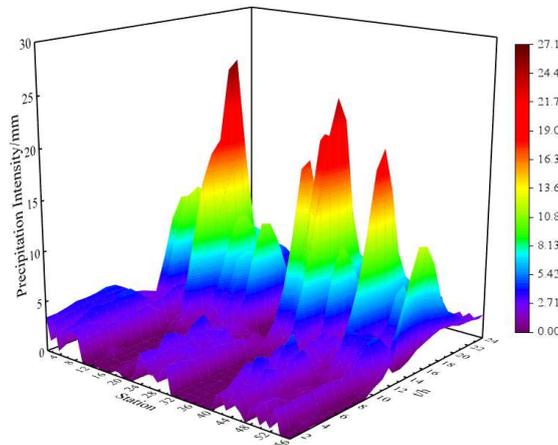


588

589

**Fig. 8.** The reconstructed precipitation field before spatiotemporal normalization.

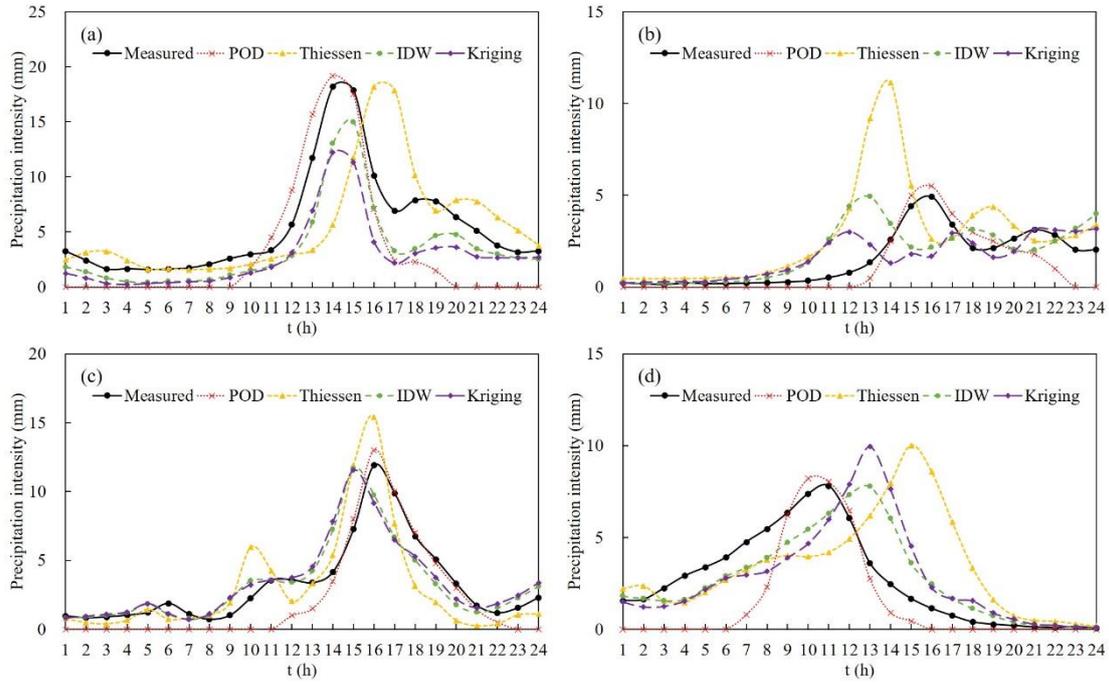
590



591

592

**Fig. 9.** The reconstructed precipitation field after removing spatiotemporal variations.

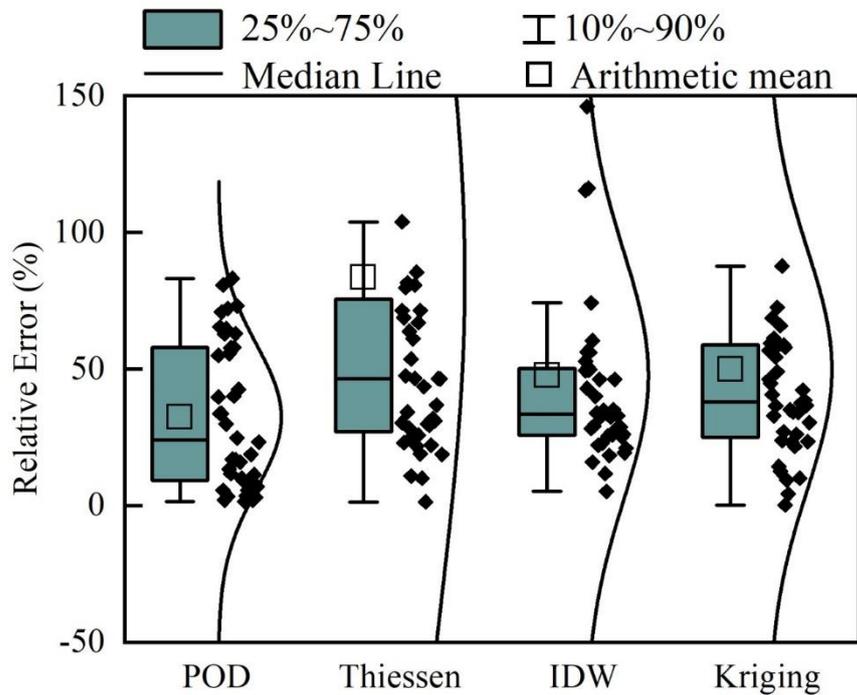


593

594 **Fig. 10.** A 24-h rainfall process obtained by different methods in ungauged basins: (a)

595 Hongshan District, (b) Jianli County, (c) Sui County, and (d) Xuanen County.

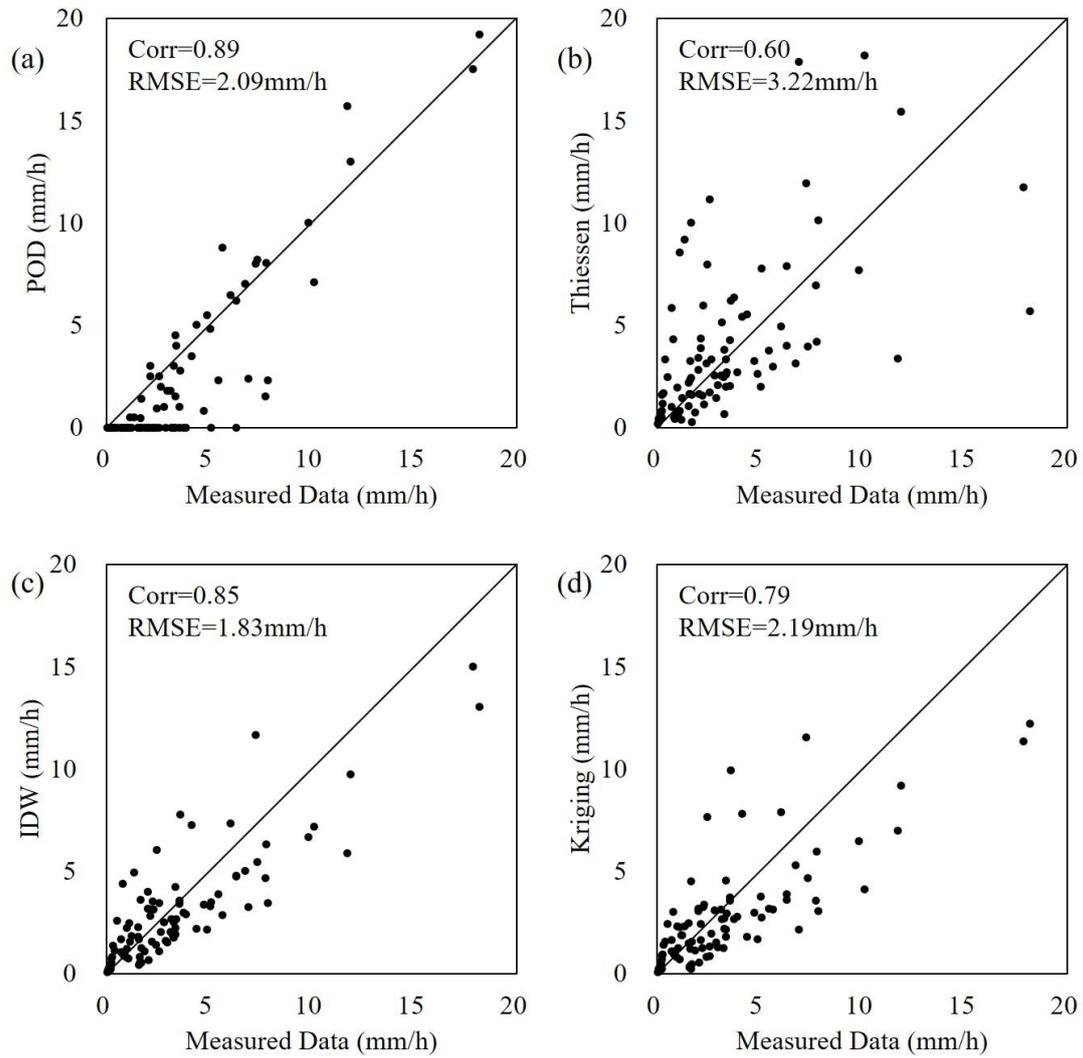
596



597

598 **Fig. 11.** Relative error box plots of the peak portion of simulated rainfall obtained by

599 different methods.



601

602 **Fig. 12.** Correlation diagram of rainfall obtained by different methods and measured

603 rainfall.