Study on the Method of Solving the P-III Curve Fitted by Precipitation Frequency Based on Excel

Hongliang Liu¹, Shanhai Li²*, and Peng Qin¹

¹Shandong Geological Environmental Monitoring, 17 Jingshan Road, Jinan, China
²School of Mathematic and Quantitative Economics, Shandong University of Finance and Economics, 40 Shungeng Road, Jinan, China

* Corresponding author: lsh751@163.com

Abstract. The built-in function of Excel was used to calculate the precipitation frequency, the least square method was used to fit the appropriate line, the statistical parameters of hydrological variables were optimized, and the charting function was used to draw the precipitation frequency curve to meet the requirements of production practice.

1. Introduction
There are lots of built-in functions with strong data processing power in Excel. Using Excel built-in function calculation can replace programming calculation, which has a wide application prospect in production practice. At the same time, excel has rich drawing ability, by using its drawing function, it can draw some complex figures, such as Pearson III curve (p-III curve) which is widely used in hydrological frequency calculation and has the characteristics of convenient drawing and visual display.

2. Data samples and statistical parameters

2.1 The data sample
In this paper, annual precipitation data from 1959 to 2018 are used for calculation and verification. Firstly, the serial data obtained in continuous sequence are entered into the Excel workbook in the form of two columns of data: annual data and precipitation data. Precipitation data are ranked in descending order.

2.2 Experience in frequency
We calculate the experience frequency by

\[ P = \frac{i}{n+1} \]

Calculated using the built-in function ROW( ) / (ROWS(series)+1), ROW( ) subtracts the number of previously unnumbered ROWS when data is not entered from the first ROW.

2.3 Statistical parameter
The sample data is a series of samples in continuous order. The mean value \( \bar{x} \) is calculated by using the AVERAGE series and the mean variance \( \sigma \) is calculated by using STDEVA (series). The
dispersion coefficient is calculated by \( C_v = \frac{\sigma}{\bar{x}} \) and the coefficient of skew \( C_s \) is calculated by \( C_s = K \times C_v \). We obtain \( \bar{x} = 692.0 \text{mm} \), \( C_v = 0.269 \).

\[3. \text{Calculation of theoretical frequency precipitation}\]

\[3.1 \text{P-III type curve function}\]

From the existing data, P-III type curve match hydrological distribution of a random variable. Its probability density function is:

\[ f(x) = \frac{\beta^n}{\Gamma(\alpha)} (x-a_n)^{\alpha-1} e^{-\beta(x-a_n)} \]  

Here \( \alpha, \beta, a_0 \) are three parameters of curve shape, \( \Gamma(\alpha) \) is the gamma function for \( \alpha \). The relationship between \( \alpha, \beta, a_0 \) and \( \bar{x}, C_v, C_s \) as follows:

\[ \alpha = \frac{4}{c_i^2} = \frac{4}{K^2 C_v^2} \]  

\[ \beta = \frac{2}{x C_v C_s} = \frac{2}{x K C_s} \]  

\[ a_0 = \bar{x} \left( 1 - \frac{2 C_v}{C_s} \right) = \bar{x} \left( 1 - \frac{2}{K} \right) \]  

\[3.2 \text{Specifies the precipitation variable for probability} \ P \]

In practice, the probability distribution of the random variable is required. Generally, we need the value of \( X_P \) of the random variable corresponding to the specified probability \( P \).

\[ P = P(x > x_p) = \bar{x} \int_{x_p}^{\infty} f(x) dx = \frac{\beta^n}{\Gamma(\alpha)} \int_{x_p}^{\infty} (x-a_n)^{\alpha-1} e^{-\beta(x-a_n)} dx \]  

Let \( t = \beta(x-a_n) \), we have

\[ P = \frac{1}{\Gamma(\alpha)} \int_{t_p}^{\infty} e^{-t} dt = 1 - \frac{1}{\Gamma(\alpha)} \int_{0}^{t_p} e^{-t} dt \]  

\[ 1 - P = \frac{t_p}{\Gamma(\alpha)} \]  

Where, \( 1 - P \) is the precipitation frequency, that is, the precipitation assurance rate.

\[ t_p = \beta(x_p - a_n) \]  

(8)

In hydrological calculation, the random variable \( x_p \) can be calculated as follows:

\[ x_p = \bar{x} \left( 1 + \phi_p \right) \]  

(9)

According to formula (8), we obtain the coefficient of deviation mean

\[ \phi_p = \frac{x_p - \bar{x}}{x C_v} \]  

(10)

According to formulas (2), (3), (4), (8), we obtain

\[ \phi_p = \frac{KC_v t_p}{2} - \frac{2}{KC_v} \]  

(11)

Substitute equation (11) into equation (9), we get

\[ x_p = \sqrt{1 + C_v \left( \frac{KC_v t_p}{2} - \frac{2}{KC_v} \right)} = \sqrt{1 + \left( \frac{KC_v t_p}{2} - \frac{2}{K} \right)} \]  

(12)
The integral type in the right of (7) is the standard $\Gamma(\alpha)$ distribution function for $\beta = 1$, and $t_p$ is the standard $\Gamma$ distribution function. Using the built-in function GAMMA.INV in Excel, we can return the standard $\Gamma(\alpha)$ cumulative distribution function of inverse function, that is, quantile $t_p$, where, the three parameters are respectively: $1 - P$, $\frac{4}{K^2}$, 1. Then we have $t_p = \text{GAMMA.INV} \left( 1 - P, \frac{4}{K^2}, 1 \right)$. Substitute into (12), we can obtain $x_p$.

4. The least square method fits the right line
The characteristic of fitting method is that under certain fitting criterion, the optimal statistical parameters of frequency curve are obtained by fitting the theoretical data with the observed data. In this paper, the least square method is used to calculate the solution, that is, the minimum criterion of square sum of relative dispersion.

$$ S = \sum_{i=1}^{n} (x_i - x_p)^2. \tag{13} $$

In equation (12), $C_r$ has been obtained through statistics. The empirical frequency $P_i$ of precipitation in each year is calculated respectively. Set the initial $K_r = 2.0$, calculate the empirical frequency $P_i$ corresponding to $X_p$ according to $K_j = K_0 + 0.1 \times j$ (where $j$ is a natural number), and obtain a series of theoretical data under different $C_r/C_r$ ratios. The relative dispersion squared $(X_i - X_p)^2$ was calculated respectively, and the sum of the relative dispersion squared $S_i$ was obtained. If $S$ increase, then $K_j = K_0 - 0.1 \times j$, according to the above method, is to find the minimum relative deviation square $S_i$ to corresponding $K_j$ is for $C_r/C_r$ ratio, all hydrological parameters, P - III also determine the type of curve, can be obtained the theoretical precipitation $X_p$ corresponding to different precipitation frequency $P_i$, rainfall frequency are shown in table 1. When $K = 1.5$, the relative dispersion sum of squares is the smallest, then the ratio of $C_r/C_r$ is determined to be 1.5, and $C_r = 1.5 \times 0.269 = 0.404$.

| No. | Year | Precipitation (mm) | Cs/Cv ratio K | Sum of the relative dispersion squared S |
|-----|------|--------------------|---------------|-----------------------------------------|
| 1   | 1962 | 1147.4             | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 2   | 1964 | 1118.8             | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 3   | 2004 | 1091.0             | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 4   | 1973 | 994.4              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 5   | 1961 | 992.1              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 6   | 2003 | 987.0              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 7   | 2005 | 963.0              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 8   | 1981 | 386.0              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 9   | 1989 | 365.5              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 10  | 1986 | 347.0              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| 11  | 1968 | 314.0              | 1.4 1.5 1.6   | 1.4 1.5 1.6                            |
| Total|      |                    |               | 22222 22124 22154                       |

Serial Numbers 8~64 data are hidden.
5. Draw the precipitation frequency curve

5.1 P-Ⅲ type curve vertical and horizontal coordinates
The ordinate of P-Ⅲ type curve is evenly divided, it can be automatically generated by the Excel charts function, the corresponding values are precipitation. The abscissa of P-Ⅲ type curve is not evenly divided, and the frequency value is related to the quantile of standard normal distribution. The quantile of standard normal distribution is 0 when P = 50%, and the abscissa of P-Ⅲ type curve is 0 when P = 0.01%. Excel built-in function NORM.S.INV to return to the standard normal cumulative distribution function of inverse function value (quantile), so the P-Ⅲ type curve calculation formula can be expressed as the abscissa values:

\[ LP = \text{NORM.S.INV}(P) - \text{NORM.S.INV}(0.0001) \tag{14} \]

5.2 The vertical and horizontal coordinate grid line
The horizontal grid lines of P-Ⅲ type curve are shown in Fig. 1. We can calculate the abscissa value \( L_p \) of each grid line separately, that is, the values of \( LP \) when \( P = 0.0001, 0.0002, \ldots, 0.0009, 0.001, 0.002, \ldots, 0.009, 0.01, 0.02, \ldots, 0.99, 0.999, 0.9999 \).

The vertical coordinate grid line is automatically generated by Excel, and we can set the maximum and minimum values. The minimum value is 0 and the maximum value is determined by referring to \( X_0.0001 \), that is, the maximum value is greater than \( X_{0.0001} \), and take the whole hundred. Set the spacing of the main grid line is 100 or 200. Each \( L_p \) value corresponds to two precipitation values, 0 and Max.

5.3 Abscissa mark
In an XY scatter plot of an Excel chart, the abscissa is the \( L_p \) value and the abscissa in P-Ⅲ type curve is the frequency value \( P \). So we can set 1 data series, add data label, manually modify data label to identify abscissa. Identify the data, the abscissas \( L_p \) are the values of \( L_p \), when \( P = 0.0001, 0.001, 0.01, 0.05, 0.1, \ldots, 0.9, 0.95, 0.99, 0.999, 0.9999 \), and corresponding precipitation is 0.

5.4 Draw the precipitation frequency curve
Through the above steps, four data series can be obtained, one is the abscissa grid line \((L_p,0)\) and \((L_p,\text{Max})\) data series, the other is the abscissa identification \((L_p,0)\) data series. Empirical \( P_i \) corresponds to \( L_p \) and precipitation \( X_i \) series, \( L_p \) and theoretical precipitation \( X_p \) series. What needs to be explained is that in order to prevent the ordinate grid line data series from connecting to a line, a blank line should be inserted between different \( L_p \) data to make the connection of ordinate data discontinuous. When the precipitation frequency curve is drawn, the XY scatter diagram is inserted first, and then select the ordinate elevation data series and add the data label. Put the data label at the bottom, and the default value is Y value. Manually change it to the precipitation frequency value under the percentage. Then adjust the settings to make the figure beautiful, the precipitation frequency curve drawn is shown in Fig. 1.
6. Calculation results
The example of this paper is the series data of precipitation in Jinan City from 1951 to 2019, and the average precipitation for many years is obtained, with the guarantee rates of 20%, 50%, 75% and 95%. Using the above method, the above values are obtained, and the previous results are compared, as shown in Table 2. The results of this calculation show that the average annual precipitation increases significantly, and the statistical data of this time is significantly higher than that of the former, the results are more objective, and the annual precipitation with different guarantee rates also increases significantly compared with the previous data.

Table 2. Comparison of calculation results and previous data

| Data series   | Perennial average precipitation (mm) | Coefficient of variation Cv | Cs/Cv | Different guarantee rate annual precipitation |
|---------------|--------------------------------------|----------------------------|-------|-----------------------------------------------|
| 1956-1988*    | 637.0                                | 2.8                        | 2.0   | 780.3 619.8 508.3 375.8                      |
| 1951-2019     | 692.0                                | 0.27                       | 1.5   | 844.2  679.5  560.4  408.1                    |

* 1991, data of Jinan hydrology bureau.

7. Conclusion
By comparison, the method calculated the annual precipitation calculated by this method increased significantly under different assurance rates, which was consistent with the variation trend of annual average precipitation of objective statistical data, indicating that the calculated results were reliable and the method was feasible. Therefore, it can be seen that the built-in rich functions in Excel software are used for the calculation of precipitation frequency, and the fitting line of least square method is completely feasible. Use its chart function freely, can output and display precipitation frequency curve completely and quickly, the chart surface is beautiful. The method can be used in production practice to replace the original graphical calculation and programming calculation. After one-time calculation, only need to manually complete the least square method data fitting, and then change the sample length setting in the chart, the work that needs several days in the past can be completed in more than ten minutes, greatly improving the production efficiency.

References
[1] Huang tinglin, ma xueni. Hydrology (5th edition). Beijing: China architecture industry press. M (2014).
[2] Excel Home. Excel 2016 function and formula application. Beijing: Peking University press. M (2018).

[3] Huang zejun. Research on hydrological frequency calculation method using Excel. Journal of science and technology entrepreneurship. J. 7 (2006).

[4] Meng caixia, wang pingyi, zhang xiaowei, et al. Comparison of three fitting evaluation methods for calculating hydrological frequency. South-to-north water diversion and water conservancy technology. J. 6 (2015).

[5] Zhang kunyi. Application of VB and Excel in hydrologic frequency calculation. Jilin water conservancy. J.12 (2013).

[6] Wu mingguan, Zhong hai, Zhou qingxin, etc. Application of Excel in hydrologic frequency calculation. Hydrology. J. 5 (2001).
