Entropy Theory and Pearson Type-3 Distribution for Rainfall Frequency Analysis in Semi-arid Region

Xiangming Kong¹*, Zhenda Hao² and Ying Zhu³

¹ School of Fundamental Science, Beijing Polytechnic, Beijing 100176, China
² Central University of Finance and Economics, Beijing 100081, China
³ School of Environmental and Municipal Engineering, Xi’an University of Architecture and Technology, Xi’an 710055, China

*E-mail: xiangming0913@126.com

Abstract. Rainfall distribution has become erratic due to the changing climate, the human activities and the environmental damage. Therefore, frequency analysis of rainfall is important for water-resources system planning and mitigation of hydrological extremes events. In this study, the entropy theory and Pearson type-3 distribution have been applied to monthly rainfall frequency analysis of Tongliao City in semi-arid region. The comparison between two methods shows that (1) the entropy theory and Pearson type-3 distribution are both able to capture the shape of PDF of monthly rainfall in Tongliao City; (2) the entropy theory is more suitable for the description of the statistical probabilities with small corresponding errors than the Pearson type-3 distribution; (3) the key statistical properties of historical records of Tongliao City are able to be preserved by the entropy-based distribution without any assumptions. The results could provide scientific basis for planning and design for water resources systems, which is of great theoretical and practical significance.

1. Introduction
In water resources management system, synthetic rainfall data are required to represent the observed distributional characteristics of the history records and to objectively reflect the situations of rainfall in a given area. Many methods have been developed for the frequency analysis of hydrological variables. Among them, Pearson type-3 distribution has been widely used for describing rainstorm, flood and drought events [1]. Recently, entropy theory has been proposed for constructing the univariate distribution of hydrological variables [2, 3]. For example, Hao and Singh applied entropy theory for simulation on single-site monthly streamflow, which could preserve key statistics of marginal distribution of historical streamflow data [4]. Kong et al. proposed a maximum entropy-Gumbel-Hougaard copula method for simulating monthly streamflow, in which the entropy method was used to construct the marginal distribution of data records in Xiangxi River [5]. The advantage of entropy theory is that no assumptions need to be considered in the marginal distribution. However, the effectiveness of entropy theory and Pearson type-3 distribution are still not demonstrated for frequency analysis of monthly rainfall in a semi-arid region of China.

Therefore, this paper would apply the entropy theory and the Pearson type-3 distribution for describing monthly rainfall of Tongliao City. The goodness-of-fit statistical tests, consisting RMSE and K-S test would be used to validate the availability of two methods.
2. Methodology

2.1. Entropy Theory
If the random variable \( X \) is continuous, then the Shannon entropy \( H(x) \) can be expressed as:

\[
H(X) = -\int_a^b f(x) \ln f(x) \, dx.
\]  

(1)

Where \( f(x) \) is the probability density function (PDF) of \( X ; x \in (a,b) \). The constraints can be expressed as:

\[
C_0 = \int_a^b f(x) \, dx = 1.
\]  

(2)

\[
C_i = \int_a^b h_i(x) f(x) \, dx = \bar{h_i}(x) \quad i = 1, 2, ..., m.
\]  

(3)

Where \( x \) is monthly rainfall value; \( \bar{h_i}(x) \) is the expectation of \( h_i(x) \). In accordance with the principle of maximum entropy, the most probable PDF for random variable \( X \) can be obtained by maximizing Equation (1) with constraints (2) and (3).

\[
f(x) = \exp[-\ln(\int_0^1 \exp(-\sum_{i=1}^m \lambda_i h_i(x)) \, dx) - \sum_{i=1}^m \lambda_i \bar{h_i}(x)].
\]  

(4)

One can write the objective function as:

\[
Z(\lambda) = \ln(\int_a^b \exp\left(-\sum_{i=1}^m \lambda_i h_i(x)\right) \, dx) + \sum_{i=1}^m \lambda_i \bar{h_i}(x)
\]  

(5)

Thus, the Lagrange multipliers \( \lambda_i (i = 1, 2, ..., m) \) in Equation (5) can be determined using the Conjugate Gradient method, which is of superlinear convergence, simple recurrence formula and less calculation.

2.2. Pearson Type-3 Distribution
The Pearson type-3 distribution was greatly convenient and useful for the frequency analysis in hydrology [1]. It is assumed that the random variable \( X \) is Pearson type-3 distributed, then the distribution can be described by the mean \( \bar{x} \), the coefficient of dispersion \( C_v \) and the coefficient of skew \( C_s \). The PDF for \( X \) can be obtained.

\[
f(x) = \frac{1}{\beta \Gamma(\gamma)} (x-\alpha)^{\gamma-1} e^{-\left(\frac{x-\alpha}{\beta}\right)}.
\]  

(6)

Where \( \Gamma(\gamma) \) is gamma function.

\[
\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t} \, dt.
\]  

(7)

\[
\bar{x} = \beta \gamma + \alpha.
\]  

(8)

\[
C_v = \frac{\sqrt{\gamma + \alpha \gamma}}{\beta}.
\]  

(9)

\[
C_s = \frac{2}{\sqrt{\gamma}}.
\]  

(10)

One can write the objective function as:

\[
\ln L = \sum_{i=1}^n \ln f(x_i) = -\sum_{i=1}^n \frac{x_i - \alpha}{\beta} - n \gamma \ln B - n \ln \Gamma(\gamma) + (\gamma - 1) \sum_{i=1}^n \ln(x_i - \alpha).
\]  

(11)

Thus, the parameters \( \alpha, \beta \) and \( \gamma \) in Equation (11) can be determined using the maximum likelihood estimation.
3. Application

Tongliao City is located at the east of Inner Mongolia. The importance of the location makes this city become a significant joint in promoting the joint-construction of the "Belt and Road" initiative. With a semi-arid climate affection, the annual precipitation of Tongliao City ranges from 305mm to 485mm. The precipitation is the replenishment source of surface water. Therefore, the frequency analysis of rainfall is significant for the water resources management in Tongliao City. Daily rainfall records from 2000 to 2018 at the Tongliao meteorological station were provided. The entropy theory and the Pearson type-3 distribution would be applied to the frequency analysis of rainfall data in Tongliao City. A comparison would be performed for the distributions generated by two methods.

3.1. Monthly Rainfall Data

The monthly rainfall data was scaled to [0,1] for computational convenience [5].

\[
SD = (OD - (1 - d)MN) / [(1 + d)MX - (1 - d)MN]
\]

where \(SD\) and \(OD\) are scaled and original values, respectively; \(MX\) and \(MN\) are maximum and minimum values, respectively; \(d\) is the parameter in the original domain.

3.2. Comparison of Generated Distributions

The PDFs for monthly rainfall in Tongliao City are generated by the entropy theory and the Pearson type-3 distribution, respectively. The generated PDFs are compared with observed rainfall data in Figure 1. The result shows that the theoretical PDF generated by the entropy theory fitted the observed data better than that generated by the Pearson type-3 distribution.

![Figure 1. Comparison between the generated PDFs and the observed data](image)

The Kolmogorov-Smirnov (K-S) and Root Mean Square Error (RMSE) and the goodness-of-fit test would be used for assessing whether the generated distribution is valid. The RMSE can be expressed as:

\[
RMSE = \sqrt{\frac{\sum_{k=1}^{K} (x_k^{\text{obs}} - x_k^{\text{th}})^2}{K}}
\]
Where $x_k^{\text{est}}$ and $x_k^{\text{obs}}$ are the estimated and observed values, respectively; $K$ is the sample size. The K-S test statistic can be stated as [6]:

$$T = \sup_x \left| F^*(x) - F_n(x) \right|$$

(14)

Where $F^*(x)$ and $F_n(x)$ are the generated and empirical distributions, respectively. The results of goodness-of-fit test are shown in Table 1. The P-values of January, February, March, November and December calculated from K-S test for Pearson type-3 distribution are lower than the significant level 0.05. Only the P-value of January calculated from K-S test for entropy theory is lower than the significant level 0.05. These results indicate that the entropy-based distribution could appropriately represent the observed rainfall records in Tongliao City. Moreover, the RMSE indicate that the errors of entropy-based distribution are much smaller than that of Pearson type-3 distribution. It also indicates the accuracy of entropy theory in capturing the shape of rainfall distribution in Tongliao City.

| Month | Entropy theory | Pearson type-3 distribution |
|-------|----------------|-----------------------------|
|       | K-S            | RMSE                        | K-S          | RMSE          |
|       | $T$ P-value    |                            | $T$ P-value  |               |
| 1     | 0.291 0.032    | 0.26 0.329                 | 0.329 0.012  | 0.93          |
| 2     | 0.186 0.238    | 0.44 0.273                 | 0.273 0.048  | 0.70          |
| 3     | 0.134 0.465    | 0.96 0.262                 | 0.262 0.061  | 1.25          |
| 4     | 0.087 0.711    | 1.93 0.089                 | 0.089 0.702  | 2.45          |
| 5     | 0.094 0.674    | 4.15 0.120                 | 0.120 0.536  | 5.15          |
| 6     | 0.128 0.497    | 11.53 0.141                | 0.141 0.436  | 11.79         |
| 7     | 0.059 0.841    | 7.60 0.132                 | 0.132 0.476  | 6.05          |
| 8     | 0.099 0.648    | 13.07 0.172                | 0.172 0.291  | 20.29         |
| 9     | 0.098 0.653    | 5.23 0.145                 | 0.145 0.410  | 5.60          |
| 10    | 0.111 0.584    | 4.85 0.155                 | 0.155 0.365  | 5.96          |
| 11    | 0.195 0.207    | 3.12 0.343                 | 0.343 0.008  | 5.27          |
| 12    | 0.165 0.322    | 0.57 0.294                 | 0.294 0.030  | 0.60          |

Figure 2 shows the comparison of mean, skewness, standard deviation and kurtosis between observed and generated monthly rainfall. The results indicate that the entropy theory can better reflect the statistical characters of monthly rainfall than the Pearson type-3 distribution. Moreover, the entropy theory has an advantage that it can construct the distribution of random variable without any assumptions.
4. Concluding Remarks
In this study, the entropy theory and the Pearson type-3 distribution are applied to the rainfall frequency analysis of Tongliao City, China. The following conclusions are drawn: The entropy theory and the Pearson type-3 distribution are both able to capture the shape of PDF of monthly rainfall in Tongliao City. The RMSE and K-S test investigate that the entropy theory is more suitable for the description of the statistical probabilities than the Pearson type-3 distribution. Moreover, the entropy-based distribution can better preserve the key statistical properties of monthly rainfall.

Acknowledgments
This research was supported by the Technological Major Projects of Beijing Polytechnic (2019Z002-007-KXB).

References
[1] Mojca Šraj, Bezak N, Brilly M. Bivariate flood frequency analyses using Copula function. EGU General Assembly 2013, Geophysical Research Abstracts, 2013, 15, EGU2013-5105
[2] Serinaldi F, Zunino L, Rosso O A. Complexity–entropy analysis of daily stream flow time series in the continental United States. Stochastic Environmental Research & Risk Assessment, 2014, 28(7):1685-1708
[3] Dariane A. B., Farhani M., Azimi S. Long term streamflow forecasting using a hybrid entropy model. Water Resources Management, 2018, 32(4):1439-1451
[4] Hao Z, Singh VP (2012) Entropy-copula method for single-site monthly streamflow simulation. Water Resour Res 48(6):W06604.
[5] Kong X M , Huang G H, Fan Y R, Li Y P. Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. Stochastic Environmental Research and Risk Assessment, 2015, 29(3):833-846
[6] Kong X M , Zeng X T, Chen C, Fan Y R, Huang G H, Li Y P, Wang C X. Development of a Maximum entropy-Archimedean Copula-based Bayesian Network method for streamflow frequency analysis—A case study of the Kaidu River Basin, China. Water, 2019, 11, 42