CLIMATE EXERCISE 1: Intensity-Duration-Frequency Curves

Xian Haoyi

October 31, 2024

1 Question and Data

With the rainfall information provided, construct the IDF curves for the return periods of 10 and 100 years.
Upload a ZIP file with your name and containing a plot of the curves, an excel table with the results and the code that you used to construct the curves.
Data is houly-precipitation data lasting from 1998 to 2016.

2 Answer

Steps to solve the exercise:

first, plot a time series of the data to assess its quality. The analysis reveals some unreasonable values below zero. Additionally, there are substantial amounts of missing values, predominantly in 1989 and 2015, while other years have fewer missing data points.

Subsequently, perform a data aggregation operation on the original dataset to obtain data at different time steps. For this, develop an *aggregation* function and remove missing values.

In the third step, create the function *get_annual_max* to extract the annual maximum values from the aggregated data obtain in the previous step, preparing it for the Generalized Extreme Value (GEV) analysis.

Using $MATLAB$'s $gevfit$ and *gevcdf* functions, fit the annual maximum values to the GEV distribution. Then derive empirical points from 1 to 24 hours for precipitation intensities with 10-year and 100-year return periods.

Finally, apply four analytical formulas to fit the empirical points using IDF function. Analyze the fitting results of these analytical formulas, compare and discuss the performance of each formula, and reflected on the challenges encountered and insights gained during this study.

2.1 Data Pre-Processing

Plot the time series and review the quality of raw data. As shown in Fig.1, there are some negative values. Besides, there are many NAN values in 1998 and 2015, remove these data in following step.

Because the time interval of the original data is 1 hour, data aggregation is needed to obtain data for other time steps.In this paper, use the duration D from 1 to 24 hours.
Firstly, generate the corresponding time series according to D, and then use a loop to take out data slices from the original data, take the average of the slices to get the precipitation data with new time intervals.
This step can also use the sliding average function *movmean* in $MATLAB$ to achieve unrepeated sliding averages for different window lengths (D).

To perform a frequency analysis, need a series with the maximum rainfall intensity in each of the years. Since the final goal is to derive an IDF curve, repeat this process for several storm durations. The function get_annual_max accomplishes this goal.

2.2 Fit GEV Distribution

Use the GEV distribution. When applied to exclusively positive values such as precipitation, the GEV distribution is:

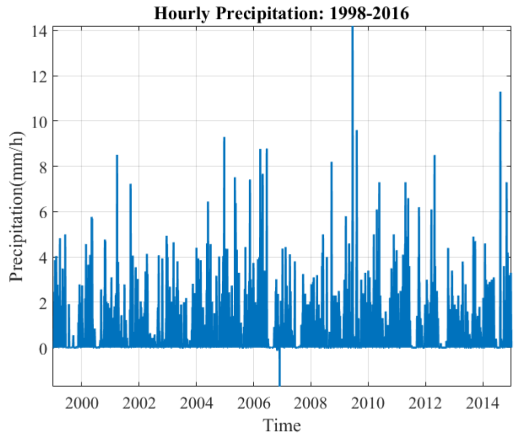

Figure 1: Precipitation Raw Data.

$$F(s,\xi) = e^{-(1+\xi s)^{-1/\xi}} \quad \forall \xi > 0$$

$$s = \frac{x-\mu}{\sigma} \quad \sigma > 0$$

Where s is the study variable standardised by the location parameter μ and the scale parameter σ, and ξ is the shape parameter. So the GEV distribution has three parameters to be fitted. The fitting results are shown in Fig.2.

Fitting quality was evaluated using Mean Squared Error (MSE) and it was found that $D = 1, 8, 10, 17$ had the worst fit with MSE greater than 0.003 and $D = 9, 15$ had the best fit with MSE less than 0.001.

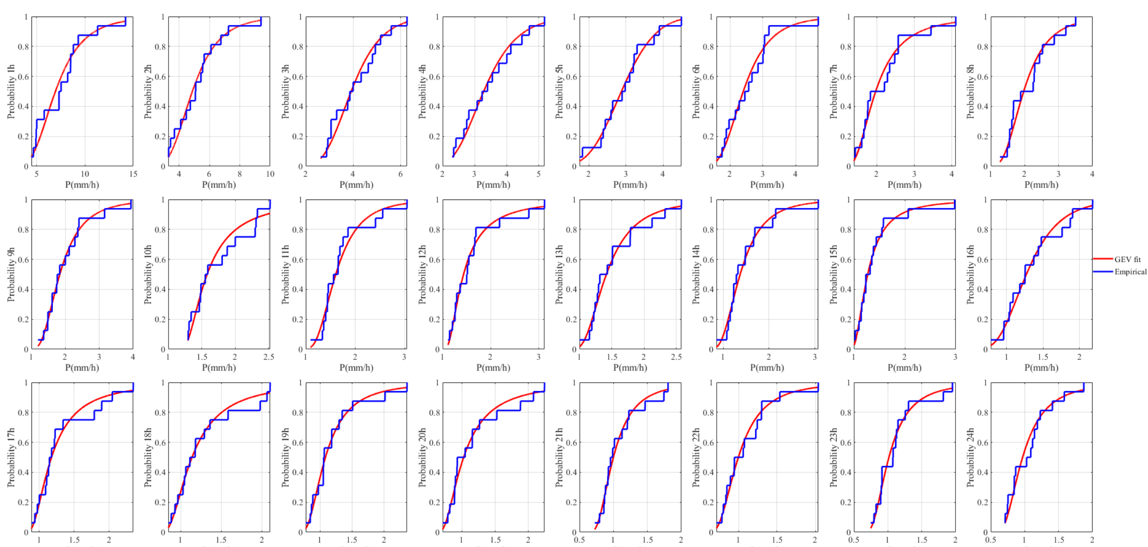

Figure 2: GEV Fit Result

2.3 Analytical Formulas Fitting

This paper uses four analytical formulas for interpolation between empirical points:

$$I = \frac{aR + b}{(D + c)^d} \tag{1}$$

$$I = \frac{aR + b}{D^c + d} \tag{2}$$

$$I = \frac{aR^b}{(D + c)^d} \tag{3}$$

$$I = \frac{aR^b}{D^c + d} \tag{4}$$

where I is the precipitation intensity, D is storm duration and R is return period; a, b, c and d are parameters. Optimize these parameters to data so that the analytical curves fit the empirical points.

The result is shown in Fig3. Analytical formulas with return periods of 10 and 100 years are shown in blue and red, respectively, and diamond icons are empirical data.

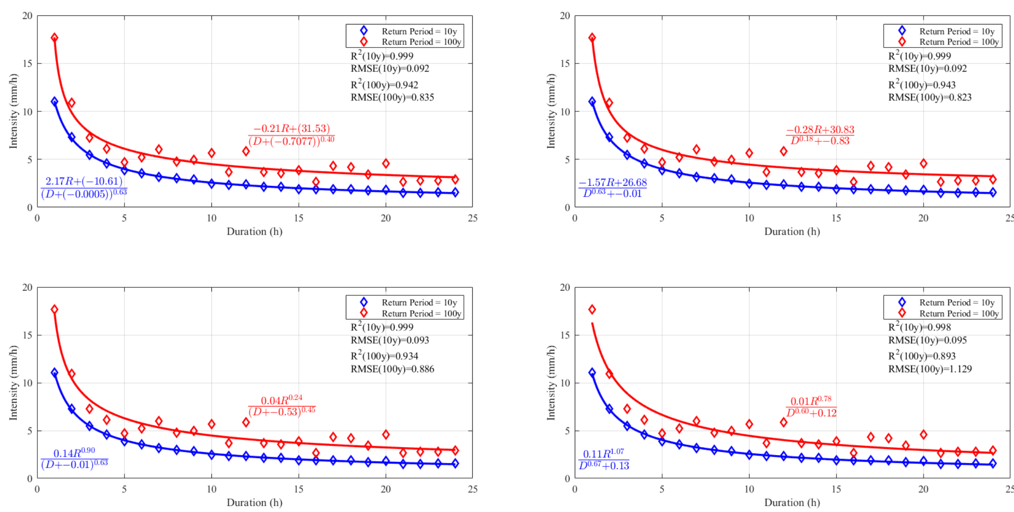

Figure 3: IDF Curves

In practice, the fitting results change with the position of the initial point, in addition, the analytical formula (1) needs to set the range of the parameters, otherwise there will be a plural error, the detailed settings can be viewed in the $IDFfunction$. Setting the range of parameters as well as the initial point position can be done using $MATLABcftool$ to assist in the setup. Table 1 and table 2 show the values of parameters.

Type	a	b	c	d
1	2.17	-10.61	-0.0005	0.63
2	-1.57	26.68	0.63	-0.01
3	0.14	0.90	-0.01	0.63
4	0.11	1.07	0.67	0.13

Table 1: Analytical formulas parameters for $R = 10y$

2.4 Comment

The flow rate for $D = 1h$ is $11.0461mm/h$ and $17.6544mm/h$ for storm events with a return period of 10 and 100 years, respectively.

Type	a	b	c	d
1	-0.21	31.53	-0.71	0.40
2	-0.28	30.83	0.18	-0.83
3	0.04	0.24	-0.53	0.45
4	0.01	0.78	0.60	0.12

Table 2: Analytical formulas parameters for $R = 100y$

In all four analytical formulas, the data with a 10-year return period fitted better with a larger R^2 (0.99 on average) compared to the 100-year data. In particular, the $RMSE$ is much smaller for the 10-year reproduction period than for the 100-year one, with mean values of 0.094 and 0.92, respectively. Comparing the fitting results of the different formulas, it can be found that formula (2) fits the best, with the highest R^2 among the four formulas at 0.99 and 0.94, and formula (4) is the worst, with the highest $RMSE$ at 0.095 and 1.129.