Acquisition of rainfall in ungauged basins: a study of rainfall distribution heterogeneity based on a new method

Ye Zhao¹,² · Xiang Zhang¹,² · Feng Xiong³ · Shuying Liu¹,² · Yao Wang¹,² · Changmei Liang⁴

Received: 7 January 2022 / Accepted: 2 June 2022 / Published online: 29 June 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
High-density rainfall data is always desired to capture the heterogeneity of precipitation to accurately describe the components of the hydrological cycle. However, equipping and maintaining a high-density rain gauge network involve high costs, and the existing rain gauges are often unable to meet the density requirements. The objective of this study is to provide a new method to analyze the spatiotemporal variability of the precipitation field and to solve the problem of insufficient site density. To this end, the proper orthogonal decomposition (POD) method is proposed, which can analyze the spatial distribution characteristics of rainfall fields to solve data shortages. To demonstrate the feasibility and advantages of the proposed methodology, four districts and counties (Hongshan District, Jianli County, Sui County, and Xuanen County) in Hubei province in China were selected as case studies. The principal results are as follows. (1) The proposed method is effective in analyzing the spatiotemporal variability of the rainfall field to reconstruct rainfall processes in ungauged basins. (2) Compared with the commonly used Thiessen polygon method, the inverse distance weighting method, and the Kriging method, POD is more accurate and convenient, and the root mean squared error is reduced from 3.22, 1.83, 2.19 to 2.09; the correlation coefficients are improved from 0.60, 0.85, 0.79 to 0.89, respectively. (3) The POD method performs particularly well in simulating the peak value and the peak time and can offer a meaningful reference for analyzing the spatial distribution of rainfall.

Keywords Rainfall · Proper orthogonal decomposition · Spatiotemporal variations · Ungauged basin · Interpolation

* Xiang Zhang
zhangxiang@whu.edu.cn

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China
² Hubei Key Lab of Water System Science for Sponge City Construction, Wuhan 430072, China
³ Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China
⁴ Hubei Water Resources Research Institute, Wuhan 430070, China
1 Introduction

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

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

With the development of remote sensing technology, satellite rainfall monitoring has become a viable option to complement rain gauge observations. Several high-resolution satellite rainfall estimates are increasingly available at a quasi-global scale (Km et al. 2021). These products have the potential to address the temporal and spatial sampling limitations of rainfall stations. However, remotely sensed rainfall products are subject to systematic bias and random error (Kidd 2010), and their performance is highly affected by the topography and climatic conditions of the investigated regions (Romilly and Gebremichael 2010; Worqlul et al. 2014; Baez-Villanueva et al. 2018; Mekonnen et al. 2020). In addition, to represent the full spatial distribution of rainfall, massive satellite data are required; however, due to satellite orbit limitations, many satellite measurements are not continuous, and long records of satellite data are extremely limited (Segond et al. 2006).

The fusion of measured and satellite rainfall data is a new trend in quantitative rainfall analysis, and the consistent spatial and temporal resolution of multi-source data is a prerequisite for data fusion. In addition, different engineering projects have different demands for rainfall data with different spatial and temporal resolutions. Therefore, the interpolation method of precipitation, i.e., the spatial variation analysis of precipitation, has been widely studied (Dey and Mujumdar 2019). There are various spatial interpolation methods for rainfall data, such as the Thiessen polygon method, the inverse distance weighting method (IDW) method, and the Kriging method, which are commonly used (Jin and Heap 2011). The Thiessen polygon method proceeds as follows (Croley and Hartmann 1985): The interpolated points are connected two by two and the vertical bisectors of the connecting lines are made. The vertical bisectors intersect to form several polygons, thus dividing the large watershed into several subregions, each with one rainfall station, and the rainfall value in each region is represented by the measured value of the rainfall station within the region. The Thiessen polygon method assumes that rainfall varies abruptly across the boundary of the polygon and rainfall is uniformly distributed within the subregion, which is unphysical. The IDW method assumes that the interpolated points are more influenced by the
closer stations than the far ones, and thus uses the reciprocal of the squared distance as the weighting factor (Chen et al. 2015). The disadvantage of this method is that the interpolation error is large when the data is significantly different from the neighboring points. The Kriging method (Armstrong and Matheron 1986) considers not only geographical features but also spatial correlation and variability of rainfall data and uses a covariance function to make the variables unbiased and optimal. However, when there are few rainfall stations, it is quite difficult to obtain the covariance function necessary for the Kriging method, which will also affect the interpolation effect to some extent (Gilbert and Simpson 1985). In addition, these methods are interpolated for a single time point, and the acquisition of rainfall processes for long time series requires iterative operations, which is very time-consuming.

To accurately and efficiently simulate the rainfall processes with high temporal resolution in large basins, it is crucial to fully consider the heterogeneity of rainfall fields. Many factors may contribute to the variability of precipitation, including tropical disturbances, El-Niño southern oscillation, Madden Julian oscillation, and so on, while the influence of spatial factors is the most significant (Pariyar et al. 2020). Variations in rainfall processes at different locations are often directly reflected in the temporal distribution of rainfall; therefore, they can be considered as “spatiotemporal variations”. For example, owing to the different locations of sites B and A, rainfall at site B might occur later than at site A, and the rainfall pattern at site B might be gentler (steeper) owing to weakening (enhancement) of the air mass (Palynchuk and Guo 2011; Lin and Jhong 2015). Several scholars have tried to represent rainfall in continuous space and time mathematically. For instance, Northrop (1998) developed a spatial–temporal model based on the Bartlett–Lewis process for the temporal evolution of storm cells and the Neyman–Scott process for spatial arrivals. Wilks (1998) proposed an extension from a single station to multiple stations by driving single-station models at various locations in New York with temporally independent but spatially correlated random numbers. Cowpertwait et al. (2002) proposed a spatial–temporal model based on the Neyman–Scott mechanism and tested the model on nine stations in Italy. Yang et al. (2005) used generalized linear models for the generation of multisite rainfall series in Southern England. Abas et al. (2014) presented a stochastic rainfall model for the generation of hourly rainfall data in an urban area in Malaysia. However, due to the influence of topography and climate, there is a large spatial gradient in rainfall intensity, such models currently assume spatial stationarity is not credible. (Alam and Elshorbagy 2015; Kwon et al. 2016). Improvements to include nonstationary properties of precipitation, to represent (1) the temporal characteristics of rainfall and (2) the spatial features of the data such as the location of high values of rainfall, topographic effects, or spatial continuity, are still under development (Wheater et al. 2005).

Proper orthogonal decomposition (POD) is an effective dimensionality reduction method that uses basis functions extracted from a large number of known data to describe the original data in an optimal approximation (Aubry 1991). Because of its ease of interpretation and broad applicability to data from both simulations and experiments (Kutz et al. 2016), the POD technique is now used widely in fields such as fluid mechanics, aerodynamics, economics, statistics, and psychology to identify and analyze spatial distribution characteristics of some variables. For example, Meric et al. (2012) applied the POD method to extract information on the dynamic spatiotemporal evolution of global stock markets during the 2008 economic crisis, Cheveigné and Simon (2007) applied the POD method to neuroscience to extract the evolutionary structure of neurons in the human brain, and Varraso et al. (2012) used POD to extract and analyze the main spatial transmission of epidemic characteristics. However, this method has rarely been used in rainfall analysis. Considering that the structure of the vibration, flow, and aerodynamic fields is similar to that of the rainfall field, these fields also change
constantly with time and space. Besides, the velocity in the flow field and the pressure in the aerodynamic field are similar to the rainfall intensity in the precipitation field. Therefore, the POD method also has potential for applicability in the field of rainfall simulation, where it can be used to analyze the spatial distribution characteristics of rainfall fields and as an interpolation-like method to obtain rainfall data in ungauged basins.

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

2 Methodology

The POD method, also known as the Karhunen-Loève procedure (Karhunen 1946; Loève 1955), is an effective dimensionality reduction method that can extract modes based on optimizing the mean square of the field variables being examined (Taira et al. 2017). It provides an objective algorithm to decompose a set of data into a minimal number of modes, each reflecting the influence of one or some physical factors on the field being decomposed, and these modes can be used to make the optimal approximate description of the original data. In this study, POD is implemented to decompose the precipitation field \( (X \in \mathbb{R}^{m \times n}) \) into several orthogonal modes that reflect the influence of several physical factors on the rainfall field, and the modes reflecting the influence of spatiotemporal factors can be found through the subsequent comparative analysis. By reconstructing the rainfall field with the remaining modes after removing the modes reflecting spatiotemporal characteristics, the rainfall field structure independent of spatiotemporal factors can be obtained, and the rainfall processes in the ungauged basins can be reconstructed indirectly. The detailed steps and physical meanings of the variables involved in using the POD method to obtain the spatiotemporal structure of the rainfall field and to obtain data for ungauged basins are described below (Fig. 1).

2.1 Sampling and preprocessing of data

To study the influence of spatiotemporal factors on the rainfall field structure, two data sampling sessions are required. The first sample matrix \( X_A \) consists of a certain rainfall process of sufficient duration at multiple stations in the study area, and the other sample matrix \( X_B \) is manually spatiotemporally normalized based on the previous sample. By manual alignment (Roja et al. 2021), the rainfall processes at different stations start and end simultaneously in \( X_B \). With the above data collation, the original determinant \( X_A \) has spatiotemporal variations, and the processed determinant \( X_B \) can be considered to be uniformly distributed in time and space. \( X_A \) and \( X_B \) can be denoted as follows:

\[
X_A = [u_A(x, t_1), u_A(x, t_2), \ldots, u_A(x, t_n)] \in \mathbb{R}^{m \times n} \\
X_B = [u_B(x, t_1), u_B(x, t_2), \ldots, u_B(x, t_n)] \in \mathbb{R}^{m \times n},
\]

where \( n \) is the length of the rainfall duration and \( m \) reflects the number of rain gauges in the study area. In fluid dynamics, the measured state \( u(x, t_n) \) is related to the physics, i.e.,
the velocity, vorticity, and stream functions at each discretized spatial grid point. For precipitation data collected from a rainfall field, the rainfall intensity at different rain gauges at time \( t_n \) can be denoted as 
\[
\frac{u}{u(x, t_n)} = \frac{u}{u(x, t_n)} - u(x),
\]
and the fluctuating rainfall intensity after removing \( u(x) \) at time \( t_n \) can be denoted as follows:
\[
\tilde{X}_A = [v_A(x, t_1), v_A(x, t_2), ..., v_A(x, t_n)] \in \mathbb{R}^{m \times n},
\]
\[
\tilde{X}_B = [v_B(x, t_1), v_B(x, t_2), ..., v_B(x, t_n)] \in \mathbb{R}^{m \times n},
\]
where \( v(x, t_i) \) stands for the \( i \)th fluctuating rainfall intensity at different stations in the study area.

\section{2.2 POD calculation on the sample matrices}

Various methods can be used to calculate POD, and the method selected for use in this study is the popular SVD (singular value decomposition)-based approach (Eckart and Young 1936). SVD can decompose \( X_A \) and \( X_B \) to get the modes and time factors. The mode represents the distribution structure of the rainfall parameter (intensity) in a certain physical sense, and the corresponding time factor reflects the trend of the mode over time. The purpose of applying POD in this study is to find the mode(s) that reflect spatiotemporal variations of the rainfall field, and to achieve this purpose, \( \tilde{X}_A \) and \( \tilde{X}_B \) are input to POD, respectively.
In matrix form, the data matrix $\mathbf{X}_A$ and $\mathbf{X}_B$ can be decomposed with SVD as follows (Fig. 2):

$$
\mathbf{\tilde{X}}_A = \Phi_A \Sigma_A \Psi_T^A,
$$

$$
\mathbf{\tilde{X}}_B = \Phi_B \Sigma_B \Psi_T^B,
$$

(4)

where the $m \times n$ matrices $\Phi_A = [\phi_{A1}, \phi_{A2}, \cdots, \phi_{An}]$ and $\Phi_B = [\phi_{B1}, \phi_{B2}, \cdots, \phi_{Bn}]$ are composed of the modes $\phi_A$ and $\phi_B$. $\phi_A$ reflects the spatial distribution information of rainfall intensity containing rainfall spatial variability, and $\phi_B$ reflects the structure of the rainfall field after removing rainfall spatial variability. The $n \times n$ matrix $\Psi_A = [a_{A1}, a_{A2}, \cdots, a_{An}]$ and $\Psi_B = [a_{B1}, a_{B2}, \cdots, a_{Bn}]$ contain coefficients representing the temporal evolution of the modes $\Psi_A$ and $\Psi_B$, respectively. The superscript T represents a conjugate transpose matrix operation (Higham et al. 2017). Matrix $\Sigma$ holds singular values ($\sigma_1, \sigma_2, \cdots, \sigma_n$) along its diagonal. $\lambda_j (\lambda_j = \sigma_j^2)$ is the energy contribution of the $j$th mode to the total variance, indicating the influence of the $j$th mode on the overall rainfall field. $\lambda$ is listed in descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, the larger the value of $\lambda$, the greater the mode energy contribution. Furthermore, the percentage rainfall intensity energy contribution ($E_i$) of each POD mode can be obtained as follows:

$$
E_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100\%.
$$

(5)

### 2.3 Reconstruction of rainfall processes in ungauged basins

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

$$
\sum_{j=1}^{r} \lambda_j / \sum_{j=1}^{n} \lambda_j \approx 1.
$$

(6)

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

$$v(x, t) = \sum_{j=1}^{r} a_j(t)\phi_j(x),$$

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

$$X_{POD}(x, t) = \bar{u}(x) + \sum_{j=1}^{r} a_j(t)\phi_j(x).$$

Because the temporal and spatial variations of the precipitation field are eliminated, the errors between precipitations at close locations are greatly reduced, and the precipitation processes in the ungauged basin can be represented by precipitation processes at an adjacent station (Fig. 3).

2.4 Verification methods

To assess the performance of the methods used in this study, we used three performance statistics widely used to compare the model performance such as root mean square error (RMSE), relative error (RE), and correlation coefficient (Corr). Based on the results, performance of these methods is presented in detail.

2.4.1 Root mean squared error (RMSE)

The RMSE (Eq. (9)) estimates the standard deviation between the observed value and the simulated value. A small RMSE value indicates better performance, whereas a higher value indicates a poor performance.
\[ \text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (Q_t^o - Q_t^m)^2}, \]  

where \(Q_o\) and \(Q_m\) refer to the observed and simulated values, respectively, and \(T\) is the total number of observations.

2.4.2 Relative error (RE)

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

\[ RE = \frac{1}{T} \left[ \sum_{t=1}^{T} \left| 1 - \frac{Q_t^m}{Q_t^o} \right| \right] \times 100\%. \]  

2.4.3 Correlation coefficient (Corr)

The Corr is a statistical indicator to reflect the closeness of the correlation between variables and its value ranges between −1 and 1 representing a perfect negative and positive correlation, respectively. The correlation coefficient is calculated by the product-difference method and is based on the deviation of the two variables from their respective means. The formula for calculating correlation can be represented as:

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

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

3 Study area and data

Hubei Province is located in central China, between latitude 29° 01’ 53” -33°6’ 47” N and longitude 108°21’ 42”-116° 07’ 50” E with a total area of 185900km². Mountains, hills, and plain lakes account for 56%, 24%, and 20% of the total area of Hubei province, respectively. Except for the alpine climate in the high-mountain areas, most parts of the province belong to the subtropical monsoon humid climate, with sufficient light energy, abundant rainfall, and long frost-free periods. The average annual precipitation in Hubei province is between 800 and 1600 mm. The regional distribution of precipitation shows a decreasing trend from south to north (Fig. 4), with a maximum of 1400–1600 mm in the southwest and a minimum of 800–1000 mm in the northwest. The seasonality of precipitation distribution is obvious, with the most in summer (300–700 mm) and the least in winter (30–190 mm), respectively. The rainy season is from mid-June to mid-July, with the largest rainfall and intensity. To study the applicability of the method recommended in this paper, we selected several extreme rainstorms in Hubei Province for analysis. Hourly rainfall data at 56 stations in Hubei province are downloaded from European Centre for Medium-Range Weather Forecasts.
4 Results and discussion

4.1 Reconstruction of rainfall processes at ungauged basins

Because of the large spatially varying gradients of precipitation in Hongshan District, Jianli County, Sui County, and Xuanen County in Hubei province, these locations are assumed to be the ungauged basins for rainfall reconstruction (Table 1). Several extreme rainfall events in Hubei Province in 2020 were used as the original input ($X_A$) of POD to obtain the modes with spatiotemporal variability, and the manually normalized data ($X_B$) were input to obtain the modes with spatiotemporal variability removed. Figure 5 shows the energy contribution and the cumulative energy contribution of the modes obtained by decomposing $X_A$ and $X_B$ as POD inputs, respectively. The energy contribution of the modes derived from the decomposition of $X_A$ shows a greater descending trend, where the cumulative energy contribution of the first two modes is approximately 10% higher than that of the modes from the $X_B$ decomposition. This indicates that the $X_A$-decomposed modes are more representative, which is consistent with the conjecture that the unprocessed data contain more spatiotemporal information. From the fifth mode onward, the energy contribution

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      |
of the $X_B$-decomposed modes becomes larger than that of the $X_A$-decomposed ones. And after the ninth mode, the cumulative energy contribution of the modes obtained from the POD of these two data sets is almost the same. In both cases, the cumulative energy contribution of the first six modes exceeds 80%, so these modes are selected as the predominant modes for subsequent analysis and reconstruction of the precipitation field.

Figure 6 shows the predominant modes and their corresponding time factors before spatiotemporal normalization, i.e., raw data $X_A$ are used as input to the POD. As the energy contribution of the mode decreases, a higher proportion of null values is observed for modes with values close to zero, which is in line with the physical meaning of the energy contribution. In addition, as the mode’s energy contribution decreases, the time factor corresponding to the mode fluctuates more frequently, which indicates that the mode becomes more unstable.

Because the energy contribution of the first mode exceeds 27.44%, it can be assumed that the reconstruction based on the first mode reflects the main structure of the precipitation field. The values of the first mode are mostly negative and its corresponding time factor is in the inverted triangular shape; therefore, a positive triangular rainfall intensity process can be obtained by multiplying the first mode and its corresponding time factor (Eq. (7)). The majority of the values of the second mode are positive, with the corresponding time factor appearing in the shape of a triangle. A similar rainfall process that increases firstly and then decreases can be obtained by multiplying the two. This can be interpreted as the structure of this precipitation field being a single-peak rainfall process. The values of the third to sixth modes are mostly around zero, and their corresponding time factors fluctuate frequently. This indicates that they play little role in the reconstructed rainfall field, but there may still be modes reflecting spatiotemporal factors in them.

To find the modes characterizing the variations of time and space, the processed data $X_B$ are used as input to POD to get the modes after spatiotemporal normalization. Similar to the modes obtained from the original data as input, the proportion of null values and the fluctuation frequency of the time factors increase with the decrease of the energy contribution of the modes. However, some variations are observed in the predominant modes and their corresponding time factors due to different inputs (Fig. 7). Compared with the first mode before spatiotemporal normalization, the values of the first mode after spatiotemporal normalization are also mostly negative, decreasing from northwest
to northeast and slightly higher in the southeast. The main pattern of precipitation after spatiotemporal normalization is also a single-peak rainfall with a backward peak, which indicates that the main pattern of precipitation does not change significantly after the normalization of time and space. In addition, the distribution of the values of the fourth mode is almost the same before and after normalization; the values of the fifth mode show an increasing trend from west to east and reach the maximum at northeast, and the trends of time factors of the fifth mode before and after the normalization are the same. Therefore, it can be assumed that the first, fourth, and fifth modes do not reflect the influence of spatiotemporal factors on the structure of the precipitation field. However, the second, third, and sixth modes change significantly due to the normalization of time and space. Therefore, it can be assumed that the superimposed effect of the second, third, and sixth modes reflects the spatiotemporal variations of the precipitation field. Therefore, the reconstructed precipitation field after removing these three modes can be regarded as the result of removing the spatiotemporal variations. Figure 8 shows the reconstruction of precipitation fields before and after spatiotemporal normalization. The comparison shows that the rainfall processes at different stations are more consistent and the peaks are more pronounced after the removal of spatiotemporal variations. This confirms that modes 2, 3, and 6 contain information on the spatial and temporal distribution of rainfall, and modes 1, 4, and 5 better reflect the peak characteristics of rainfall.
4.2 Comparative analysis of different methods

Based on the results obtained in the previous section, the rainfall processes at the ungauged sites are reconstructed according to the method introduced in Sect. 2.3. To illustrate the
implementability and superiority of the proposed method, three most commonly used interpolation methods in current hydrological engineering: the Thiessen polygon method, the IDW method, and the Kriging method, were selected for comparative study. The rainfall data obtained by different methods as well as the measured rainfall data in the ungauged stations are shown in Fig. 9. It can be seen that a 24-h rainfall varies in magnitude and time distribution in different regions of Hubei Province, and the interpolation results of different methods applied in different regions also have great differences.

In comparison, the Thiessen polygon method performs poorly in the case study of this paper, especially in Jianli County where the rainfall gauge distribution is relatively sparse, because for district- and county-scale regions, the simplification of assuming an abrupt change in the spatial distribution of rainfall at a specific boundary is very rough. The accuracy of the interpolation results for rain intensity above 5 mm/h is low, with a Corr of only 0.6 and an RMSE of 3.22 mm/h (Fig. 10, Table 2). In addition, the interpolation results of peak time in Hongshan District and Sui County are late, which is unfavorable to the prevention and control of rain floods. Therefore, the Thiessen polygon method is only an alternate interpolation option for small areas where other methods cannot be employed. The interpolation performance of the Kriging and IDW methods is slightly better overall, but the performance of the peak value and the peak time is not satisfactory.

The peak value of the rainfall obtained by the IDW method shows a significant improvement compared to the Thiessen polygon method, with an RMSE of only 1.83 mm/h and a Corr of 0.85 (Fig. 10, Table 2). However, the interpolation of the IDW method for rainfall exceeding 15 m/h tends to be small and the performance in terms of peak time is not satisfactory. For example, there are large errors in the simulation of peak time in Jianli and Xuanen counties. This may be attributed to the fact that the IDW method mainly focuses on the magnitude of rainfall at the adjacent stations and does not give enough attention to the peak time.

![Fig. 9 A 24-h rainfall process obtained by different methods in ungauged basins: a Hongshan District, b Jianli County, c Sui County, and d Xuanen County](image_url)
The interpolation results of the Kriging method are mostly a little smaller than the measured values, and the peak time is not accurate. This is unfavorable from the perspective of practical engineering safety. The Corr of the Kriging method is 0.79 and the RMSE is 2.19 mm/h, outperforming the Thiessen polygon method but worse than the IDW method (Fig. 10, Table 2). This large error of the Kriging method may be triggered by establishing the covariance matrix with insufficient dense sites in this study area.

The POD method exhibits a more accurate simulation of the peak value and peak time. The median and arithmetic mean of the relative errors of the POD simulations for the peak portion of the rainfall are minimal (Fig. 11), which is very meaningful for hydrological design. But perhaps because the extracted modes contain not only the spatiotemporal characteristics of the rainfall field but also part of the initial and final rainfall, the POD method is inaccurate in this case in simulating rainfall less than 5 mm/h at the beginning and end of the rainfall period, resulting in an RMSE of 2.09 mm/h. However, the POD method is more accurate for rainfall over 10 mm/h and the Corr exceeds 0.89 (Fig. 10, Table 2).

### Table 2: Statistical indicators for validation of interpolation results of different 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               |
5 Conclusions

Spatial interpolation of rainfall has been regarded as the most common method to obtain rainfall data in ungauged basins, and the nonconsistency of rainfall spatial distribution is the source of error in the interpolation method. In this study, POD is introduced to analyze the spatiotemporal variations of rainfall fields and is used as an interpolation-like method to obtain rainfall data in ungauged basins.

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

In the application to the case study area in Hubei Province, POD is proved an effective and reliable approach for obtaining rainfall data in ungauged basins. Compared with the Thiessen polygon method, the IDW method (Goovaerts 2000), and the Kriging method (Haberlandt 2007), the first advantage of the proposed method is that it has a significant improvement in the simulation of the peak value and peak time. Second, the use of the POD method to remove spatiotemporal correlation modes to reconstruct rainfall in ungauged basins is once and for all, and it is more efficient to obtain long series of rainfall data with the POD method than other methods that require repeated interpolations. Despite the above advantages, this method has the potential for further improvement. For example, how to separate the parts of the same mode that may contain different characteristics of the rainfall field and how to quantitatively identify the modes that characterize spatial and temporal properties remain to be discussed. It is hoped that this study will serve as a stepping stone for the readers to become familiar with POD, analyze complex issues associated with spatial heterogeneity of rainfall fields, and further advance the developments of this.
method in analyzing the effects of other factors (e.g., distance between stations and the ocean, slope, slope direction, wind speed) on the spatial distribution of rainfall data.

Acknowledgements The data sets used in this study are available on the corresponding websites of Era-5 and NCEP. We express our gratitude to the reviewers and editors for their comments on the manuscript.

Funding The work is funded by the National Key Research and Development Program of China (No.2019YFC0408901) and the National Natural Science Foundation of China (No.41890823).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abas N et al (2014) A comparative study of mixed exponential and Weibull distributions in a stochastic model replicating a tropical rainfall process. Theoret Appl Climatol 118(3):597–607
Alam MS, Elshorbagy A (2015) Quantification of the climate change-induced variations in intensity–duration–frequency curves in the Canadian prairies. J Hydrol 527:990–1005
Alexis B, Guy D (2004) Temporal and spatial resolution of rainfall measurements required for urban hydrology. J Hydrol 299(3–4):166–179
Armstrong M, Matheron G (1986) Disjunctive kriging revisited. I. II. Math Geol 8:711-728-729–742
Aubry N (1991) On the hidden beauty of the proper orthogonal decomposition. Theoret Comput Fluid Dyn 2(5):339–352
Baez-Villanueva OM, Zambrano-Bigiarini M, Ribbe L, Nauditt A, Giraldo-Osorio JD, Thinh NX (2018) Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. Atmos Res 213:34–50
Chen C, Zhao N, Yue T, Guo J (2015) A generalization of inverse distance weighting method via kernel regression and its application to surface modeling. Arab J Geosci 8:6623–6633
Cheveigné AD, Simon JZ (2007) Denoising based on time-shift PCA. J Neurosci Methods 165(2):297–305
Cowpertwait P, Kilsby CG, O’Connell PEA (2002) space-time neyman-scott model of rainfall: empirical analysis of extremes. Water Resour Res 38(8):6–1–6–14
Crole TE, Hartmann HC (1985) Resolving thiessen polygons. J Hydrol 76(3–4):363–379
Dey P, Mujumdar PP (2019) On the uniformity of rainfall distribution over India. J Hydrol 578:124017
Eckart C, Young G (1936) The approximation of one matrix by another of lower rank. Psychometrika 1(3):211–218
Faurès JM, Goodrich DC, Woolhiserv DA, Sorooshian S (1995) Impact of small-scale spatial rainfall variability on runoff modeling. J Hydrol 173(1):309–326
Gilbert RO, Simpson JC (1985) Kriging for estimating spatial pattern of contaminants: potential and problems. Environ Monit Assess 5(2):113
Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228(1–2):113–129
Haberlandt U (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J Hydrol 332(1–2):144–157
Higham JE et al (2017) Using modal decompositions to explain the sudden expansion of the mixing layer in the wake of a groyne in a shallow flow. Adv Water Resour 107:451–459
Jin L, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Eco Inform 6(3–4):228–241
Karschnik K (1946) Zur spektral theorie stochastischer prozesse. Annales Academiae Scientiarum Fennicae 34:13
Kidd C (2010) Satellite rainfall climatology: a review. Int J Climatol 21(9):1041–1066
Km A, Amm B, Taw A (2021) Effect of temporal sampling mismatches between satellite rainfall estimates and rain gauge observations on modelling extreme rainfall in the upper awash basin Ethiopia. J Hydrol 598(7):126467
Kutz JN, Brunton SL, Brunton BW and Proctor JL (2016) Dynamic mode decomposition: data-driven modeling of complex systems. Soc Ind Appl Math Phila

Springer
Kwon H-H, Lima CH, Kim J-Y (2016) A bayesian beta distribution model for estimating rainfall IDF curves in a changing climate. J Hydrol 540:744–756
Lebel T, Bastin G, Obled C, Creutin JD (1987) On the accuracy of areal rainfall estimation: a case study. Water Resour Res 23(11):2123–2134
Lin GF, Jhong BC (2015) A real-time forecasting model for the spatial distribution of typhoon rainfall-sciencedirect. J Hydrol 521:302–313
Loève M (1955) Probability theory, princeton, N.J.: D Van Nostrand
Marani M (2005) Correction to “non-power-law-scale properties of rainfall in space and time.” Water Resour Res 41(8):323–333
Mekonnen K, Melesse AM, Woldesenbet TA (2020) Spatial evaluation of satellite-retrieved extreme rainfall rates in the upper awash river basin. Ethiopia Atmos Res 249:105297
Meric G, Lenz C, Smeltz W, Meric I (2012) International evidence on market linkages after the 2008 stock market crash. Int J Bus Fin Res 6(4):45–57
Nicótila L, Celegen EA, Rinaldo A, Marani M (2008) On the impact of rainfall patterns on the hydrologic response. Water Resour Res 44(12):15–28
Northrop P (1998) A clustered spatial-temporal model of rainfall. Proc Royal Soc: Math, Phys Eng Sci 454(1975):1875–1888
Palynchuk BA, Guo Y (2011) A probabilistic description of rain storms incorporating peak intensities. J Hydrol 409(1–2):71–80
Pariyar SK, Keenlyside N, Sorteberg A, Spengler T, Ogawa F (2020) Factors affecting extreme rainfall events in the south pacific. Weather Climate Extrem 29(6):100262
Wahidi R, Olcmen SM, Hubner JP et al (2018) Different approaches to applying pod analysis to 3d3c data in a large measurement volume. Fluid Mech Res 45(3):187–201
Roja C, Hamza V, Milind M, Ingle ST, Kuttipurath J (2021) Changes in large-scale circulation over the indo-pacific region and its association with 2018 kerala extreme rainfall event. Atmos Res 263:105809
Romilly TG, Gebremichael M (2010) Evaluation of satellite rainfall estimates over Ethiopian river basins. Hydrolog Earth Syst Sci 15(5):1505–1514
Segond ML, Onof C, Wheater HS (2006) Spatial–temporal disaggregation of daily rainfall from a generalized linear model. J Hydrol 331(3–4):674–689
Sohn BJ, Han HJ, Seo EK (2010) Validation of satellite-based high-resolution rainfall products over the Korean peninsula using data from a dense rain gauge network. J Appl Meteorol Climatol 49(4):367–370
Taira K, Brunton SL, Dawson S, Rowley CW, Colomnus T, Mckeon BJ (2017) Modal analysis of fluid flows: an overview. AIAA J 55(12):4013–4041
Terink W, Leijnse H, van den Eertwegh G, Uijlenhoet R (2018) Spatial resolutions in areal rainfall estimation and their impact on hydrological simulations of a lowland catchment. J Hydrol 563:319–335
Van de Beek CZ, Leijnse H, Torfs F, Uijlenhoet R (2010) Climatology of daily rainfall semi-variance in the netherlands. Hydrolog Earth Syst Sci 15(1):171–183
Varraso R, Garcia-Aymerich J, Monier F, Moual NL, Batlle JD, Miranda G et al (2012) Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis.2,3. Am J Clin Nutr 5:1079–1092
Wheater HS, Chandler RE, Onof CJ, Isham VS, Bellone E, Yang C (2005) Spatial-temporal rainfall modeling for flood risk estimation. Stoch Env Res Risk Assess 19(6):403–416
Wilks DS (1998) Multisite generalization of a daily stochastic precipitation generation model. J Hydrol 210(1):178–191
Worqlul AW, Steenhuis TS et al (2014) Comparison of rainfall estimations by trmm 3b42, mpeg and cfsr with ground-observed data for the lake tana basin in Ethiopia. Hydrolog Earth Syst Sci 18(12):4871–4881
Yang C, Chandler RE, Isham VS, Wheater HS (2005) Spatial-temporal rainfall simulation using generalized linear models. Water Resour Res 41(11):W11415.1-W11415.13

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.