Distribution selection for hydrologic frequency analysis using subsampling method

S Das\textsuperscript{1, 2}

\textsuperscript{1}College of Hydrometeorology, Nanjing University of Information Science and Technology (NUIST), Nanjing, China

E-mail: samirandas@gmail.com

Abstract. This paper investigates the potential utility of subsampling, a resampling technique with the aid of a goodness of fit test to select the best distribution for frequency analysis. Subsampling draws samples (of smaller size) from the original sample without replacement. The performance of the methodology is assessed by applying the methodology to an observed annual maximum (AM) hydrologic data series. Several AM discharge series of different record lengths are used as case studies to determine the performance. Overall, it is found that the methodology is suitable for a longer data series and a good performance can be obtained when the subsample size is around half of the underlying data sample. The methodology has also outperformed the standard AD test in terms of effectively discriminating between distributions. All results indicate that the subsampling technique can be a promising tool in discriminating between distributions.

1. Introduction

Hydrologic frequency analysis is concerned with the assessment of magnitudes (such as flood or heavy rainfall) of stated frequency or stated degree of rarity for use as input to the process of hydrologic risk assessment and management. The application of probability theory in a standard frequency approach requires the determination of a probability distribution and the estimation of its parameters. While several procedures related to parameter estimation, such as L-moments [1], provide satisfactory results in dealing with parameter estimation of a wide range of probability distributions, that has not been the case about the methods that select an appropriate probability distribution function. The selection of a suitable probability distribution for a particular gauged site is still a major problem in frequency analysis. The selection of a suitable distribution leads to a reduction in the error of quantile estimates. In hydrologic frequency analysis, different types of distributions have been considered in different contexts, and a list of such can be obtained from the WMO report [2]. The choice of such distributions is generally selected on the basis of graphical methods (such as a probability plot or L-moment ratio diagram) or by goodness of fit tests such as a chi-square test [3], Kolmogorov–Smirnov (KS) test [4], Anderson–Darling (AD) test [5,6], or probability plot correlation coefficient test [7]. These tests, however, have often been criticized for their inability to discriminate between statistical distributions for the same application [2,6]. At the same time, it has been acknowledged that GOF tests can sensibly be used in frequency analysis as the basis for rejecting some distributions, but not to select a best population distribution.

Thus, there is a need for alternative approaches to attempt further distinguishing between flood frequency distribution selections. In a more recent time, [8] presented an objective model selection
criterion based on Akaike information, Bayesian Information and Anderson–Darling Criterion. However, as shown by the authors, the performance in the context of three-parameter distribution selection is not satisfactory. Other notable developments in the context of flood frequency analysis are the Bayesian hierarchical model and climate-informed methods [9,10].

The objective of this paper is to assess how subsampling [11], a data intensive resampling technique [12], can be used as a tool to overcome difficulties in selecting a best population distribution for a site with relatively long record length where the discrimination between statistical distributions is challenging. Subsampling draws samples (of smaller size) from the original sample without replacement. The reason the subsampling can be used in distribution selection is due to the fact that the subsamples are actually samples from the true unknown distribution, as was the case with the original sample. So, if the traditional methods failed to discriminate between distributions for the original sample, it is expected that a huge number of subsamples drawn from the original sample would provide a pool of data series that can effectively be used to determine an outright single choice for a gauged site. The study assesses the performance by applying the methodology to observed annual maximum (AM) flow data series. Several AM flow series of different record lengths pertaining to catchments located in different regions of U.K. are used as case studies.

2. Methodology
This section details the methodology applied in the study, which includes subsampling technique, statistical models and the proposed test procedure.

2.1. Subsampling
Subsampling falls within the broad area of resampling technique. Popular resampling techniques such as bootstrap [13] and subsampling [11] are data intensive simulation methods for statistical inference. They create an ensemble of datasets, each of which is drawn from the original sample of size \( n \). In classical bootstrap, samples of size \( n \) are drawn with replacement while in subsampling, and samples of size \( b \) (\( b < n \)) are drawn without replacement. The reason the subsampling can be used in distribution selection is that the subsamples are actually samples (of smaller size) from the true distribution \( F \). On the other hand, resamples (e.g. bootstrap) are samples from an estimator of \( F \). Here, \( F \) is a probability (or cumulative) distribution function, \( F(q) = \text{prob}(Q_i \leq q), \quad \text{for} \quad i = 1, \ldots, N \), where \( q \) is any real number. A number of \( B \) i.i.d. samples \( Q^{(1)} \), \ldots, \( Q^{(B)} \) (each of size \( b \)) are drawn from the sample population consisting of the observations \( \{Q_1, \ldots, Q_N\} \). In subsampling terminology, these \( B \) i.i.d. samples are called subsamples. A good summary of subsampling can be obtained from [11].

2.2. Statistical techniques
The statistical approach associated with frequency analysis is focused on finding an appropriate form to model the underlying distribution of flood data and then estimating the parameters of this underlying distribution. In this study, two-parameter distributions, such as Extreme value I (EV1) or Gumbel (G), and three-parameter distributions, such as Generalized Extreme Value (GEV), Generalized Logistic (GLO), Pearson Type 3 (PE3), Generalized Normal (GNO) and Generalized Pareto (GPA), are considered. These distributions have been widely used in flood frequency analysis [14]. This study uses the method of L-moments to estimate distribution parameters. The theoretical expressions and the formulae for parameter estimation using L-moments of the candidate distributions are given by [14]. In this study, a goodness of fit test (GOF) based on the Anderson-Darling (AD) test statistic [6,15,16] is used to evaluate the suitability of different probability distributions. The AD test is regarded as the most powerful and preferred test available at present [6].
2.3. Test procedure
The proposed subsampling methodology can be applied to an at-site annual maximum (AM) hydrologic data series to select the most appropriate distribution for frequency analysis. The test procedure has the following steps:

- A large number of subsamples (i.e., 1000) of size \( b \) (\( b < n \)) are drawn from the original data series of size \( n \) without replacement.
- The Anderson-Darling (AD) test is then applied to each subsample drawn from the original sample. The AD test statistic is computed for each candidate distribution.
- The distribution is chosen that is accepted most frequently by the test statistic at the 5% significance level.

By the above description, one can understand that the methodology is aimed to develop a hydrologic decision tool for a longer data series. Generally, the recommended size (record length) for performing at-site frequency analysis is 20 years [17]. In this procedure, the AD test is applied to each subsample to determine their frequency distribution. Thus, the record length 20 years is considered as minimum subsample size (\( b \)). It then requires the sample size (\( n \)) to be greater than 20 in the subsampling procedure. However, the selection of an appropriate subsample size with respect to a sample is an important aspect of this methodology. This is also regarded as an important issue when determining a confidence interval with the subsampling approach [18]. This will be examined in detail in the following sections.

3. Application and results
Four gauged stations are selected as case studies to apply the subsampling methodology. They (Stations 39001, 54002, 21006, 12001) pertain to catchments located in different regions of U.K. They are different in record lengths, and the records vary from 46 to 130 years. Table 1 shows the necessary information about the data series such as location, record length and statistical characteristics in terms of L-moments. These datasets are taken from the Centre for Ecology & Hydrology (CEH), U.K.

| Station No. | Station Name | River | Record Length (yrs) | AREA (km²) | Statistical characteristics | Mean/L1 | L-CV | L-Skew | L-kur |
|-------------|--------------|-------|---------------------|-----------|-----------------------------|---------|------|--------|-------|
| 39001       | Kingston     | Thames| 130                 | 9948      |                             | 325.4   | 0.195| 0.138  | 0.168 |
| 54002       | Evesham      | Avon  | 76                  | 2210      |                             | 154.2   | 0.272| 0.164  | 0.206 |
| 21006       | Boleside     | Tweed | 46                  | 1500      |                             | 428.7   | 0.189| 0.174  | 0.140 |
| 12001       | Woodend      | Dee   | 70                  | 1370      |                             | 427.6   | 0.207| 0.075  | 0.197 |

Popular tests, namely the Anderson-Darling (AD) goodness of fit test and L-moment ratio (LMR) diagram, are applied to identify the appropriate distributions for each of the selected data series. Table 2 summarizes the AD test result for each of the AM data series. By ‘Reject’ (or ‘Accept’), it means that the distributions are rejected (or accepted) by the test statistic at the 5% significance level. For stations 39001, 54002 and 21006, the AD test rejected only one distribution, namely GPA, out of six candidate distributions, while for station 12001 the test rejected all the considered distributions. Overall, for each AM data series, the test statistic has failed to discriminate among the candidate distributions. Plots of dimensionless L-moment ratios (LMR) are shown in figure 1, which compares the observed and the theoretical relations between L-skewness and L-kurtosis for each of the AM data series. The LMR values of stations 12001 and 54002 lie above the theoretical GLO line, while for station 39001 it is positioned between the GEV and GLO line. In the case of 21006, the value lies between the GNO and PE3 line, and the value is very close to the GEV line and the EV1 (G) point.
all cases, there is enough doubt to get an outright choice for respective data series. Thus, in this context, the L-moment ratio diagrams cannot be regarded as suitably reliable.

**Table 2.** Anderson-Darling goodness of fit test results at the 5% significance level applied to the selected stations.

|        | NOR | EV1 | GEV | GLO | PE3 | GNO | GPA |
|--------|-----|-----|-----|-----|-----|-----|-----|
| ST39001| Reject | Accept | Accept | Accept | Accept | Accept | Reject |
| ST54002| Reject | Accept | Accept | Accept | Accept | Accept | Reject |
| ST21006| Accept | Accept | Accept | Accept | Accept | Accept | Reject |
| ST12001| Reject | Reject | Reject | Reject | Reject | Reject | Reject |

Figure 1. L-moment ratio diagram showing the point values for selected AM series.

The proposed subsampling methodology is applied to each of the AM data series. A large number of subsamples (e.g. 10000) were drawn from the original data series, taking a set of different subsample size (b) without replacement. The set of different b was chosen for particular AM series according to their record lengths. This has been indicated in figure 2. The algorithm in [19] was used to draw subsamples from the original data series without replacement. The Anderson-Darling (AD) test was then applied to each subsample, and a summary result has been produced for each distribution. Figure 2 shows the results in terms of percentages of accepted distributions for each station. Different values are obtained for different distributions and for different subsample sizes. It is noted that for the distributions that are accepted at the 5% significance level for the original data series, the acceptances of those distributions increase with the increase of b. However, at certain subsample sizes, discrimination is difficult to achieve. For example, at b= 100 for station 39001, an 100% acceptance rate was observed for four distributions: GEV, GLO, PE3 and GNO. However, for other subsample sizes such as b= 20, 40 and 60, the discrimination was good enough to choose the most appropriate distribution. A similar kind of conclusion can be obtained for station 54002. Due to its short record length (46), the results of station 21006 are reported for b= 20, 25, 30, 35 and 40. The discrimination is difficult in this context, as it was also shown by the AD test and LMR diagram with the original data. However, at b=20 the GNO was accepted slightly more than the GEV and EV1. A similar observation was also noted for b=25. Although none of the distributions were accepted by the AD test for station 12001, the test procedure allows the identification of an appropriate distribution. The results so
obtained were then used to produce frequency curves. Figure 3 shows four probability plots: the GLO fit for stations 39001, 54002 and 12001 and the GNO fit for station 21006. The flood data are plotted against the return periods corresponding to the logistic variate for GLO and the normal reduced variate for GNO. The observed data are presented as small circles and the flood quantiles are presented as lines. There are two relatively high events (see the probability plot for station 12001 in Figure 3), which might have caused the distributions to perform so poorly for station 12001, but with this test procedure, the discrimination among distributions was achieved. The relatively poor performance of the test method in differentiating between distributions for station 21006 may be owing to reasons such as short record length and the characteristics of L-moment statistics.

One of the key observations from the case studies is that the finding of an appropriate size of b is an important task. It is found that the discrimination was recognized when the size of the subsample was around half of the sample size. Thus, taking \( b = \frac{1}{2} \) of \( n \) is a good choice to start a subsampling experiment in this context. The methodology also provides a degree of acceptance that is absent for a GOF test. This can allow ranking of the distributions or selection of an appropriate distribution in case none of the considered distributions are accepted by a GOF test.

![Figure 2](image)

**Figure 2.** Bar charts showing the results of the Anderson-Darling goodness of fit test applied to the subsamples of different sizes drawn from the selected AM data series. Each panel represents the station number, and different colors indicate different types of distribution considered in this study.
Figure 3. Probability plot: GLO fit for stations 39001, 54002, and 12001 and GNO fit for station 21006. The observed data are presented as small circles and the flood quantiles are presented as lines. The flood data are plotted against the return periods ($T$).

4. Discussion and conclusion

This study was conducted to investigate the potential utility of a subsampling technique to discriminate between distributions for use in hydrologic frequency analysis. The concept is applicable to the fact that the subsamples are actually samples from the true unknown distribution, as was the case with the original sample. The idea is to analyse a huge number of subsamples obtained from the original sample to better discriminate between distributions that otherwise may not be possible with the original sample. The methodology can be thought of as the framework opposite to the regional analysis [20, 21]. In a regional analysis framework, the flood data from different sites in a homogeneous region are assumed to come from a population distribution, and the distribution that is accepted in the highest number of cases or a similar sort of arrangement is chosen for frequency analysis. In this context, a large number of subsamples are drawn from a sample, and the distribution that is accepted by a goodness of fit test in the highest number of cases is chosen for frequency analysis.

To demonstrate the performance, the methodology was applied to observed data. Four gauged stations pertaining to catchments in U.K. were selected as case studies to apply the subsampling methodology. When applied to observed data, it was found that with the proper choice of subsample size, the methodology performed better in discriminating among distributions. Generally, the performance of the methodology decreased as the subsample size approached the original sample size. The methodology performed better when the subsample was around half of the sample size. Thus for practical purposes, the use of $n/2$ can be endorsed, but it is recommended to try a range of sizes and see what the overall behaviour is. It was also found that a longer data series produced better results compared to a shorter record length. The method also provides a degree of acceptance, which can be useful for practitioners in decision making.

However, it is recognized that in some cases it may not perform as a full proof tool because it depends on the power of AD testing, which has its own limitations. Then again, in comparison with the standard AD tests, the proposed methodology performs considerably better, and a sensible use of
the methodology can give good results. A good operational strategy could be to use the AD and LMR tests in combination with the subsampling technique. The first two methods can be used to eliminate inappropriate distributions from the candidate distributions. If two or more are selected by the tests, then the proposed methodology can be applied to select the best population distribution. Overall, the methodology can be regarded as an interesting complement to standard statistical testing because it presents the advantage of identifying the best probability distribution with a degree of acceptance.

Acknowledgement
The author would like to acknowledge the financial support in the form of a faculty start-up fund made available by the College of Hydrometeorology, Nanjing University of Information Science and Technology (NUIST), Nanjing, China. The observed annual maximum flood data series of selected stations was obtained freely from the Centre for Ecology & Hydrology (CEH), UK website: http://www.ceh.ac.uk/data/nrfa/data.

References
[1] Hosking J R M 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics Journal of the Royal Statistical Society Series C (Applied Statistics) 52 105–24
[2] Cunnane C 1989 Statistical distributions for flood frequency analysis (Geneva: Operational Hydrology Report (WMO))
[3] Das S, Millington N and Simonovic S P 2013 Distribution choice for the assessment of design rainfall for the city of London (Ontario, Canada) under climate change Can. J. Civ. Eng. 40 121–9
[4] Chowdhury J U, Stedinger J R and Lu L 1991 Goodness–of–fit tests for regional generalized extreme value flood distributions Water Resources Research 27 1765–76
[5] Ahmad M I, Sinclair C D and Spurr B D 1988 Assessment of flood frequency models using empirical distribution function statistics Water Resources Research 24 1323
[6] Laio F 2004 Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters Water Resources Research 40 1–10
[7] Vogel R W and McMartin D E 1991 Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution Water Resources Research 27 3149–58
[8] Laio F, Di Baldassarre G and Montanari A 2009 Model selection techniques for the frequency analysis of hydrological extremes Water Resources Research 45 1–11
[9] Yan H and Moradkhani H 2014 A regional Bayesian hierarchical model for flood frequency analysis Stochastic Environmental Research and Risk Assessment
[10] Renard B and Lall U 2014 Regional frequency analysis conditioned on large-scale atmospheric or oceanic fields Water Resources Research 50 9536–54
[11] Anon (Springer Series in Statistics) Dimitris N P, Joseph P R and Michael W 1999 Subsampling-Springer
[12] Politis D N 1998 Computer-intensive methods in statistical analysis ieee Signal Processing Magazine 15 39–55
[13] Efron B 1979 Bootstrap methods: another look at the jackknife The Annals of Statistics 7 1–26
[14] Hosking J R M and Wallis J R 1997 Regional frequency analysis: an approach based on L-moments (Cambridge University Press)
[15] Stephens M 1986 Tests based on EDF statistics. In Goodness-of-fit techniques ed R.B. D’Agostino and M A Stephens (New York: Marcel Dekker, Inc.)
[16] Das S and Simonovic S P 2012 Assessment of uncertainty in flood flows under climate change impacts in the upper Thames River Basin, Canada British Journal of Environment & Climate Change 2 318–38
[17] Institute of Hydrology 1999 Flood Estimation Handbook (Wallingford, U.K.: Institute of Hydrology)
[18] Bickel P J and Sakov A 2008 On the choice of m in the m out of n bootstrap and confidence bounds for extrema *Statistica Sinica* **18** 967–85
[19] Green B F 1977 Fortran subroutines for random sampling without replacement *Behavior Research Methods & Instrumentation* **9** 559–59
[20] Das S and Cunnane C 2011 Examination of homogeneity of selected Irish pooling groups *Hydrology and Earth System Sciences* **15** 819–30
[21] Das S and Cunnane C 2012 Performance of flood frequency pooling analysis in a low CV context *Hydrological Sciences Journal* **57** 433–44