

Article



Spatiotemporal Modes of Short Time Rainstorms Based on High-Dimensional Data: A Case Study of the Urban Area of Beijing, China

Wei Liu¹, Sheng Chen² and Fuchang Tian^{1,*}

- ¹ School of Civil Engineering, Tianjin University, Tianjin 300072, China; waterliuwei@tju.edu.cn
- ² China Institute of Water Resources and Hydropower Research, Beijing 100038, China; chensheng@iwhr.com
- * Correspondence: tianfuchang@tju.edu.cn; Tel.: +86-22-27401156

Abstract: The identification of the characteristics of short time rainstorms in urban areas is a difficult problem. The traditional rainfall definition methods, using rainfall graph or a GIS map, respectively reflect the temporal or spatial variations of a rainfall process, but do not regard a rainfall as one complete process including its temporal and spatial dimension. In this paper, we present an approach to define typical modes of rainfall from the temporal and spatial dimensions. Firstly, independent rainfall processes are divided based on the continuous monitoring data of multiple rainfall stations. Subsequently, algorithms are applied to identify the typical spatiotemporal modes of rainfall and reconstruction of the process of modes, including dimensionality reduction, clustering, and reconstruction. This approach is used to analyze the monitoring data (5 min intervals) from 2004 to 2016 of 14 rainfall stations in Beijing. The results show that there are three modes of rainstorms in the Beijing urban area, which account for 31.8%, 13.7%, and 54.6% of the total processes. Rainstorm of mode 1 moves from the northwest to the center of Beijing, then spreads to the eastern part of the urban area; rainstorm of mode 2 occurs in the southwestern region of the urban area, and gradually northward, but there is no rainfall in the mountainous northwest; rainstorm of mode 3 is concentrated in the central, eastern, and southern regions. The approach and results of this study can be applied to rainstorm forecasting or flood prevention.

Keywords: rainstorm mode; high dimension; dimension reduction; cluster

1. Introduction

The security of water resources in this changing environment has become a research focus, due to the fact that climate change causes variations in rainfall at a large scale, and human activities influence the spatiotemporal characteristics at the regional scale [1,2]. Big cities are especially concerned as they are the hub of human activities. Human activities lead to the frequent occurrence of extreme rainstorms, through the urban heat island effect and air pollution [3,4]. Statistics show that 60% of the cities in China suffered from waterlogging from 2014 to 2016 [2]. Studies showed that urban waterlogging is directly related to the temporal and spatial distribution of rainstorms [5]. For example, the flood peak of triangle rainfall with a rain-peak in the central or rear is 30% larger than that of even rainfall, irrespective of whether the average rainfall is the same [6]. It is of great significance to study the spatial and temporal modes of rainfall to prevent waterlogging [7].

At present, there are two main approaches to the definition of urban rainfall process. First, different rainfall types are applied to describe various rainfall processes, such as single peak and double peaks, the method based on site monitoring data. Studies of this method focus on monitoring data of a single station or the average of multiple stations. Pilgrim and Cordery [8] put the time of a rain peak at the most likely position, and the proportion of the rain peak in the total rainfall is the average of the proportion of the rainfall peaks in each field. Keifer and Chu [9] designed the Chicago-mode according



Citation: Liu, W.; Chen, S.; Tian, F. Spatiotemporal Modes of Short Time Rainstorms Based on High-Dimensional Data: A Case Study of the Urban Area of Beijing, China. *Water* **2021**, *13*, 3597. https:// doi.org/10.3390/w13243597

Academic Editors: Yaohuan Huang, Yesen Liu, Runhe Shi and Hongyan Ren

Received: 9 November 2021 Accepted: 5 December 2021 Published: 14 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). to the strength, diachronic, and frequency of rainfall. Huff [10] designed four modes of rainfall, according to the location of the rain peak and duration of rainfall in Illinois, USA. These methods only consider the total rainfall and extreme rainfall of a single station [11]. However, the data of a single station cannot reflect the spatial characteristics of rainfall, especially in metropolitan areas with obvious spatiotemporal variations of the background environment, such as temperature and wind direction [12]. This method is widely used in urban planning or urban construction, but it is increasingly criticized for neglecting the spatial variations of rainfall, especially in large cities. The second method defines rainfall from the spatial perspective based on geographic theory. The general method is to use a spatial interpolation algorithm to interpolate the monitoring data from stations into spatial distribution data, such as the software ANUSPLIN, developed by Hutchinson, of the Australian National University [13]. In recent years, with the advancement of satellite and radar technology, rainfall spatial data can be obtained more directly, with higher accuracy, for example with Global Precipitation Climatology Project (GPCP) and Tropical Rainfall Measuring Mission (TRMM). These spatial distribution data are more suitable for analyzing the distribution of total rainfall or cumulative rainfall in a certain period, but it is difficult to express the correlation between rainfall distributions in different periods. However, in the real rainfall process, the rain belt usually moves rapidly. The spatial distribution of total rainfall or the time history distribution of rainfall intensity at a single point are not enough to accurately describe the dynamic temporal and spatial distribution characteristics of a rainfall, which is very important for the risk emergency management of rainstorm.

For the purpose of reducing the urban storm disaster effectively, it is necessary to express a complete "rainfall process", which includes not only the rainfall graph of stations at different locations, but also the spatial relationship of rainfall of these stations at different period. So, rainfall is a multidimensional data that includes time and spatial characteristics. Some scholars try to integrate the temporal and spatial characteristics of rainfall with multidimensional data. For example, based on the 3-dimensional mosaic reflectivity data from 10 S-band Doppler radars in Guangdong province, an artificial intelligence (AI) algorithm for automatic hail detection and nowcasting is developed in the light of the machine learning (ML) technology [14]. The temporal and spatial distribution characteristics of short duration rainfall in Shenzhen are analyzed and extracted by LLE algorithm [15]. It is necessary to conduct further study on the rainfall process with temporal and spatial dimensions.

In contrast to taking the rainfall data of a single station as the research object, this paper used the ML algorithm to extract the temporal and spatial distribution characteristics of rainfall from the rainfall data of all rainfall stations in the whole research area. The continuous monitoring data of rainfall stations from 2004 to 2016 were divided into different rainfall processes in Beijing. The rainfall processes were taken as the research objects, then, algorithms, such as dimensionality reduction, clustering, and reconstruction were applied to identify the typical spatiotemporal process of rainfall, and then simulate the rainfall process of different modes.

2. Study Area and Data Process

According to previous studies, Beijing has become one of the most urbanized cities in China in the past 30 years [16,17]. In 2013, urban population accounted for 86.3% of the total population in Beijing, far higher than the average of 53.73% for China, causing a rapid expansion of built-up areas [17]. In recent years, severe waterlogging has frequently occurred in Beijing [18]. For example, floods and waterlogging caused serious casualties on 21 July 2012 [19].

The Beijing urban area is 396 km². Fourteen meteorological monitoring stations have continuously recorded the rainfall data of the Beijing urban area in recent years. The continuous monitoring data of 14 rainfall stations were selected from the database, which contains data from 2004 to 2016, at intervals of 5 min. The 14 rainfall stations were evenly distributed in urban areas, as shown in Figure 1.



Figure 1. Study area.

Statistics show that 89 rainstorms occurred during 2004 and 2016, which is defined as rainfall of 1 hour exceeding 30 mm [20]. For the comparability of rainfall processes, it was necessary to standardize the rainfalls with different duration. According to previous studies, the method of 1-h moving average was used to deal with the rainfall processes, as shown in Figure 2. The data in the red box were selected as the standardized rainfall.



Figure 2. Rainfall of 1-h moving average.

Figure 2 shows the data of a rainstorm process at 14 stations. In Figure 2, "cumulative rainfall" represents the sum of rainfall at 14 stations at a certain period. For example, "527" represents the sum of rainfall at all 14 stations from the beginning of rainfall to 1 h. The range surrounded by red dotted/dashed outlines indicates that the "cumulative rainfall" was the largest in this hour compared with any other hour.

According to the early warning standard issued by Beijing Flood Control Office, when the rainfall exceeds 50 mm in one-hour, yellow warning signals will be issued to the public, indicating that serious urban disasters may occur. In 2004–2016, 22 rainfall events had a maximum sliding rainfall of more than 50 mm in one hour, accounting for 24.7% of the total rainfalls. In this paper, the 22 rainstorms were selected as samples, which are called "short duration storm". The rainfall data of these 22 rainstorms are shown in Table 1.

| | Instantaneous Rainfall (mm) Every 5 min | | | | | | | | | | | | |
|-----------------|---|------|------|------|------|------|------|------|------|------|------|------|-------|
| Date | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | Total |
| 2006-6-27 23:00 | 4.2 | 7.4 | 22.2 | 8.2 | 8.6 | 8.6 | 10.3 | 10.8 | 12.1 | 11.1 | 7.7 | 5 | 116.2 |
| 2006-6-30 22:45 | 3.8 | 3.3 | 3.8 | 4 | 3.4 | 3 | 7 | 19.7 | 18.9 | 18.4 | 5.4 | 8.9 | 99.6 |
| 2006-7-12 4:25 | 3 | 3 | 3 | 3 | 3.5 | 4.1 | 13.6 | 19.3 | 14.2 | 22.2 | 20.6 | 13.1 | 122.6 |
| 2006-7-12 19:05 | 4.5 | 4.5 | 7.3 | 20.3 | 24.2 | 16.3 | 11 | 13.2 | 14.8 | 11.6 | 9.6 | 8.6 | 145.9 |
| 2006-7-13 22:50 | 3.6 | 3.3 | 3.3 | 4.2 | 9.9 | 11.5 | 16 | 11.4 | 7.3 | 7.8 | 6.3 | 4.3 | 88.9 |
| 2006-7-31 8:50 | 10.6 | 7.1 | 3.3 | 9.7 | 12.7 | 16.7 | 20.7 | 28.7 | 18.2 | 17.2 | 15.7 | 11.7 | 172.3 |
| 2007-6-27 13:05 | 8.4 | 12.4 | 16.4 | 15.9 | 23.9 | 28.9 | 24.9 | 20.4 | 18.4 | 13.9 | 11.1 | 9.6 | 204.2 |
| 2008-6-23 14:55 | 4.8 | 3.5 | 5.8 | 4 | 9.2 | 11.7 | 16.3 | 9.2 | 10.3 | 7.7 | 10.7 | 8.5 | 101.7 |
| 2008-7-18 9:10 | 8.4 | 6.8 | 8.8 | 7.8 | 4.8 | 6.8 | 5.8 | 2.8 | 4.8 | 3.8 | 3.8 | 7.1 | 71.5 |
| 2009-7-22 17:15 | 6.5 | 11.5 | 16 | 16.2 | 8.5 | 20.4 | 24.9 | 19.9 | 17.9 | 12.9 | 5 | 6.5 | 166.2 |
| 2009-7-23 15:50 | 2 | 1 | 13.5 | 18.7 | 16.5 | 28 | 23 | 19.4 | 16.6 | 4.6 | 3.6 | 4.6 | 151.5 |
| 2009-8-7 16:20 | 2.1 | 3.4 | 2.7 | 6.6 | 9.3 | 8.6 | 9 | 9.6 | 6 | 3.5 | 5 | 2.2 | 68 |
| 2011-7-1 7:05 | 4 | 5.3 | 4.2 | 8.6 | 4.8 | 12.2 | 16.5 | 19.1 | 19.9 | 19.1 | 19 | 14.5 | 147.2 |
| 2013-7-31 19:20 | 5.1 | 0.6 | 5.1 | 1.1 | 13.6 | 0.1 | 5.6 | 0.6 | 3.6 | 1.1 | 11.1 | 7.8 | 55.4 |
| 2014-6-10 14:05 | 4.9 | 6.9 | 4.6 | 7.5 | 16.5 | 26.9 | 52.5 | 18.8 | 39.3 | 25.2 | 16.5 | 21.2 | 240.8 |
| 2014-6-15 18:55 | 8.5 | 2.3 | 6.2 | 9.6 | 14.2 | 13.1 | 8.7 | 8.4 | 8.6 | 9.1 | 11.2 | 7.3 | 107.2 |
| 2014-7-16 18:55 | 16.3 | 8.8 | 22 | 6.3 | 5.8 | 8.5 | 7.3 | 15.1 | 15.6 | 25 | 26.9 | 20 | 177.6 |
| 2014-8-23 22:30 | 0.3 | 1.8 | 7.9 | 14.3 | 16.8 | 12.7 | 18.3 | 22.8 | 12.3 | 7.8 | 8.4 | 4.8 | 128.2 |
| 2014-8-30 21:50 | 4.9 | 9.4 | 7.9 | 14.6 | 7.7 | 13.4 | 8.4 | 10.9 | 12.8 | 7.7 | 11.2 | 13.9 | 122.8 |
| 2015-8-23 14:10 | 2.5 | 4.5 | 5.7 | 9.5 | 15.6 | 7.4 | 3.7 | 2 | 5.3 | 9.9 | 4.2 | 5.1 | 75.4 |
| 2016-8-6 22:05 | 7.5 | 7 | 11 | 7 | 5.5 | 5.1 | 6 | 7.1 | 5.5 | 5.5 | 6.2 | 17.3 | 90.7 |
| 2016-9-7 18:25 | 5 | 3.7 | 1.3 | 21.3 | 26 | 32.5 | 38.1 | 22.8 | 6.8 | 18.3 | 9.5 | 5.3 | 190.6 |

Table 1. Data for 22 rainstorms.

Data source: Beijing Municipal Land and Water Protection Station.

3. Methodology

3.1. Flowchart for Extraction of Rainfall Temporal and Spatial Modes

Rainstorm is described according to the duration, intensity, total amount, and frequency by most researchers. In this paper, a new method is introduced to describe a rainstorm. The spatiotemporal process of a rainstorm was constructed as a high-dimensional array, and then principal component analysis (PCA), dynamic clustering (k-means), and reconstruction were applied to analyze the array [21]. The flowchart is shown in Figure 3.

In Figure 3, "m" is the number of samples, which represents the number of rainfalls in the manuscript. "n" is the dimension of the samples which represents the dimension of the rainfall matrix. "k" represents the dimension of rainfall samples after dimensionality reduction.

- (1) The rainstorm events were digitized and structured. High-dimensional arrays were established from temporal and spatial dimension perspectives.
- (2) Principal component analysis was used to map high-dimensional array to low latitude array.
- (3) Dynamic clustering was used to categorize samples to typical modes for describing the temporal and spatial distribution of rainstorms.

(4)

of each rainstorm modes.

- rain-2 rain-1 rain-3 rain-x [m×n] [m×n] [m×n] [m×n] PCA ᡟ rain-2 rain-1 rain-3 rain-x [m×k] [m×k] [m×k] [m×k] K-means] mode-2 mode-1 mode-y [m×k] [m×k] m×k PCA PCA PCA mode-1 mode-2 mode-y [m×n] [m×n] [m×n]

3.2. Construction of High-Dimensional Array for Rainstorms

Figure 3. Flowchart for extraction of rainfall temporal and spatial modes.

Rainstorm was characterized as a spatiotemporal process. The continuous monitoring data was divided into independent rainfall periods, where the duration of no rainfall was longer than 120 min [20]. When the maximum rainfall in 1 h was greater than 30 mm, it was called a rainstorm.

With the inverse calculation of principal component analysis, the low dimensional

array was reduced to a high-dimensional array to express the spatiotemporal process

We take the monitoring data from multiple stations in a rainstorm as a multi-dimensional array (shown in Figure 2). Based on the processing method introduced by "2. Study Area and Data process" in this paper, the multi-dimensional array was converted to matrix with same rows and columns, the number of which is 14×12 in this paper. Fourteen is the number of stations and twelve is the number of periods. The matrix forms one sample in the database of rainstorms (Ω). This is shown in Equation (1).

$$\Omega = \{Q_1, Q_2, \dots Q_m\}, Q_i = \begin{bmatrix} r_{1t1} & r_{2t1} \cdots & r_{st1} \\ r_{1t2} & r_{2t2} \cdots & r_{st2} \\ \vdots & \vdots & \vdots \\ r_{1tn} & r_{2tn} \cdots & r_{stn} \end{bmatrix}$$
(1)

where Q_i represents rainstorm i, m is the number of rainstorms, r_{stn} is the rainfall at tnperiod of s rainfall station, s = 1,2,3... S, tn = 1,2,3... N, S is the number of stations, and N is the number of periods. The main objective of this paper is to study the "mode" of rainfall, or may be called "structure", so the monitoring rainfall data is standardized by Equation (2).

$$x_{jtn} = \frac{r_{stn}}{\sum_{j=1}^{s} r_{jtn}} \times 100$$
⁽²⁾



where x_{jtn} represents the ratio of *j* station rainfall to all station at *tn* period, *S* is the number of stations.

Figure 4 is a sample of visual representation of the standardized rainstorm matrix.

| Station | | | | | | Tin | ne | | | | | | |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------------------|
| Code | t1 | t2 | t3 | t4 | t5 | t6 | t7 | t8 | t9 | t10 | t11 | t | |
| No.1 | 2.9 | 3.4 | 5.2 | 3.3 | 2.2 | 1.9 | 1.7 | 1.3 | 1.9 | 3.1 | 2.0 | 1.9 | |
| No.2 | 0.9 | 3.3 | 4.3 | 1.4 | 1.8 | 1.4 | 2.8 | 5.6 | 6.2 | 7.3 | 10.0 | 5.5 | |
| No.3 | 0.3 | 0.3 | 1.1 | 0 | 0 | 0.1 | 0.1 | 1.1 | 2.8 | 4.0 | 3.7 | 3.4 | |
| No.4 | 0 | 0 | 0.3 | 0.2 | 2.4 | 1.8 | 1.2 | 1.7 | 0.4 | 1.4 | 3.8 | 8.2 | |
| No.5 | 0 | 0.1 | 0.2 | 2.7 | 4.9 | 3.3 | 3.5 | 3.6 | 13.9 | 9.2 | 6.5 | 8.3 | |
| No.6 | 0 | 0 | 0 | 1.7 | 9.1 | 3.1 | 5.7 | 5.2 | 2.4 | 9.1 | 5.7 | 8.4 | |
| No.7 | 3.6 | 5.3 | 8.4 | 5.6 | 5.8 | 5.4 | 3.9 | 2.4 | 3.1 | 2.3 | 3.8 | 4.5 | |
| No.8 | 4.0 | 4.7 | 4.0 | 5.1 | 3.1 | 1.2 | 1.5 | 2.3 | 0.8 | 1.4 | 7.6 | 8.2 | |
| No.9 | 2.6 | 2.0 | 1.8 | 0 | 5.5 | 6.6 | 7.0 | 8.7 | 7.0 | 11.1 | 11.1 | 14.5 | |
| No.10 | 3.8 | 7.0 | 6.7 | 7.4 | 7.0 | 20.5 | 4.3 | 11.1 | 8.3 | 10.5 | 7.9 | 6.7 | |
| No.11 | 6.7 | 7.8 | 13.2 | 12.0 | 9.9 | 17.2 | 28.3 | 27.0 | 15.9 | 15.6 | 6.8 | 6.7 | |
| No.12 | 10.7 | 15.4 | 13.9 | 22.4 | 19.1 | 18.0 | 15.9 | 12.1 | 17.7 | 14.6 | 14.5 | 10.8 | Legend |
| No.13 | 19.3 | 15.1 | 11.3 | 6.9 | 5.7 | 10.5 | 7.1 | 3.2 | 3 | 1.7 | 2 | 4.3 | 2.9 Rainfall (%) |
| No | 45.0 | 35.1 | 28.9 | 30.8 | 22.6 | 8.3 | 16.3 | 13.9 | 3.4 | 8.0 | 3.0 | 4.3 | |
| | | | | | | | | | | | | | |

Figure 4. Sample of rainstorm matrix.

3.3. Dimensionality Reduction of High Dimensional Array

As shown in Figure 4, the Q_i is a $S \times N$ dimensional array. It was necessary to map the high-dimensional array to low dimensional space, in order to cluster and analyze arrays in Ω [22]. Principal component analysis (PCA) was used to reduce the dimensions of high-dimensional arrays [23].

Taking U represents the arrays in Ω , which is an n×m matrix, where n is the number of rainstorms, and m represents the number of characteristics, which equal S × N in this paper, including the monitoring data of each station at all periods. $Y_{n \times k}$ is the matrix converted from $X_{n \times m}$ by the transformation matrix $V_{m \times k}$, which means that the original data is converted from m dimensions to k dimensions (k \ll m). The steps are as follows:

- (1) The transformation matrix $X_{n \times m}$ was obtained by centralization of the matrix U,
- (2) Calculation of covariance matrix $\sigma_{m \times m}$ of $X_{n \times m}$,
- (3) Calculating the eigenvalues and eigenvectors of the variance matrix of M, M = $\sigma_{m \times m}$,
- (4) The dimension *k*, which can retain more than 90% information of original data, was obtained. K eigenvectors constitute the transformation matrix $V_{m \times k}$ as column vectors,
- (5) Descending dimensions by Equation (3).

γ

$$X_{n \times k} = X_{n \times m} \times V_{m \times k}$$
 (3)

 $Y_{n \times k}$ is the low dimensional matrix after transformation, which *n* is the number of samples and *k* is the number of dimensions of new matrix.

3.4. Clustering and Feature Selection

After the dimensionality reduction of high-dimensional samples, the k-means clustering algorithm is used to classify low dimensional samples of $Y_{n \times k}$ [24].

(1) *r* initial cluster centers are set up: $Z_1(p), Z_2(p), \ldots, Z_r(p)$, where *p* is the number of iterations.

- (2) Calculating the distance from samples $x \ (x \in X)$ to each cluster center, if $D_x(j) = min\{D_x(i)\}\ i = 1, 2, \dots r$, then $x \in S_j$, where S_j represents cluster j with the center of Z_j .
- (3) The new center of each cluster is calculated. The new center of Z_j is calculated by Equation (4).

$$Z_j(p+1) = \frac{1}{N} \sum_{i=1}^N x_i, j = 1, 2, \cdots r$$
(4)

where *N* is the number of samples contained in the cluster S_j , and x_i is the sample in S_j . Using $Z_j(p + 1)$ as the new cluster center, and the clustering criterion function can be minimized (Equation (5)).

$$J_j = \left[\sum_{x \in S_j(k)} x - z_j(k+1)^2\right]^{\frac{1}{2}}$$
(5)

where $j = 1, 2, \cdots K$.

(4) If $Z_j(p+1) \neq Z_j(p)$, j = 1, 2, ..., r, then go to step (2); if $Z_j(p+1) = Z_j(p)$, j = 1, 2, ..., r, the calculation is over.

In this paper, different r values were calculated, and the initial values of different cluster centers were selected for the k-means cluster. Finally, the rainstorms were divided into three modes. The mean value of each rainstorm was taken as a typical mode of the rainstorm.

3.5. Reconstruction of Rainstorms

With the inverse calculation of principal component analysis, the low dimensional array was reduced to a high-dimensional array to express the spatiotemporal process of each rainstorm. The i clustering centers are reconstructed into $i \times m$ matrices (Equation (6)).

$$X_{app} = Z_{i \times k} \times V_{m \times k}' \tag{6}$$

where *i* is the number of rainstorm modes and *m* is the dimension of the original data.

4. Results

The method of extracting rainfall temporal and spatial modes was applied to analyze the monitoring data of 22 rainstorms in the urban area of Beijing. It was found that the spatiotemporal distribution can be divided into three modes, and there are obvious differences in the center, spatial distribution, and occurrence time of three modes. The movement process of rainfall centers in different modes at different periods is shown in Figure 5. The centroid coordinates for a certain period was obtained by the method of calculating the geographical center with the weight of rainfall.

As shown in Figure 5, there are obvious differences in the three modes. The geographical centers of mode 1, mode 2, and mode 3 are located in northwest, southwest, and southeast, respectively. Mode 1 moves from the northwest to the urban center, mode 2 mainly spreads from the southwest and south to the north and the urban center, and mode 3 is basically concentrated in the urban center.

The matrixes of 3 modes are shown in Figure 6, which shows the characteristics of rainfall distribution. For example, the rainstorms belonging to mode 2 are more centralized than that belonging to modes and mode 3, which shows rainfall concentrated at several stations of 30748000, 30523900, 30504030, and 30523650. The rainfall spatial distribution at different periods of each mode is shown in Figures 7–9, which represents the percentage of rainfall at all rainfall stations in the period by the depth of the color in the location of the station.

As shown in Figure 6, the proportion of rainfall at each station is significantly different. In mode 1, rainfall is mainly concentrated in six stations, which are stations 3, 6, and 4. The spatial non-uniformity of rainfall in mode 2 is the most obvious, and only four stations account for a large proportion of rainfall, which are stations 5, 9, 12, and 4. In mode 3, almost all stations have obvious rainfall, except station 1 and station 2.

In mode 1, as shown in Figures 5–7, rainstorm spreads from the northwest mountainous area to the central area of the city and the eastern part of the city. Rainfall began at the beginning of the northwest mountain area, and the rest of the city did not have any rainfall. Rainfall gradually dispersed and rainfall occurred at all stations. There are seven rainstorms of this mode, accounting for 31.8% of all sample. Among them, the single peak and homogeneity stations are 43%, and the double-peak type accounts for 14%.



Figure 5. Centroids of three modes at different periods.



Figure 6. Matrixes of 3 modes. (a) Mode 1; (b) mode 2; (c) mode 3.



Figure 7. Spatiotemporal process of mode 1.

In mode 2, as shown in Figures 5, 6 and 8, the main rainfall was concentrated in the southern and southwestern regions of the city, gradually spreading to the northern and urban central areas, and there was no rainfall in the northwest mountain areas. There are three rainstorms of this mode, accounting for 13.7% of the total sample. Two of them are unimodal and one is homogeneous. In addition, this type of rainfall occurs between 14:00 and 16:00.



Figure 8. Spatiotemporal process of mode 2.

In mode 3, as shown in Figures 5, 6, and 9, the main rainfall was concentrated in the central area of the city and the eastern and southern parts of the city, which basically did not move. There was no rain in the northwest mountain areas. There were 12 rainstorms of this mode, accounting for 54.6% of total sample, among them, 62% were single peak type and 38% were homogeneity type. This mode was the main rainstorm type in summer in Beijing, which mainly occurred from afternoon to evening.



Figure 9. Spatiotemporal process of mode 3.

5. Discussion

How to identify and extract valuable information from multidimensional and massive rainfall monitoring data is a problem faced by many researchers. In this paper, a new approach for rainfall mode recognition is introduced, and different rainfall modes are identified from the massive monitoring data through the algorithms of dimensionality reduction, clustering, and reconstruction of a high dimensional array. This approach can be applied to both multidimensional data analysis and spatiotemporal data mining.

The results show that there are three modes of rainstorms in the Beijing urban area. Rainstorms of mode 1 moved from the northwest to the center of Beijing, then spread to the eastern part of the urban area; rainstorms of mode 2 occurred in the southwestern region of the urban area, and gradually northward, but there was no rainfall in the mountainous northwest; rainstorms of mode 3 were concentrated in the central and eastern regions, and basically did not move. The results are consistent with the actual rainstorm process. This approach provides a framework for analysis, and there is uncertainty in some respects. For example, the result of the restructured rainfall has certain randomness and uncertainty in spatiotemporal distribution, which should be paid attention to in future. Because of the abundance of available data, this paper only selected the rainfall data of 14 years in Beijing. It should be noted that urban rainfall is a part of a larger range of rainfall in most cases, the temporal and spatial distribution characteristics of rainfall extracted from the rainfall data of 14 stations may have a certain randomness and uncertainty. More extensive rainfall data should be collected, considering terrain, climate, etc., according to the availability of data. In addition, it is necessary to distinguish the historical evolution of rainfall modes in different periods due to the climate changes in the city.

There are several suggestions for practice, First, the spatial and temporal resolution for rainfall data needs to reflect the temporal and spatial differences of a rainfall. Secondly, the study area should include the urban area and surrounding areas as much as possible, which is mainly to maintain the integrity of rainfall process.

6. Conclusions

In this paper, a high dimensional array is introduced for the study of the spatiotemporal distribution of rainfall, which describes rainfall by storing continuous rainfall monitoring data of all rainfall stations.

Through the establishment of high dimensional arrays of each rainstorm and algorithms, such as dimensionality reduction, clustering, feature extraction, and reconstruction, the spatiotemporal distribution of rainstorms in the flood season of Beijing city was analyzed. It was found that there were three spatiotemporal modes of rainstorms in the urban area of Beijing from 2004 to 2016. Rainstorms of mode 1 moved from the northwest mountain area to the central district, and further spread to the eastern part of the area. Rainstorms of mode 2 concentrated in the southwest of the urban area, gradually spreading to the northern and urban central areas; the northwest mountainous area basically had no rain. Rainstorms of mode 3 concentrated in the central area of the urban area and the eastern and southern regions, and basically did not move. The variation of the centroids of different modes shows a significant difference between the modes. The approach and conclusions in this paper can be applied to the study of rainfall modes in other cities or regions at a different scale, so as to provide assistance for rainfall forecasting and flood prevention.

The limitation of current machine algorithms is that it is too dependent on the number and quality of learning samples. If the rainfall stations are dense and the rainfall data quality is accurate, this method can achieve good results. If it is in an area with insufficient data and sparse rainfall stations, results may not be satisfactory. With the increasing density of rainfall stations, the improvement of rainfall data quality, and improvement of machine learning algorithm, the machine learning algorithm will get more reasonable and objective results.

Author Contributions: The study was designed by W.L. and F.T. The data were collected and analyzed by W.L. The paper was written by S.C. and F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Willems, P.; Arnbjerg-Nielsen, K.; Olsson, J.; Nguyen, V. Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmos. Res.* **2012**, *103*, 106–118. [CrossRef]
- 2. Xia, J.; Wei, S. Perspective on water security issue of changing environment in China. J. Hydraul. Eng. 2016, 47, 292–301.
- 3. Kong, F.; Shi, P.J.; Fang, J.; Lu, L.; Fang, J.; Guo, L. Advances and Prospects of Spatiotemporal Pattern Variation of Extreme Precipitation and its Affecting Factors under the Background of Global Climate Change. J. Catastrophology **2017**, *32*, 165.
- 4. Wang, B.; Zhao, L.; Gong, Y. Characteristics of temporal pattern and return period of short-duration rainfall at Beijing Observatory. *Torrential Rain Disasters* **2015**, *35*, 302–308.
- Rafieeinasab, A.; Norouzi, A.; Kim, S.; Habibi, H.; Nazari, B.; Seo, D.-J.; Lee, H.; Cosgrove, B.; Cui, Z. Toward high-resolution flash flood prediction in large urban areas—Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. *J. Hydrol.* 2015, 531, 370–388. [CrossRef]
- 6. Zhangchun, W.; Guoping, C.; Zhimin, A. Experimental study on overland flow. J. Hydraul. Eng. 1995, 26, 84–89.
- Guan, M.; Sillanpää, N.; Koivusalo, H. Storm runoff response to rainfall pattern, magnitude and urbanization in a developing urban catchment. *Hydrol. Process.* 2016, 30, 543–557. [CrossRef]
- 8. Guoping, C.; Jin, S.; Rongsheng, F. Research on Rainfall Pattern of Urban Design Storm. Adv. Water Sci. 1998, 9, 41–46.
- 9. Keifer, C.J.; Chu, H.H. Synthetic Storm Pattern for Drainage Design. J. Hydraul. Div. 1957, 83, 1332-1–1332-25. [CrossRef]
- 10. Kibler, D.F. Urban Stormwater Hydrology; American Geophysical Union: Washington, DC, USA, 1982.
- 11. Song, X.; Zhang, J.; Zou, X.; Zhang, C.; AghaKouchak, A.; Kong, F. Changes in precipitation extremes in the Beijing metropolitan area during 1960–2012. *Atmos. Res.* 2019, 222, 134–153. [CrossRef]
- 12. Wei, Z.; Changhe, L.U. Assessing Changes in Rainstorms in Beijing During the Last 50 Years. J. Resour. Ecol. 2016, 7, 372–377. [CrossRef]
- 13. Hutchinson, M.F. *ANUSPLIN Version 4.3*; Centre for Resource and Environment Studies, The Australian National University: Canberra, Australia, 2004.
- 14. Wenhai, Z.; Lei, L. An initial application of artificial intelligence on the detection and nowcasting of hail weather. *ACTA Meteorol. Sin.* **2019**, *77*, 282–291.
- 15. Liu, Y.Y.; Li, L.; Liu, Y.S.; Chan, P.W.; Zhang, W.H. Dynamic spatial-temporal precipitation distribution models for short-duration rainstorms in Shenzhen, China based on machine learning. *Atmos. Res.* **2020**, 237, 104861. [CrossRef]
- 16. Zhang, Y.; Li, Q.; Huang, H.; Wu, W.; Du, X.; Wang, H. The Combined Use of Remote Sensing and Social Sensing Data in Fine-Grained Urban Land Use Mapping: A Case Study in Beijing, China. *Remote. Sens.* **2017**, *9*, 865. [CrossRef]
- 17. Liu, X.; Tian, G.; Feng, J.; Ma, B.; Wang, J.; Kong, L. Modeling the Warming Impact of Urban Land Expansion on Hot Weather Using the Weather Research and Forecasting Model: A Case Study of Beijing, China. *Adv. Atmos. Sci.* **2018**, *35*, 723–736. [CrossRef]
- 18. Zhicong, Y.; Guo, W.; Li, N.; Xie, Y. Numerical Simulation of Urban Ponding in Beijing. Meteorol. Mon. 2015, 41, 1111–1118.
- Gu, X.; Li, N.; Zhou, Y.; Wu, J. Thinking on urban waterlogging disaster defense initiated by "7 21" extraordinary rainstorm in Beijing. J. Nat. Disasters 2013, 22, 1–6.
- 20. Weidong, L.; You, H.; Ren, G.; Yang, P.; Zhang, B. Subtle Precipitation Characteristics in Beijing Area. *Clim. Environ. Res.* 2014, 19, 61–68.
- Zhang, T.; Detang, L.; Daolun, L.; Yi, D. Research on Statistical Information Reconstruction of Images Based on Multiple-Point Geostatistics Integrating Soft Data with Hard Data. J. Comput. Res. Dev. 2010, 47, 573–578.
- 22. Jolliffe, I.T. Principal Components Analysis; Wiley Online Library: New York, NY, USA, 2002.
- 23. Mitchell, T. Machine Learning; China Machine Press: Beijing, China, 2003; pp. 417–433.
- 24. Ng, R.T.; Han, J. Efficient and Effective Clustering Methods for Spatial Data Mining. In Proceedings of the VLDB Conference, Santiago, Chile, 12–15 September 1994.