

# Risk Evaluation and Prediction of Rainfall Disaster in Flood Season Based on Pearson Fitness Test and Risk Function

**Binbin Li**

*Changjiang Survey, Planning, Design and Research Co., Ltd, Wuhan 430010, China  
reben\_1987@163.com*

**ABSTRACT.** Heavy rainfall during flood season causes great losses for the cities and the farmland. In order to better describe the flood season rainfall disaster events, the risk function combining risk probability and risk loss is put forward, and the risk degree value is obtained by combining the corresponding risk loss. The probability distribution of precipitation in flood season is found out through Pearson fitness test, and the probability of precipitation in each flood season is calculated using the hydrological frequency analysis method. In this paper, taking Tongzhou district of Beijing city as an example, the above method was used for risk analysis, and the result was close to the reality. The research results have deepened the understanding of the changing law of precipitation risk in Tongzhou district and have certain reference value for urban disaster prevention decision-making.

**KEYWORDS:** risk function, risk evaluation, rainfall disaster, Pearson Fitness Test

## 1. Introduction

The heavy rainfall disaster mainly refers to the fact that the local super-standard rainfall cannot be discharged in time due to the low drainage capacity of the river channel, insufficient drainage power or being supported by the river, caused the surface water of the farmland exceeding the tolerance depth and time of crops to waterlogging, and lead to in crop yield reduction or non-harvest [1].

Although natural disaster events such as heavy rainfall are inevitable, the disaster loss can be mitigated by preventing and improving the capacity for resisting natural disasters. The effective mitigation of the loss caused by natural disasters depends on people's scientific understanding of the risk of natural disasters.

Based on the risk analysis for the disaster caused by heavy rainfall, first, water conservancy department can scientific determine disaster prevention standards and plan the construction of disaster prevention projects. Second, it can provide the technical guidance and services to control flood, reduce disaster and avoid risk for

urban construction, land planning, each industry production and operation, and people's life. Third, it can be also beneficial to make scientific decision-making for the work of flood control, avoid the blindness of the command decision-making, improve the scientific, validity, accuracy and efficiency of the command decision-making, and effectively reduce the flood disaster loss. Last, it can also be used as the basis for insurance amount of inundation and drought insurance for approval.

The study of heavy rainfall risk has great social benefits, which are directly reflected in disaster reduction benefits, and indirectly reflected in economic benefits and some social benefits. For a city or town with flood control tasks, it plays an important role in keeping one side safe, stabilizing the society and promoting the smooth development of various industries. Its huge economic and social value can hardly be reflected in figures.

Risk research mainly focuses on four aspects: (1) engineering reliability analysis; (2) disaster risk research, including flood risk research, drought risk research; (3) environmental risk analysis of water activities; (4) water supply system risk analysis. Besides, based on risk analysis and decision-making technology combined with computer technology, decision support system is developed to analyze the reliability of water supply system, the risk of water matter activity pollution, and the engineering reliability [2-5].

The concept of the risk degree based on risk probability and risk loss is put forward in this paper, which can be taken as an index to measure a city's disaster resistance. The purpose of risk analysis is to make better decisions, and the degree of risk also provides a reference for the government to make decisions on disaster prevention programs.

## 2. Methodology

### 2.1 Pearson Fitness Test

Pearson Fitness test aims at selecting the appropriate distribution curve by statistical test. It is a test of the suitability of theoretical distribution (frequency curve) and statistical distribution (empirical distribution) [6]. Its general formulation on rainfall is as follow.

Suppose that the total precipitation can be divided into  $K$  disjoint intervals according to the value of precipitation, and the precipitation with a certain value randomly appearing belongs to the  $i_{th}$  interval is denoted as event  $A_i (i=1,2,\dots,K)$ , the corresponding probability is  $P(A_i)=P_i$ .

For an observation value with a capacity of  $n$  precipitation, if  $n_i (\sum_{i=1}^K n_i = n)$  of which falls into the  $i_{th}$  interval, the frequency of the event  $A_i$  is  $n_i/n$ . Under the assumption  $H_0$ , according to the natural ratio test method, when  $n$  is large, the divergence is calculated as follows :

$$\bar{U} = -2 \ln \lambda = \sum_{i=1}^K \frac{(n_i - P_i^0)^2}{nP_i^0} \quad (1)$$

Where  $\lambda$  is the natural ratio,  $n_i$  is the actual frequency of the event  $A_i$ ,  $nP_i^0$  is the theoretical frequency of events  $A_i$ .

At the level  $\alpha$ , the critical boundary of the test is

$$\bar{U} \geq \chi_{\alpha}^2(K - r - 1) \quad (2)$$

When calculated  $\bar{U}$  by formula (1) satisfies equation (2), reject  $H_0$ , otherwise accept  $H_0$ .

Calculate the probability when  $\bar{U} \geq u$ , namely  $P\{\bar{U} \geq u\} = P$ .

The p-value can be called the fitness between the obtained data and the null hypothesis. The greater the p-value, the stronger the evidence for the null hypothesis. Given a significance level  $\alpha$ , when p-value is less than  $\alpha$ , the null hypothesis is rejected, and it is considered that the theoretical distribution and statistical distribution have a large degree of divergence and a low degree of fitness, and the probability that the theory conforms to the actual event is small. Compared various distribution curves, final choice will be made.

R software is statistical software, but also a mathematical computing environment, which provides a flexible and interactive environment to analyze, visualize and display data. The above functions can be achieved by programming with R software. The general steps are as follows [7]:

- (1) Enter data first, using scan () function;
- (2) Group data, calculate frequency of each group, and call cut () function and table () function to group and count;
- (3) Calculate the theoretical probability values of the original hypothesis (hypothesis distribution) in each cell.
- (4) Pearson chi-square test was performed. Call the chisq.test () function.

## 2.2 Risk Function

The basic meaning of risk is the uncertainty of loss. However, the definition of risk generally includes loss and uncertainty, where uncertainty includes known and unknown probability distributions of loss events. Some definitions emphasize uncertainty, some emphasize loss, and some emphasize both concepts [8]. The understanding of risk is: if a human activity may face loss, and the possibility is uncertain, then the activity is called risk activity, the loss that faces is called risk loss, the probability of loss is called risk probability, and risk is the utility function of taking risk loss and risk probability as independent variables. The specific form of

the function needs to be further determined. The places with high risk probability and high loss are considered high risk areas, the places with high risk probability and low loss are considered low risk areas, the places with low risk probability and low loss are considered high risk areas, and the places with low risk probability and low loss are considered small risk areas. Based on this understanding, a more reasonable definition of risk degree in various natural states is proposed.

If the risk degree is  $r$ , the risk loss is  $l$ , and the risk probability is  $p$ , then

$$r = f(x, l) = ap \cdot e^{cl+d}, \text{ when } 0.002 < p < 1 \quad (3)$$

$$r = f(x, l) = b \cdot e^{cl+d}, \text{ when } 0 < p < 0.002 \quad (4)$$

Where  $a$ ,  $b$ ,  $c$ ,  $d$  are the undetermined parameter (such as  $a=1$ ,  $b=0.002$ ,  $c=1$ ,  $d=0$ ),  $0 < p < 1$ ,  $0 \leq l \leq L_{max}$ ,  $L_{max}$  is the maximum risk loss.

### 3. Pearson Chi-square Test for Precipitation Distribution in Flood Season

#### 3.1 Study Area

Tongzhou district is located in the southeastern suburbs of Beijing, the north end of the Beijing-Hangzhou Grand Canal, the east end of Beijing Bailichang Street. It belongs to the alluvial plain of Yongding and Chaobai rivers and is a part of the North China Great Plain. Tongzhou area belongs to the temperate continental monsoon climate, the summer (June to August) is hot and rainy, rainfall concentration, intensity, the average precipitation is between 480 mm and 530 mm, accounting for 80% to 85% of the annual precipitation, heavy rain process occurs in the annual July 15 to August 15, which is the key period of flood control and drainage. When the precipitation exceeds 600 ~ 800 mm, "bud waterlogging" and "summer waterlogging" are prone to occur.

#### 3.2 Pearson Chi-square Test

The assumption of logarithm normal distribution, Weibull distribution and Pearson III type distribution are respectively put forward, and R software is used for fitting inspection for each hypothesis distribution, the P value result are respectively: P - value < 2.2 e-16, P - value = 0.2108 and P - value = 0.7461. It is observed that Weibull distribution and Pearson III type distribution of P values are greater than 0.05, the precipitation in this area can be considered obeying the two distributed[9].

Besides, the P value of the null hypothesis for Pearson type III distribution is close to 1, and the Pearson type III distribution has better improved significantly than the Weibull distribution. And Pearson type III distribution is chosen as flood season precipitation distribution of Tongzhou district.

**4. Risk degree calculation and risk analysis**

There are only 22 years due to the heavy rainfall caused disasters loss in Tongzhou district, and the probability value according to the 22 years precipitation with Pearson type III distribution is calculated, such as table 1.

*Table 1 The calculation probability of precipitation from 1949 to 1994*

| year | Precipitation (mm) | damage area (10 <sup>4</sup> mu) | probability | year | Precipitation (mm) | damage area (10 <sup>4</sup> mu) | probability |
|------|--------------------|----------------------------------|-------------|------|--------------------|----------------------------------|-------------|
| 1949 | 814.9              | 67.56                            | 0.0749      | 1963 | 420.7              | 11.48                            | 0.6779      |
| 1950 | 834.2              | 88.43                            | 0.0648      | 1964 | 600                | 13.85                            | 0.3054      |
| 1951 | 331.2              | 27.93                            | 0.8554      | 1966 | 481.7              | 4                                | 0.5415      |
| 1953 | 460.5              | 16.31                            | 0.5889      | 1967 | 497.4              | 2                                | 0.5069      |
| 1954 | 967.4              | 76.76                            | 0.0224      | 1969 | 741                | 49.55                            | 0.127       |
| 1955 | 873.2              | 83.5                             | 0.048       | 1973 | 498.5              | 5                                | 0.5045      |
| 1956 | 908.6              | 58.6                             | 0.0363      | 1977 | 436.6              | 7.61                             | 0.6425      |
| 1958 | 711.6              | 19                               | 0.1548      | 1978 | 640.1              | 5.2                              | 0.2426      |
| 1959 | 1024.8             | 68.28                            | 0.0137      | 1979 | 645                | 14                               | 0.2356      |
| 1960 | 346.5              | 1.34                             | 0.829       | 1988 | 571.1              | 2.5                              | 0.3567      |
| 1962 | 527                | 24.8                             | 0.4437      | 1994 | 698.8              | 5.06                             | 0.1684      |

In order to calculate further, the disaster area in the table needs to be standardized. Formula (5) is used for normalization. The calculated results are listed in table 2.

$$L_k = \frac{l_k - \min_{\forall i} \{l_i\}}{\max_{\forall i} \{l_i\} - \min_{\forall i} \{l_i\}} \quad k = 1, 2, \dots, 22 \quad (5)$$

*Table 2 The calculation results of risk degree from 1949 to 1994*

| year | Precipitation (mm) | risk loss | risk degree | year | Precipitation (mm) | risk loss | risk degree |
|------|--------------------|-----------|-------------|------|--------------------|-----------|-------------|
| 1949 | 814.9              | 0.7604    | 335.43      | 1963 | 420.7              | 0.1164    | 121.34      |
| 1950 | 834.2              | 1         | 962.06      | 1964 | 600                | 0.1436    | 62.64       |
| 1951 | 331.2              | 0.3053    | 393.71      | 1966 | 481.7              | 0.0305    | 63.08       |
| 1953 | 460.5              | 0.1719    | 139.08      | 1967 | 497.4              | 0.0076    | 52.64       |
| 1954 | 967.4              | 0.866     | 170.11      | 1969 | 741                | 0.5536    | 202.25      |
| 1955 | 873.2              | 0.9434    | 537.15      | 1973 | 498.5              | 0.042     | 62.24       |
| 1956 | 908.6              | 0.6575    | 97.14       | 1977 | 436.6              | 0.072     | 92.1        |
| 1958 | 711.6              | 0.2028    | 42.67       | 1978 | 640.1              | 0.0443    | 30.28       |
| 1959 | 1024.8             | 0.7686    | 64.12       | 1979 | 645                | 0.1454    | 48.74       |
| 1960 | 346.5              | 0         | -           | 1988 | 571.1              | 0.0133    | 38.13       |
| 1962 | 527                | 0.2694    | 170.62      | 1994 | 698.8              | 0.0427    | 20.85       |

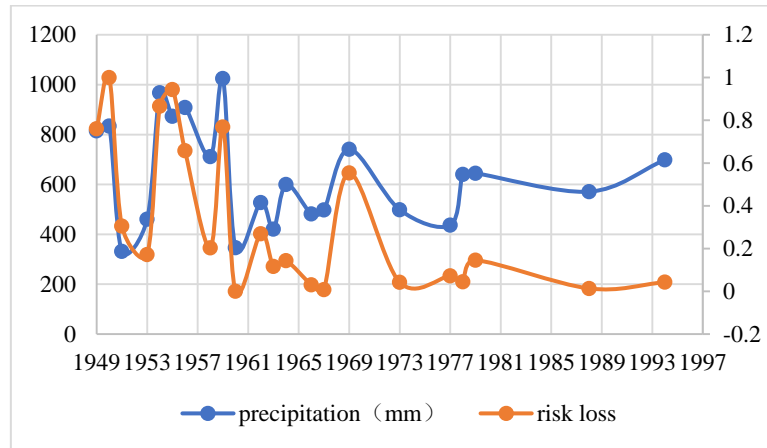


Figure 1 The relation diagram of precipitaon and risk loss from 1949 to 1994

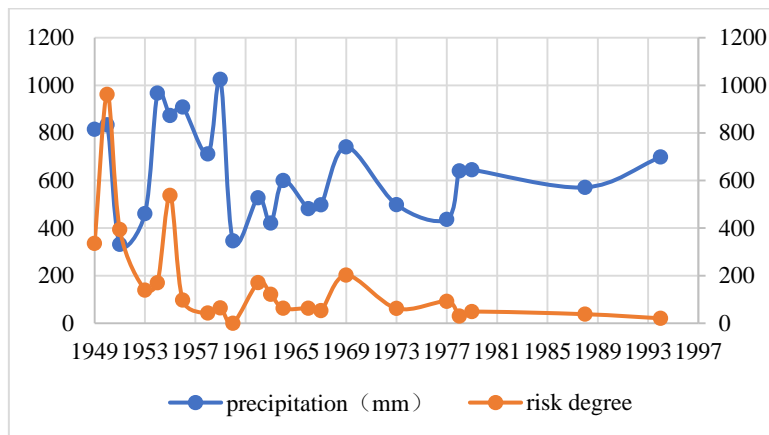


Figure 2 The relation diagram of precipitaon and risk degree from 1949 to 1994

Based on disaster prevention capacity in 1994, first, the calculated flood season precipitation with a-hundred-year return period in Tongzhou district through Pearson type III distribution is 1043.20 mm, and the corresponding risk loss according to the risk degree model is 0.62, which can be concluded that the affected area is 542400 mu.

Considering the extreme cases, the probability of one, three and five once-in-a-century rainfalls in the next 20 years can be calculated by using geometric distribution as 0.1653, 0.00096 and 0.000013. The corresponding disaster area is 0.5424 million, 1.6272 million and 2.712 million mu respectively.

## 5. Conclusion

The value of risk degree can reflect the local disaster prevention ability to some extent. Generally, the smaller the risk degree is, the stronger the disaster prevention ability is. It can be seen from the change of risk degree that Tongzhou has significantly improved its ability to prevent precipitation disasters, and the risk has remained at a very low level since 1978. During that period, a complete backbone drainage and irrigation project and field supporting project were established throughout the region, and comprehensive development and management of "farmland four-water conversion" were carried out, so that disasters were better managed. This verifies the feasibility of the risk degree model proposed in this paper to some extent.

There is still much research to be done. For example, the parameter value of the risk degree model in this study is given, but there is no complete theory to support it, and there is a little more subjectivity. Although the result is reasonable, it is not completely convincing, and further exploration is needed in theory. Besides, when using hypothesis test to select the distribution, the moment method and maximum likelihood estimation method are used to estimate the parameters. The rationality of the method is not analyzed in detail and the advantages and disadvantages of the method are not evaluated. In the future, other methods can be used to estimate the parameters and compare the different methods. In addition, it is also worth discussing how to improve the model when applying the method of this paper to risk analysis of other regions.

## References

- [1] Madhusmita Swain, Sujata Pattanayak, U.C. Mohanty. Characteristics of occurrence of heavy rainfall events over Odisha during summer monsoon season. *Dynamics of Atmospheres and Oceans*, Vol. 82 (2018), p. 107-118.
- [2] Andreas Paul Zischg, Patrick Hofer, Markus Mosimann, Veronika Röthlisberger, Jorge A. Ramirez, Margreth Keiler, Rolf Weingartner. Flood risk (d) evolution: Disentangling key drivers of flood risk change with a retro-model experiment. *Science of the Total Environment*, Vol. 639 (2018), p. 195-207.
- [3] Saeideh Kheradmand, Ousmane Seidou, Diawoye Konte, Mahaman Bohari Barmou Batoure. Evaluation of adaptation options to flood risk in a probabilistic framework. *Journal of Hydrology: Regional Studies*, Vol. 19 (2018), p. 1-16.
- [4] Hugo Carrão, Gustavo Naumann, Paulo Barbosa. Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability. *Global Environmental Change*, Vol. 39 (2016), p. 108-124.
- [5] A. Lopez-Nicolas, M. Pulido-Velazquez, H. Macian-Sorribes. Economic risk assessment of drought impacts on irrigated agriculture. *Journal of Hydrology*, Vol. 550 (2017), p. 580-589.

- [6] Xinyi Song, Fan Lu, Hao Wang, Weihua Xiao, Kui Zhu. Penalized maximum likelihood estimators for the nonstationary Pearson type 3 distribution. *Journal of Hydrology*, Vol. 567 (2018), p. 579-589.
- [7] E. González-Estrada & J. A. Villaseñor. An R package for testing goodness of fit: goft, *Journal of Statistical Computation and Simulation*, Vol. 88 (2017), 726-751.
- [8] Faisal Khan, Hangzhou Wang, Ming Yang. Application of loss functions in process economic risk assessment. *Chemical Engineering Research and Design*, Vol. 111 (2016), 371-386.
- [9] Coles, S, Pericchi, LR, Sisson, S. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, Vol. 273 (2003), p. 35-50.