Rainfall Data Reconstruction Based on Chaotic Characteristics of Meteorological Factors

Junchen Li¹*, Ke Kong¹, Chenfeng Cui² and Zhitao Zhang²

¹School of Water Conservancy and Environment, Jinan University, Jinan, China
²College of water resources and Architectural Engineering, Northwest A&F University, Yangling, China

*Corresponding author e-mail: 201607231@ujn.edu.cn

Abstract. Rainfall data is a basic hydrological data, which is indispensable for the calculation of many hydrometeorological conditions. However, the acquisition of long-sequence rainfall data often has the problem of missing data. Based on the chaotic characteristics of rainfall and temperature data changes, the paper introduces temperature data with strong correlation with rainfall data into the model based on similar phase space theory for reconstruction of rainfall data. Through experiments in the Loess Plateau, the results show that the similarity between the calculated results and the real data reaches 95.17%. Compared with the traditional method, the accuracy of data reconstruction is improved.

Keywords: Chaos Theory, Refactoring Research, Rainfall Data, Temperature Data

1. Introduction

Rainfall is an important component of regional water circulation and water balance. High-quality rainfall data is a necessary prerequisite for analyzing and calculating various phenomena in the fields of hydrology, meteorology, and agriculture. In reality, many of the original rainfall data are missing data and cannot meet the application needs. Therefore, improving the integrity of rainfall data is an important direction of current hydrological research [1].

Features based on rainfall data, Chaos theory is considered a superior high-precision method to supplement data integrity[2]. Huang Sheng and Liang Chuan used the theory of chaotic phase space to analyze and process the 20 years of rainfall data in Yanting Station from 1985 to 2004. The similarity error between the reconstructed result and the real data is between 18% and 20%[3]. Wang Weiguang used chaotic phase space theory to chaotically process rainfall data from 1961 to 2000 in Wuding River at different time scales. The similarity error after data reconstruction is 8%~15% [4]. Existing research results have shown that the rainfall data has good chaotic characteristics.

Although the reliability of the chaotic method is becoming more and more mature, but if the rainfall data has a long lack of measurement period, the accuracy of chaotic analysis will be reduced [5], which will affect the reliability of hydrological research. The paper uses the characteristics of strong correlation between temperature series and rainfall series, and reconstructs rainfall data based
on similar phase space theory to analyze and reconstruct rainfall data, which solves the problem of missing rainfall data and ensures the accuracy of rainfall data reconstruction [6].

2. Theoretical Analysis
As rainfall, temperature and other water science processes are affected by many factors, they have caused huge spatiotemporal variability, showing non-random but seemingly random characteristics, which makes it difficult for traditional deterministic mathematical models to simulate these water science processes. And the emergence of chaos theory provides new ideas for studying this highly complex system [7~8]. There is a strong correlation between temperature and rainfall data, it is assumed that their chaotic characteristics also have a strong correlation [9].

2.1. Delay Time and Autocorrelation Detection
According to the chaos theory, the decomposition coefficient is the value obtained by dividing the annual data in a year by the monthly data in each month, and its physical meaning is the distribution proportion of the annual data in the year [10]. Using the decomposition coefficient of the data, the delay time $T$ necessary for the chaotic operation can be obtained. The decomposition coefficient can better reflect the changing characteristics of the data. The data sequence can be reconstructed through the inverse operation [11].

Reasonable selection of delay can optimize the separation of adjacent orbits in the minimum phase space. If the selected delay time is too small, the effective information between different vectors is insufficient, which will cause the final correlation dimension to be seriously underestimated. But if the selected delay time is too large and the system still has chaotic characteristics. It is impossible to achieve the optimal separation of the orbits, so that the effective information is lost, the space-time average value in the space is invalid, and the correlation dimension is overestimated. The autocorrelation function method is used to determine the delay time, and the best delay time is determined by the first pass of $0$ or $\sqrt{e}$. The calculation formula is:

$$r_{\tau} = \frac{\sum_{T=t+1}^{N} (A_{T-t} - \bar{A})(A_{T} - \overline{A})}{\sum_{T=1}^{N} (A_{T} - \overline{A})^2}$$

From the calculation result of the decomposition coefficient autocorrelation function, it can be seen that the value of the rainfall data delay time is $\tau=2$, and the value of the temperature data delay time is $\tau=2.1$, and the duration is basically consistent with the average period of the orbit.

After comparing and analyzing the delay time of the two sets of data, the delay result is substituted into the G-P equation and the chaotic characteristics of the two sets of data are continuously analyzed and compared using the saturated correlation dimension method.

2.2. Compare the Chaotic Characteristics of the Two Sets of Decomposition Coefficients by Using the Saturation Correlation Dimension Method
After phase space reconstruction, the delay time of the decomposition coefficient is obtained, which can be estimated by the correlation dimension

\[ c(r) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} H \left( r - \| y_i - y_j \| \right) \]

Where \( H(x) \) is the Heaviside function (Heaviside), and has:

\[ H(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases} \]

Select a different scale \( r \) to get the relationship diagram between \( \ln(C(r)) \) and \( \ln(r) \). If there is a scale-free area, that is a straight line segment, it indicates that the time series distribution has fractal characteristics, and the slope of the straight line segment is the correlation dimension (attractor dimension) of the time series [12]. Figure 4 shows the relationship between \( \ln(r) \) and \( \ln(C(r)) \) under different embedding space dimensions \( m \) (\( m = 3, 4, ..., 11 \) in this study).

Choose a different scale \( r \), Get the relation chart between \( \ln(C(r)) \) and \( \ln(r) \), if there is a scale-free zone, which is, a straight line segment, it indicates that the time series distribution has fractal characteristics, and the slope of the straight line segment is the associated dimension of the time series (the attractor dimension)[12]. Figure 4 The relationship between \( \ln(r) \) and \( \ln(C(r)) \) under different embedding space dimensions \( m \) (\( m = 3, 4, ..., 11 \) in this study).

Figure 3 Rainfall \( \ln(r) \) and \( \ln(C(r)) \) relation chart

Figure 4 Relationship between fractal dimension (d) and embedded phase space (m)

When the embedding phase space dimension \( m = 11 \), the correlation dimension reaches a saturation value \( d = 1.32 \) (also called the saturated correlation dimension), so take the minimum embedding dimension \( m = 11 \), which represents the number of effective degrees of freedom of the dynamic system. In other words, when the embedded phase space of the precipitation time series reaches 11 dimensions, the system has a stable attractor dimension of 1.32.

Figure 5 Temperature \( \ln(r) \) and \( \ln(C(r)) \) relation chart

Figure 6 Relationship between fractal dimension (d) and embedded phase space (m)

The minimum embedding dimension \( m = 12 \) and the saturation correlation dimension \( d = 0.98 \) of the temperature series are also calculated. In other words, when the embedded phase space of the
temperature time series reaches 12 dimensions, the system has a stable attractor dimension of 0.98. This shows that the temperature change sequence in the area of the experiment also has chaotic characteristics. By calculating, analyzing and comparing the autocorrelation function and chaotic characteristic curve of the two sets of meteorological factor data, it is fully proved that the two sets of data have very similar change characteristics and spatial orbit characteristics, which proves that the temperature data can reconstruct the rainfall data chaotically.

3. Reconstruction between Meteorological Factors

3.1. Principles and Methods of Factorization Reconstruction Based on Chaotic Attractors

For a chaotic attractor, all the moving points near its orbit have similar motion states. Therefore, when it is difficult to explore the change characteristics of the chaotic attractor itself, the motion of the approximate point function relationship in its field can be calculated, and using the motion characteristics of these similar trajectories to predict subsequent motion patterns.

Let $S(t)$ be the state point of the decomposition coefficient of the time series at $t$, then $S(t+T)$ is the decomposition coefficient at the time $(t+T)$. We take $S(t+T)$ as the predicted point of this analysis, and $S(t)$ as the known point. The core of using known points to predict unknown points is to fully understand the relationship between the two. When the chaotic characteristics of the experimental sequence are fully verified, $S(t+T)$ can be predicted by $S(t)$. And before the prediction, considering the possible impact of the non-synchronized long point on the route on its prediction, it is often an influencing factor that has a particularly large impact on the result. Based on the above assumptions, a prediction model based on the weights of steps of different similarities is proposed:

$$S(t+T) = \frac{\sum_{i=1}^{k} S(t_i + T)}{\sum_{i=1}^{k} \frac{R - d_i}{R - d_m}}$$

In this formula: $S(t+T)$ is the decomposition coefficient at time $(t+T)$. $S(t_i)$ is each point in the neighborhood of the predicted point $S(t)$; $k$ is the number of the neighborhood radius; $D_i$ is the spatial distance from the phase prediction point to the prediction point $S(t_i)$, $d_m$ is the shortest spatial distance from the phase prediction point to the prediction point $S(t_i)$; $R$ is the neighborhood radius;

More comprehensively, this prediction model not only considers the principle of similar phase space points between attractors, but also reflects the possible impact of the non-synchronized long point on the route on its prediction, and finally it can be concluded that the different phases in the route The position of the point has a different effect on the final sequence.

3.2. Using Temperature Data to Reconstruct the Missing Precipitation Data

It is proved that comparing the two sets of data has obvious chaotic characteristics. Before we use monthly temperature data to reconstruct the blank monthly rainfall data based on the weighted average method, we first macroscopically compare the correlation between the two sets of temperature and precipitation data, and we draw the relationship Figure to determine the correlation between the two.

![Figure 7 Precipitation temperature data correlation diagram](image_url)
The above figure is the temperature data corrected by monthly precipitation data and correction coefficient. It can be found that the two meteorological factor data have an obvious correlation, so this provides a considerable possibility for reconstructing and analyzing the precipitation data using temperature data. Subsequent experiments will test and fit the correlation between the two sets of data decomposition coefficients. When the error accuracy is below the acceptable range, it becomes possible to use temperature data for precipitation reconstruction analysis.

**Figure 8** Relationship between uncorrected and corrected two sets of decomposition coefficients

The two sets of decomposition coefficients have a certain correlation without any pretreatment, but the degree of fit is low. And its total error and average error both reached 56.5%. This kind of data cannot replace the precipitation data decomposition coefficient for reconstruction analysis due to its poor correlation. Therefore, before reconstructing the precipitation data, it is necessary to correct the decomposition coefficient of the temperature data before proceeding to the next step.

**Table 1** Decomposition coefficient data relationship table

| Time | 1.00 | 2.00 | 3.00 | 4.00 | 5.00 | 6.00 | 7.00 | 8.00 | 9.00 | 10.00 | 11.00 | 12.00 |
|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|
| Rainfall | 100.2 | 37.83 | 72.75 | 7.51 | 13.37 | 6.43 | 6.48 | 13.04 | 3.79 | 10.81 | 126.1 | 97.15 |
| Temperature | 75.21 | 88.56 | 43.86 | 22.35 | 16.49 | 13.07 | 12.62 | 13.97 | 17.88 | 15.68 | 135.4 | 42.99 |

The correction coefficient was revised several times and the correction coefficient $K=2.3$ was finally determined, and the above decomposition coefficient relationship diagram and data relationship table were obtained. After the correction, the total error and average error of the two decomposition coefficients are reduced to 0.52%. Using the aforementioned chaotic analysis results of decomposition coefficients, the weighting method is used for calculation, and the revised temperature decomposition coefficient sequence is used to predict the corresponding annual precipitation decomposition coefficient sequence.
We reconstructed the annual precipitation data from 2010 to 2013 using the revised weighted predicted temperature decomposition coefficient. The weighted average relative error of the monthly precipitation from 2010 to 2013 obtained by this model is 9.49%, 7.96%, 3.21% and 3.32% respectively.

4. Conclusions
The paper proves that the variation of rainfall sequence in the Loess Plateau is chaotic and predictable. In the experimental area, it is assumed that the rainfall data is missing. Considering the obvious correlation between the two meteorological factors, the two hydrological and meteorological factor sequences are eliminated. After the zero-value interference, the corrected decomposition coefficient of the temperature series is used to reconstruct the blank period of the rainfall data, and the fitting error on the average value fluctuates within 5%. The experimental results fully show that when the rainfall sequence data is missing in the Loess Plateau, it is completely feasible to use the temperature sequence to reconstruct the data based on chaos theory.

Acknowledgments
This work was financially supported by Farmland water-saving and emission-reducing salt control technology and its application fund.

References
[1] CAO Xiao lei, ZHOU Zu hao, WANG Zi ru, Effects of precipitation down scaling method in evapotranspiration simulation of land surface model, [J], South-to-North water Transfers and Water Science & Technology 2015, 13(6): 1031–1039. (in Chinese)
[2] YANG Wei hua, SUN Wen ye, XU Zhi xin, et al, Research on Problem of Scale in Hydrology, [J], Pearl River, 2016, 37(5): 51–53. (in Chinese)
[3] LIU Ting ting, ZHANG Li wu, WANG Xi qin, Analysis of Research Progress of Precipitation Evolution Characteristics, [J], Technology Innovation and Application, 2013, 27(4): 126. (in Chinese)
[4] Aristita Busuioc, Deliang Chen, Cecilia Hellström. Performance of statistical down scaling models in GCM validation andregional climate change estimates. application for Swedish precipitation. [J], International Journal of Climatology, 2002, 17(6): 12–13.
[5] WANG Zhong ya, WEN Yu hua, DONG Jia gen, Research on the Seasonal Period Classification of Taihu Basin Floods Based on Fractal Theory, [J], China Rural Water and Hydropower, 2015, 28(2): 118–122. (in Chinese)
[6] PANG Shu jiang, WANG Xiao yan, MA Wen jing, Research on Parameter Uncertainty for the HSPF Model Under Different Temporal Scales, [J], environmental Science, 2018, 39(5): 2030–2038. (in Chinese)
[7] YAN Bai zhong, XIAO Chang lai, QIAO Yu, et al, Multiple Timescale Chaos Identification of Groundwater Depth and Subdivision of Jilin City Based on 0-1 Test,[J],Transactions of The Chinese Society of Agricultural Machinery,2015,41(1):138–147. (in Chinese)

[8] WEI Run chu, XIAO Chang lai, ZHANG Yu qing, Chaos Identification of Precipitation Time Series and Subdivision Based on the 0-1 Test, Journal of Northeastern University (Natural Science), 2014, 35(12):1792–1796. (in Chinese)

[9] Sun N, Hong B, Hall M Assessment of the SWMM model uncertainties within the generalized likelihood uncertainty estimation (GLUE) framework for a high-resolution urban wellshed[J]. Hydrological Processes, 2014, 28 (6): 3018-3034.

[10] Wen X, Hu D, Cao B, et al. Dynamics change of honghu lake's water surface area and its driving force analysis based on remote sensing technique and TOPMODEL model[J]IOP Conference Series: Earth and Environmental Science, 2014, 17 (1) :012130.

[11] Worku T, Khare D, Tripathi S K. Modeling runoff-sediment sponse to land use/land cover changes using integrated GIS and SWAT model in the Beressa watershed[J], Environmental EarthSciences,2017,76 (16): 550.