Environment, Resource and Ecology Journal (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/erej.2022.060104
ISSN 2616-3756 Vol. 6 Num. 1

Research on Extreme Precipitation Analysis Model Based on Robust Mahalanobis Distance and Nonlinear Fitting

Runmo Wang[1], Zonghong Han[1], Shiqin Li[2], Jing Chen[3]

[1]*School of Electrical Engineering and Automation, Qilu University of Technology, Jinan, Shandong, 250353, China*

[2]*School of Information Science and Engineering, Harbin Institute of Technology (Weihai), Weihai, Shandong, 264209, China*

[3]*School of Mathematics and Statistics, Qilu University of Technology, Jinan, Shandong, 250353, China*

Keywords: robust Mahalanobis distance, nonlinear least squares method, precipitation analysis

Abstract: This paper presents a quantitative and qualitative analysis of extreme precipitation. First, it is used to filter abnormal precallofil values using robust Mahalanobis distance. At the same time, consider the influence of related variables on the annual precipitation. It has been high in precipitation in 2003, 2016 and 2021, and higher precipitation is 1964, 1983 and 1992. Using nonlinear least squares method to fit the correlation between the annual precipitation and other climate indicators, the PRCP has an unstable oscillation in July, and DEWP and TEMP have an oscillation rise trend, and SLP is attenuated.

1. Introduction

With the rise in global temperatures and the massive emissions of carbon dioxide, global precipitation has increased. Many parts of the country have suffered from heavy precipitation and its rarity, which poses a great threat to people's life and safety. Therefore, it is important to develop predictive models for cities with different potential extreme precipitation events and mathematical models for the quantitative analysis of their losses. [1]

2. Outlier Analysis Based on Robust Mahalanobis Distance

The Mahalanobis distance [2] represents the covariance distance between data, which can effectively represent the similarity of unknown samples, and the Mahalanobis distance is an important way to detect anomalies for data.

Step 1: Set up a matrix $X_{n\times p}$ with n rows and p columns, and colraw h samples in the matrix.

Step 2: Determine the value h of the sample data, the less the sample data h, the stronger its resistance to outliers, too few h values will result in the inability to distinguish between outliers, generally by default.

$$h = 0.75 \times n$$

If the sample is small, then take:

$$h = 0.9 \times n$$

Step 3: Randomly select p+1 samples from n samples to form the covariance matrix, calculate its determinant, if its value is zero, then randomly add a sample until the value of the determinant is not zero, at this time the covariance matrix is the initial covariance matrix S_0, using random samples to calculate the initial sample mean T_0

If n is small, calculate it using T_0, S_0 and start iterating to obtain S_3, repeating the process to obtain multiple S_3, choosing the smallest 10 and iterating again until it starts to converge. n is large, divide it into smaller parts and iterate separately. Finally return the value of the smallest determinant of T_{MCD} and S_{MCD}.

Calculate the Mahalanobis distance $d(i)$ of the sample from T_{MCD}, S_{MCD}

$$d(i) = \sqrt{(x_i - T)' \times S^{-1} \times (x_i - T)}$$

If $d(i) > \sqrt{\gamma^2}_{P,0.975}$, write W as 0, otherwise 1. Recalculate from W.

$$\begin{cases} T = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i \\ S = \sum_{i=1}^{n} w_i (x_i - T)(x_i - T)' / \left(\sum_{i=1}^{n} w_i - 1 \right) \end{cases}$$

At this point S is the stable covariance matrix, on which the robust Mahalanobis distance is obtained.

Firstly, the data visualization of the annual precipitation from the meteorological stations in Zhengzhou City was carried out, and for each year the annual precipitation was subtracted from the average precipitation for 56 years to obtain a precipitation visualization of the difference Figure 1.

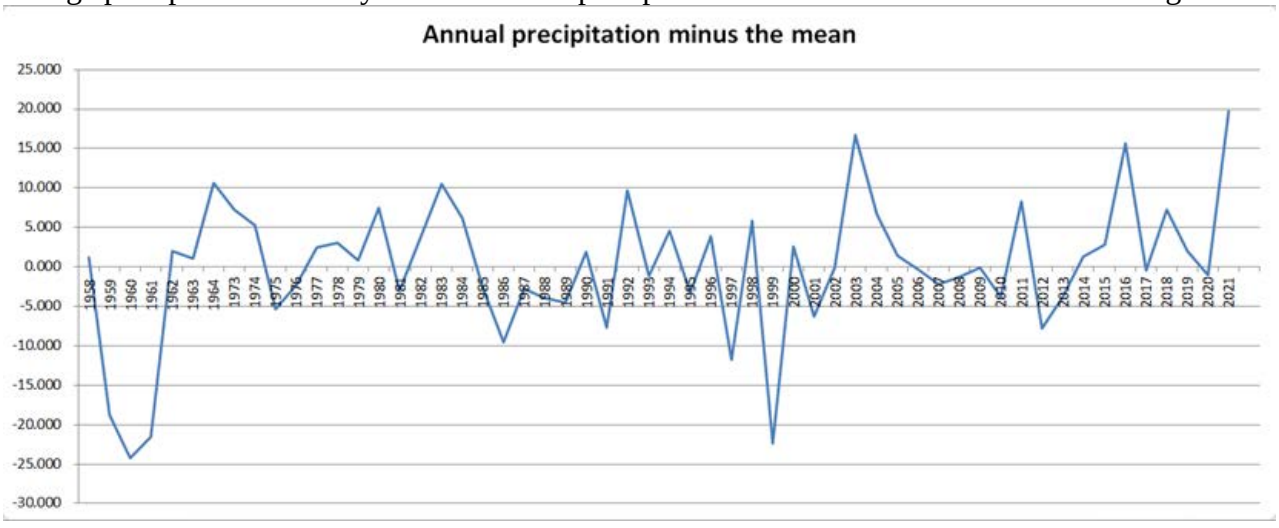

Figure 1: Visualization of precipitation-mean precipitation values

It is clear from Figure 1 that there are precipitation peaks as well as precipitation valleys from 1958 to 2021, and that as time goes on, precipitation peaks occur more frequently and precipitation valleys occur less frequently, with precipitation basically remaining greater after 2002, which is related to global warming and the large amount of carbon dioxide emissions.

Using the Mahalanobis distance to find years with high precipitation, and adjusting up the threshold of the Mahalanobis distance to find outliers, 11 outliers can be obtained using MATLAB, with points of high and low precipitation, which are divided into four categories according to the degree of outlier anomalies. It is possible to obtain extremely high and much higher than average precipitation in 2003, 2016 and 2021, with higher precipitation in 1964, 1983 and 1992.

3. Precipitation Model Based on Nonlinear Least Squares Fitting

Step 1: Through the basic processing of the data can be derived from the function trend can be selected from different functions. The fitting function is as follows:

$$f(x) = f(x, a_1, a_2, \ldots, a_n)$$

Step 2:For a set of data (x_i, y_i) (i = 1,2, … , n) find a $y = f(x)$ such that the curve fits best to all data points under certain conditions, that is, the curve fits well, and let δ_i be the residual of the curve at x_i

$$\delta_i = f(x_i) - y_i = 1,2, \ldots, n$$

Using least squares, i.e. the smallest sum of squares of deviations, as a criterion for determination.

$$J_{\min} = \min(f(x_i) - y_i)^2$$

The determined $f(x)$ from Step1 is brought into (2), thus changing the problem to one of minimising a non-linear function.

Step 3: Using non-linear optimization to find the parameter $a_1, a_2, \ldots, a_n$.

Seven meteorological indicators, PRCP, DEWP, SLP, TEMP, VISIB, WDSP and MXSPD, were analysed and processed to create a one-month time series for July in Zhengzhou city, where heavy rainfall persisted in July.

The PRCP data in Zhengzhou during July were analysed and the PRCP as a function of time was obtained by fitting the precipitation using Sum of Sine.

$$\begin{aligned} y_2 = {} & 1.803\sin(0.1337t - 0.8368) + 0.9408\sin(0.7108t + 0.082) + 0.6986\sin(0.407t - 0.322) \\ & + 0.4107\sin(1.571t - 3.634) + 0.4895\sin(1.027t - 6.033) + 0.4568\sin(2.129t - 2.952) \\ & + 0.5411\sin(1.31t - 5.71) + 0.3944\sin(2.756t - 3.437) \end{aligned}$$

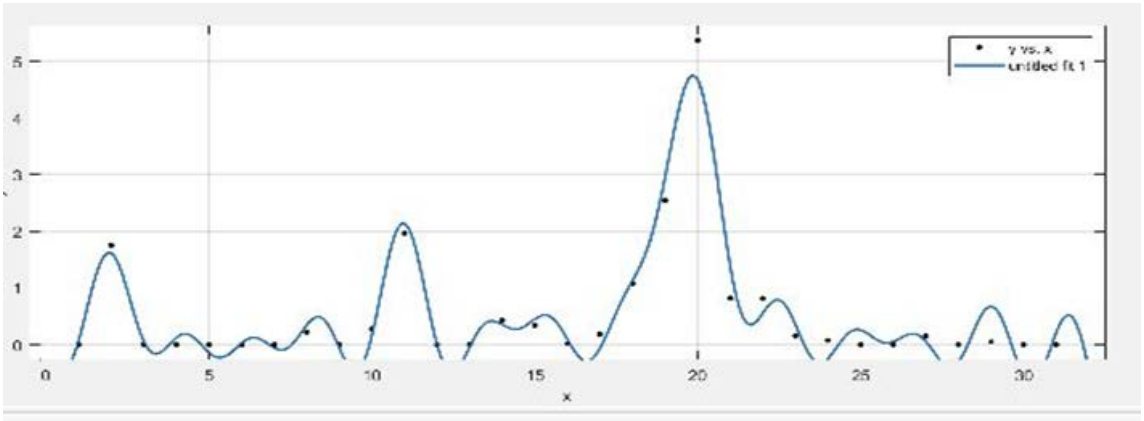

Figure 2: PRCP fit results

According to Figure 2 it can be seen that the Sum of Sine function fits well for July precipitation with an R-square of 0.937.

Similarly for y1 (DEWP), y3 (SLP), y4 (TEMP), y5 (VISIB), y6 (WDSP) and y7 (MXSPD) the fit analysis was carried out.

$$y_1 = 0.7635t^4 + 0.9008t^3 - 3.572t^2 - 0.007124t + 73.43$$

$$y_3 = -2.4t^7 + 0.7359t^6 + 14.15t^5 - 3.608t^4 - 27.24t^3 + 3.082t^2 + 15.12t + 1003$$

$$y_4 = 1.36e^{-7}t^8 - 1.61e^{-5}t^7 + 0.0008t^6 - 0.02t^5 + 0.2795t^4 - 2.226t^3 + 9.368t^2 - 16.93t + 86.51$$

$$y_5 = -2.438t^9 + 2.294t^8 + 15.5t^7 - 12.73t^6 - 35.59t^5 + 21.5t^4 + 36.4t^3 - 9.929t^2 - 14.76t + 8.024$$

$$y_6 = 5.834\sin(0.0024t + 1.014) + 0.6852\sin(0.3t + 2.466) + 0.7659\sin(0.6873t - 4.945) + 0.639\sin(0.8213t - 2.742) + 0.5119\sin(2.193t + 2.89) + 0.4662\sin(1.602t + 1.78)0.4632\sin(2.497t + 2.721)$$

$$y_7 = 11.39\sin(0.049t + 1.248) + 3.877\sin(0.09t + 4.482) + 0.252\sin(0.54t - 0.88) + 0.88\sin(2.71t - 1.484) + 1.136\sin(0.645t + 1.436) + 0.904\sin(2.493t - 2.912) + 1.087\sin(2.149t - 2.993) + 0.712\sin(1.691t - 0.60)$$

The average variance of the fitting effect using non-linear least squares is around 0.85, which is a good fit. The fitting function can better represent the precipitation situation in Zhengzhou in July, and it can be seen from the fitting function that the PRCP in July shows an unstable oscillation, DEWP and TEMP show an oscillating upward trend, and SLP shows a decaying oscillating trend.

4. Conclusion

This paper analyzes the annual variation of Zhengzhou precipitation characteristics, screening a lot of precipitation, and quantitative analysis of Zhengzhou flood event in 2021.

References

[1] Maining, Ren Zhihua, Wang Wei, Liu Na, Cao Ning. Comparative analysis of parallel observation of national precipitation weather [J / OL]. Study on arid zone: 1-11 [2022-01-05]. Http: //Kns.cnki.net/kcms/detail/65.1095.x. 20211224.1422.002.html.

[2] Bidush Ranjan Swar et al. Genetic Diversity Studies in MAGIC Population of Soybean (Glycine max (L.) Merrill) Based on Mahalanobis D2 Distance [J]. International Journal of Plant & Soil Science, 2021: 18-25.