Some best-fit probability distributions for at-site flood frequency analysis of the Ume River

Samara Kousar | Abrar Raza Khan | Mahmood Ul Hassan
Zahra Noreen | Sajjad Haider Bhatti

Abstract
At-site flood frequency analysis is a direct method of flood estimation at a given site. The choice of an appropriate probability distribution and parameter estimation method plays a vital role in at-site frequency analysis. In the current article, flood frequency analysis is carried out at five gauging sites of the Ume River in Sweden. Generalised extreme value, three-parameter log-normal, generalised logistic and Gumbel distributions are fitted to the annual maximum peak flow data. The L-moment and the maximum likelihood methods are used to estimate the parameters of the distributions. Based on different goodness-of-fit tests and accuracy measures, the three-parameter log-normal distribution has been identified as the best-fitted distribution by using the L-moments method of estimation for gauging sites Harrsele Krv, Gardiken and Överuman Nedre. The generalised extreme value distribution with the L-moments estimation provided the best fit to maximum annual streamflow at gauging sites Solberg and Stornorrfors Krv. Finally, the best-fitted distribution for each gauging site is used to predict the maximum flow of water for return periods of 5, 10, 25, 50, 100, 200, 500, and 1000 years.

Keywords
generalised extreme value, generalised logistic, Gumbel, L-moments, log-normal, maximum likelihood

1 INTRODUCTION

A flood is an overflow of water from the river, lake, or stream channel that submerges land which is usually dry. Floods are one of the most common natural hazards that cause disasters like mortality, economic and communication infrastructure destruction and damages in environmental and agricultural assets. The most common reasons for floods include heavy rainfall, melting snow around the year and the lack of capacity of natural watercourses to convey excess water. Floods are natural phenomena that occur from time to time in rivers. The purpose of flood frequency analysis is to estimate the flood discharge at a particular site. The prediction of flood magnitude is important for a wide range of engineering problems such as dams, spillways, road and railway bridges, culverts, and flood control works, and so forth.

The choice of an appropriate probability distribution and parameter estimation method is important to
describe the flood frequency at a particular site. In the present study, we have selected generalised extreme value (GEV), three-parameter log-normal (LN3), generalised logistic (GLO) and Gumbel (GUM) as candidate distributions to fit the annual flow data at five gauging sites of Ume River. These distributions are commonly traced as the best-fitted distributions for at-site and regional flood frequency analysis in hydrology literature (Cunnane, 1989). Castellarin et al. (2012) reviewed applied statistical methods for flood frequency analysis in Europe. According to their review, the GEV was among the recommended distributions for Austria, Italy, Germany and Spain. Gumbel (GUM) and GEV were recommended for Finland and Turkey, respectively, for modelling annual maximum flood flow. Rahman et al. (2013) conducted at-site frequency analysis in Australia and identified the log-Pearson Type III, GEV, and generalised Pareto as best-fitting probability distributions for 127 sites. Saf (2009) recommended GLO distribution for the Upper-West Mediterranean sub-region in Turkey.

Several methods have been developed to estimate the parameters of hydrologic frequency models. The method of moments (MOM), probability weighted moments (PWM), maximum likelihood (ML) method and L-Moments (LM) method are the most commonly used methods in hydrology for parameter estimation. The ML method is widely accepted as the most powerful method for the estimation of parameters. It is an efficient method as it produces parameter estimates with the smallest sampling variance (Hamed & Rao, 1999). Recently, the LM method has received considerable attention in the field of statistical hydrology and its use is rapidly growing for parameter estimation in flood frequency analysis. It is a less complicated and computationally easier method for estimation of parameters. In the current article, we used the ML and LM methods to estimate the parameters of the candidate probability distributions for analysis of annual maximum peak flow data.

The performance of fitted probability distributions with a particular estimation method is compared using different accuracy measures and goodness-of-fit (GOF) tests.

This is the first study focused on at-site flood frequency analysis of Ume River. The objective of the study is to identify the best suitable distribution among the candidate distributions for annual maximum peak flow data of five gauging sites separately using the ML and LM methods. In addition, another objective is to estimate the quantiles for different return periods (like 5, 10, 25, 50, 100, 200, 500, and 1000 years) with non-exceedance probability using the estimated parameters of the best-suited probability distribution for each gauging site.

The rest of the article is organised as: Section 2 describes the area of study and data used for the analysis. Section 3 deals with the model description, parameter estimation methods and performance indicators used for comparison. Section 4 provides results and discussion of the application of LM and ML based flood frequency analysis at five gauging sites of the Ume River. Finally, Section 5 concludes the article.

2 | STUDY SITES AND DATA

The Ume River is one of the main rivers of northern Sweden and flows from south to east from its source, Överum lake in the Scandinavian mountain range. It is about 449 km long with a catchment area of 26815 km². The annual maximum flow data from five gauging sites of the Ume River (Swedish: Umeälven) are studied in this article. A map of the Ume river with the gauging sites used for analysis is presented in Figure 1. The data have been collected from the Swedish meteorological and hydrological institute (n.d.) (www.smhi.se). The range of the data series varies from 52 to 104 years. Gauging sites' characteristics are summarised in Table 1. Time plots of annual maximum flow (AMF) at different gauging sites are presented in Figure 2 for the better visual representation of data.

3 | METHODOLOGY

3.1 | Candidate probability distributions

The choice of an appropriate probability distribution is important to describe flood frequency at a particular site. We have considered GEV, GLO, LN3, and GUM distributions for the analysis of flood frequency at five gauging sites of the Ume River. These distributions are recommended for flood frequency analysis in many countries (Castellarin et al., 2012; Cunnane, 1989; Guru & Jha, 2015; Rahman et al., 2013; Vogel & Wilson, 1996). The probability density function (pdf) and summarised in Table 2. We have used the LM and the ML method to estimate the parameters of the candidate probability distributions. A brief description of these methods is given in the following sub-sections.

3.2 | ML method

The ML method is one of the most applied methods for parameter estimation of probability distributions. The ML estimates are obtained by maximising the likelihood function or log-likelihood function of the probability distribution. Let we have a random sample of n independent identically quantile function of these distributions are distributed observations $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n$ for
FIGURE 1  A map of the Ume River showing the considered gauging stations

TABLE 1  Characteristics of gauging sites

| Station name     | Station no. | Time period | Longitude | Latitude | Area    | Elevation |
|------------------|-------------|-------------|-----------|----------|---------|-----------|
| Harsele Krv      | 1733        | 1958–2015   | 64.02     | 19.57    | 13386.20| 200 m     |
| Stornorrfors Krv | 1734        | 1958–2014   | 63.86     | 20.05    | 26567.80| 100 m     |
| Överuman Nedre   | 1856        | 1964–2015   | 65.96     | 15.03    | 651.50  | 600 m     |
| Gardiken         | 20011       | 1961–2014   | 65.48     | 15.86    | 4362.30 | 400 m     |
| Solberg          | 436         | 1912–2015   | 65.75     | 15.40    | 1083.56 | 500 m     |

FIGURE 2  Time series plots of five gauging stations of the Ume River
TABLE 2 Probability density and quantiles functions of the probability distributions

| Distributions | Probability density function $f(\tilde{z})$ | Quantile function $\phi(F)$ |
|---------------|-------------------------------------------|-----------------------------|
| GEV           | $\frac{1}{\alpha} \left[ 1 - k \left( \frac{\tilde{z} - \mu}{\alpha} \right) \right]^{(1/k-1)} \exp \left( -1 - k \left( \frac{\tilde{z} - \mu}{\alpha} \right) \right)$ | $\mu + \frac{\alpha}{k} \left[ 1 - (-\log(F))^k \right]$ |
| LN3           | $\frac{1}{\alpha \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\tilde{z} - \mu}{\sigma} \right)^2 \right] \left[ 1 + \frac{1}{2} \left( \frac{\tilde{z} - \mu}{\sigma} \right) ^2 \right]$ | The explicit analytical form is not available |
| GLO           | $\frac{1}{\alpha} \left[ 1 - k \left( \frac{\tilde{z} - \mu}{\alpha} \right) \right]^{(1/k-1)} \left[ 1 + \frac{1}{2} \left( \frac{\tilde{z} - \mu}{\alpha} \right) ^2 \right]$ | $\mu + \frac{\alpha}{k} \left[ 1 - (1-F)/F \right]^k$ |
| GUM           | $\frac{1}{\mu \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\tilde{z} - \mu}{\sigma} \right)^2 \right]$ | $\mu - \alpha \left[ \log(-\log(F)) \right]$ |

Note: $\mu$, $\alpha$, and $\kappa$ represent the location, scale and shape parameters of the probability distributions.

which the probability density function (pdf) of each $\tilde{z}_i$ is $f(\tilde{z}_i|\theta)$, then, the likelihood ($L(\theta)$) of the data set is the product of the pdf evaluated at each of the observed value of the data and can be written as $L(\theta) = \prod_{i=1}^{n} f(\tilde{z}_i|\theta)$. The log-likelihood function ($l(\theta)$) can be expressed as $l(\theta) = \sum_{i=1}^{n} \log f(\tilde{z}_i|\theta)$, where $\theta$ is an unknown parameter (or parameter vector). The ML estimate of $\theta$ is the value of $\theta$ that maximises $l(\theta)$ for given data $\tilde{z}$. Maximisation of $l(\theta)$ is achieved by some numerical optimisation technique (like, Nelder and Mead, differential evolution, Newton-Raphson method, etc.). In this article, we have used the Nelder and Mead (1965) method for numerical optimization.

3.3 | LM method

The LM introduced by Hosking (Hosking, 1986; Hosking, 1990) are linear functions of probability weighted moments (PWM). The LM are more convenient than PWM as they can be directly interpreted as a measure of the scale and shape of the distribution. It is convenient to begin with the PWM to find the LM (Hosking & Wallis, 2005). For a given distribution that has a quantile function $\phi(F)$, the PWM $\beta_r$ is defined as

$$\beta_r = \frac{1}{0} \phi(F) F^r dF \quad r = 0, 1, 2, ...$$

The unbiased sample estimators of $\beta_r$, the first four PWM for given distribution can be computed as

$$b_0 = n^{-1} \sum_{i=1}^{n} \tilde{z}_{j,n}$$
$$b_1 = n^{-1} \sum_{i=2}^{n-1} \frac{(j-1)}{(n-1)} \tilde{z}_{j,n}$$

Hosking (1990) described the LM ratios as L-coefficient of skewness ($r_1 = \frac{\lambda_2}{\lambda_1}$) and L-coefficient of kurtosis ($r_4 = \frac{\lambda_3}{\lambda_1}$), where $\lambda_1$ represents the measure of central tendency and $\lambda_2$ is a measure of dispersion.

Similar to the method of moments, in the method of LM, parameter estimates are obtained by equating the sample LM with the corresponding theoretical LM and solving the resulting system of equations for unknown parameters.

3.4 | SE of parameter estimates

The SE indicate the reliability of estimated parameters and the performance of estimation methods. The SE of estimated parameters are obtained by using a parametric bootstrap method (Meylan, Favre, & Musy, 2012). The parametric bootstrap method is very similar to the Monte Carlo simulation. The process to obtain the SE of the
estimated parameters for each distribution using a simulation study is summarised below:

- We have drawn 1,000 samples of size equal to the length of data from each distribution at each gauging site using estimated parameters with both LM and ML methods.
- For each simulated sample, the ML and LM estimates are obtained.
- The SE are obtained by taking the standard deviation of the sampling distribution of ML and LM estimates over 1,000 simulated samples.

### 3.5 GOF tests

The GOF test is one of the ways that can be used to assess how well a particular distribution fits the given data. These tests describe the differences between the empirical (or sample) cumulative distribution function (CDF) of data and theoretical (or expected) CDF of the distribution being tested. The Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramér–von Mises (CVM) GOF tests are used in this article.

The test statistics for KS, CVM, and AD test are defined as,

\[
\text{KS} = \text{Max}\{ |F(\hat{z}_i) - \hat{F}(\hat{z}_i)| \}.
\]

\[
\text{CVM} = \frac{1}{12n} + \sum_{i=1}^{n} \left( \frac{2i-1}{2n} - F(\hat{z}_i) \right)^2.
\]

\[
\text{AD} = -n - \sum_{i=1}^{n} \left( \frac{2i-1}{n} \left( \frac{\log(1-F(\hat{z}_{i-1}+1)) + \log(F(\hat{z}_i))}{\log(1-F(\hat{z}_i))} \right) \right).
\]

In all the above GOF tests \( F(\hat{z}_i) \) represents the empirical CDF of the data (observed values) and \( \hat{F}(\hat{z}_i) \) is the theoretical CDF of the distribution.

### 3.6 Accuracy measure methods

Accuracy measure is another way of describing how well a particular model fits given data. In the current article, we have considered some common accuracy measures like root mean square error (RMSE), mean absolute error (MAE), root mean squared percentage error (RMSPE), mean absolute percentage error (MAPE) and correlation coefficient \( R^2 \) to evaluate how effectively a given distribution combined with particular estimation method fits the observed data.

These measures are defined as:

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F(\hat{z}_i) - \hat{F}(\hat{z}_i))^2}.
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |F(\hat{z}_i) - \hat{F}(\hat{z}_i)|.
\]

\[
\text{RMSPE} = 100 \sqrt{n} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{F(\hat{z}_i) - \hat{F}(\hat{z}_i)}{F(\hat{z}_i)} \right)^2}.
\]

\[
\text{MAPE} = 100 \frac{1}{n} \sum_{i=1}^{n} \frac{|F(\hat{z}_i) - \hat{F}(\hat{z}_i)|}{F(\hat{z}_i)}.
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (F(\hat{z}_i) - \hat{F}(\hat{z}_i))^2}{\sum_{i=1}^{n} (F(\hat{z}_i) - \hat{F}(\hat{z}_i))^2}.
\]

where \( \hat{F}(\hat{z}_i) = \frac{1}{n} \sum_{i=1}^{n} F(\hat{z}_i) \) and \( n = \) sample size of the data series. In all the above accuracy measure formulas \( F(\hat{z}_i) \) represents the empirical CDF of the data (observed values) and \( \hat{F}(\hat{z}_i) \) is the theoretical CDF of the distribution (with estimated parameters by any particular method).

### 3.7 Quantile estimation

The estimate of quantile \( \tilde{z}_T \) for a return period (T) of scientific relevance is a goal of flood frequency analysis. The return period T (recurrence interval) is the period expressed in the number of years in which the annual maximum peak flow of flood \( \tilde{z}_T \) is expected to return. A given flood return level \( \tilde{z}_T \) with a return period T may be exceeded once in T years. Therefore, \( P(\tilde{Z} \geq \tilde{z}_T) = \frac{1}{T} \), then the cumulative probability of non-exceedance \( F(\tilde{z}_T) \) is

\[
F = F(\tilde{z}_T) = P(\tilde{Z} \leq \tilde{z}_T) = 1 - P(\tilde{Z} \geq \tilde{z}_T) = 1 - \frac{1}{T}.
\]

The distribution functions can be expressed in the inverse form as \( \tilde{z}_T = \phi(F) \), hence estimate of quantile \( \tilde{z}_T \) can be directly evaluated by replacing F. If distribution cannot be expressed in inverse form \( \tilde{z}_T = \phi(F) \), then the numerical method is used to evaluate \( \tilde{z}_T \) for a given value of F which is based on numerical relationships between \( \tilde{z}_T \) and F. The expressions of quantile functions of candidate distributions is given in Table 2. The T-year estimate of quantile is calculated by substituting the value of \( F = (1 - 1/T) \) in the expressions of the quantile. There is always uncertainty in estimated flood magnitude. To
quantify such uncertainty, we estimate the SE of flood estimates and construct 90% confidence intervals of flood quantiles for different return periods by using the parametric bootstrap method. We use a parametric bootstrap method to calculate SE of estimated quantiles and to construct 90% confidence interval for flood quantiles of different return periods. This parametric bootstrap method is more precise than asymptotic computation in case of small sample size (n). The details of this parametric bootstrap method are available in Meylan et al. (2012).

4 | RESULT AND DISCUSSION

The basic descriptive statistics of five gauging sites are provided in Table 3. The basic assumptions in the statistical flood frequency analysis and fitting of probability distributions are independence, randomness, stationarity and skewness of the data series (Afreen & Muhammad, 2012; Hassan, Hayat, & Noreen, 2019; Kite, 1977). Therefore, in this article Wald-Wolfowitz (WW) and Augmented Dickey-Fuller (ADF) tests are applied to test for randomness and stationarity, respectively, while independence and skewness are judged by correlation coefficient (r) at lag-1. The results concerning the testing of above assumptions are summarised in Table 4.

It comes out that annual data series of all gauging sites are independent, random, stationary, and skewed, therefore, suitable for statistical flood frequency analysis and hence fitting of candidate probability distributions.

The ML and LM parameter estimates along with SE for each distribution at five gauging sites are reported in Table 5. For distribution comparison, we have used four GOF tests and five accuracy measures. The GOF tests and the accuracy measures indicate how well the observed flood data correspond to the fitted distribution. The distributions with the estimation method are ranked according to the p-values in each GOF test for each gauging site. The total ranked score of each combination of probability distribution and estimation method in all GOF tests is given in Table 6 for each gauging site. Similarly, each distribution with the estimation method is ranked according to each accuracy measure for each site and the total ranked score is presented in Table 6. The best-fitted distribution is selected based on the highest combined total rank score in the GOF tests and accuracy measures.

Table 6 shows the overall performance of fitted distributions based on GOF tests and accuracy measures. The LN3 distribution estimated by the LM method is the most suitable distribution as compared to other distributions for gauging sites Harrsele Krv, Gardiken and Överuman Nedre, while the GEV distribution estimated by the LM method is the best-fitted distribution for gauging sites Stornorrfors Krv and Solberg.

In lines with the findings in Hassan et al. (2019) and Ahmad, Fawad, and Mahmood (2015), from the results of present analysis, a single distribution does not emerge as the best-suited distribution for all gauging sites. The sample size is not an important factor in favour of any particular distribution or estimation method. Overall, the

| Station name       | n     | Mean  | Median | Max (year) | S      | CV    | Skewness | Kurtosis |
|--------------------|-------|-------|--------|------------|--------|-------|----------|----------|
| Harrsele Krv       | 58    | 544.98| 440.00 | 1250 (1993)| 251.69 | 0.46  | 1.22     | 0.24     |
| Stornorrfors Krv   | 57    | 1363.46| 1350.00| 2380 (1993)| 390.60 | 0.29  | 0.60     | −0.05    |
| Överuman Nedre     | 52    | 78.54 | 66.75  | 189 (1993) | 37.45  | 0.48  | 1.31     | 1.10     |
| Gardiken           | 54    | 368.35| 317.00 | 870 (1961)| 116.34 | 0.32  | 2.38     | 6.69     |
| Solberg            | 104   | 219.54| 214.00 | 360 (2005)| 56.47  | 0.26  | 0.18     | −0.45    |

Note: S indicates the SD and CV represents the coefficient of variation.

| Station name       | n     | ADF (p-value) | WW (p-value) | r   | p-value |
|--------------------|-------|---------------|--------------|-----|---------|
| Harrsele Krv       | 58    | .00           | .71          | −0.05 | .59    |
| Stornorrfors Krv   | 57    | .00           | .96          | −0.14 | .11    |
| Överuman Nedre     | 52    | .00           | .08          | 0.20  | .21    |
| Gardiken           | 54    | .00           | .61          | 0.06  | .50    |
| Solberg            | 104   | .00           | .45          | 0.02  | .73    |

Table 3 Descriptive statistics (cubic meter per second) of five gauging sites of the Ume River

Table 4 Results of testing assumptions of five gauging sites' data
### TABLE 5  Estimated parameters with ML and LM methods

| Catchment | Probability distributions | GEV | LN3 | GLO | GUM |
|-----------|---------------------------|-----|-----|-----|-----|
|           | ML | μ | α    | κ   | ML | μ | α | κ | ML | μ | α | κ |
| 1733      | 396.73 (15.73) | 98.72 (14.61) | −0.65 (0.20) | 448.16 (22.13) | 163.54 (26.29) | −1.00 (0.12) | 443.20 (22.12) | 95.42 (17.54) | −0.63 (0.09) | 438.47 (19.90) | 157.33 (15.51) |
| 1734      | 1194.58 (39.61) | 329.20 (36.46) | 0.08 (0.09) | 1308.95 (56.42) | 371.02 (37.78) | −0.29 (0.10) | 1308.27 (53.29) | 216.47 (25.33) | −0.19 (0.08) | 1181.13 (49.80) | 321.15 (37.49) |
| 1856      | 58.02 (3.09) | 19.25 (2.68) | −0.40 (0.14) | 66.67 (4.07) | 27.93 (4.04) | −0.79 (0.13) | 65.97 (4.05) | 16.44 (2.75) | −0.50 (0.08) | 62.65 (3.23) | 24.48 (2.39) |
| 20,011    | 312.70 (11.31) | 44.21 (7.89) | −0.44 (0.11) | 333.60 (13.15) | 68.11 (10.10) | −0.82 (0.07) | 329.78 (10.70) | 36.77 (6.03) | −0.49 (0.04) | 324.87 (10.25) | 61.81 (7.72) |
| 436       | 198.46 (10.11) | 54.03 (7.28) | 0.23 (0.11) | 216.99 (6.29) | 55.97 (4.34) | −0.09 (0.08) | 217.00 (6.01) | 32.67 (2.84) | −0.06 (0.06) | 191.93 (9.38) | 51.41 (7.20) |

| Catchment | Probability distributions | GEV | LN3 | GLO | GUM |
|-----------|---------------------------|-----|-----|-----|-----|
|           | LM | μ | α    | κ   | LM | μ | α | κ | LM | μ | α | κ |
| 1733      | 415.39 (19.41) | 132.55 (18.15) | −0.29 (0.13) | 462.48 (38.58) | 179.14 (74.06) | −0.79 (0.15) | 470.32 (16.23) | 103.17 (16.63) | −0.37 (0.13) | 435.99 (23.88) | 188.84 (21.18) |
| 1734      | 1190.47 (53.92) | 339.05 (38.90) | 0.07 (0.10) | 1313.92 (56.25) | 380.79 (52.61) | −0.26 (0.12) | 1318.55 (33.48) | 215.19 (27.40) | −0.12 (0.11) | 1179.60 (41.82) | 318.53 (37.66) |
| 1856      | 59.42 (3.32) | 21.99 (3.05) | −0.23 (0.12) | 67.37 (5.15) | 28.91 (7.24) | −0.69 (0.16) | 68.43 (2.77) | 16.55 (2.73) | −0.50 (0.08) | 62.01 (3.99) | 28.64 (3.19) |
| 20,011    | 312.70 (11.31) | 44.21 (7.89) | −0.44 (0.11) | 333.60 (13.15) | 68.11 (10.10) | −0.82 (0.07) | 329.78 (10.70) | 36.77 (6.03) | −0.49 (0.04) | 324.87 (10.25) | 61.81 (7.72) |
| 436       | 198.46 (10.11) | 54.03 (7.28) | 0.23 (0.11) | 216.99 (6.29) | 55.97 (4.34) | −0.09 (0.08) | 217.00 (6.01) | 32.67 (2.84) | −0.06 (0.06) | 191.93 (9.38) | 51.41 (7.20) |

Note: Karrske Krv: 1733, Stornorrfors Krv: 1734, Överuman Nedre: 1856, Gardiken: 20011 and Solberg: 436. Here, \( \hat{\mu} \), \( \hat{\alpha} \), and \( \hat{\kappa} \) represent the estimated location, scale and shape parameters, respectively. The values in parenthesis are a SE of estimated parameters.
LM method is the best estimation method for the identification of a suitable distribution (also see, Ahmad et al., 2015; Hassan et al., 2019). The site having the highest average of annual maxima of flood and the catchment area is best modelled by GEV distribution while the site having the lowest average of annual maxima of flood and the catchment area is best modelled by LN3 distribution (see Tables 1, 3, and 6). The GEV distribution is more suitable for gauging sites having negative values of the coefficient of kurtosis and low values of coefficient of variation and coefficient of skewness (see Tables 3 and 6). The results give an important geographic insight as well, that is, the gauging site close to the Gulf of Bothnia is in favour of the GEV distribution while the LN3 distribution is suitable for a gauging site that is extremely away from the Gulf of Bothnia.

One of the objectives of flood frequency analysis is to estimate the quantile in the extreme upper tail of the best-fitted distribution at each gauging site. The quantile estimates for the return periods of 5, 10, 25, 50, 100, 200, 500, and 1000 years are calculated by using quantile function and parameter values of the best-fitted distribution for any particular gauging site. Quantile estimates $\xi_T$ with non-exceedance probability $F$ based on the

| Distribution | Method | (Harrsele Krv) 1733 | (Gardiken) 20011 |
|--------------|--------|---------------------|-------------------|
|              |        | GOF | AM | Total | GOF | AM | Total |
| GEV          | ML     | 23  | 11 | 34    | 20  | 35 | 55    |
| LN3          | ML     | 18  | 33 | 51    | 10  | 18 | 28    |
| GLO          | ML     | 22  | 19 | 41    | 18  | 29 | 47    |
| GUM          | ML     | 3   | 12 | 15    | 5   | 7  | 12    |
| GEV          | LM     | 12  | 28 | 40    | 21  | 26 | 47    |
| LN3          | LM     | 15  | 37 | 52    | 17  | 40 | 57    |
| GLO          | LM     | 9   | 22 | 31    | 13  | 17 | 30    |
| GUM          | LM     | 6   | 17 | 23    | 4   | 8  | 12    |

| Distribution | Method | (Stornorrfors Krv) 1734 | (Solberg) 436 |
|--------------|--------|-------------------------|---------------|
|              |        | GOF | AM | Total | GOF | AM | Total |
| GEV          | ML     | 15  | 17 | 32    | 19  | 29 | 48    |
| LN3          | ML     | 16  | 26 | 42    | 15  | 26 | 41    |
| GLO          | ML     | 7   | 8  | 15    | 11  | 18 | 29    |
| GUM          | ML     | 10  | 28 | 38    | 6   | 7  | 13    |
| GEV          | LM     | 22  | 40 | 62    | 24  | 40 | 64    |
| LN3          | LM     | 23  | 35 | 58    | 20  | 35 | 55    |
| GLO          | LM     | 10  | 7  | 17    | 10  | 17 | 27    |
| GUM          | LM     | 5   | 19 | 24    | 3   | 8  | 11    |

| Distribution | Method | (ÖVeruman Nedre) 1856 |
|--------------|--------|-----------------------|
|              |        | GOF | AM | Total |
| GEV          | ML     | 14  | 22 | 36    |
| LN3          | ML     | 21  | 37 | 58    |
| GLO          | ML     | 19  | 15 | 34    |
| GUM          | ML     | 3   | 8  | 11    |
| GEV          | LM     | 13  | 28 | 41    |
| LN3          | LM     | 23  | 38 | 61    |
| GLO          | LM     | 9   | 20 | 29    |
| GUM          | LM     | 6   | 12 | 18    |

Note: The bold values are rank score for best-fitted distribution.

TABLE 6  The rank score of distributions based on GOF tests and accuracy measures
### TABLE 7 Quantile estimates of the flood with 90% confidence interval for five gauging sites of the Ume River

| Gauging site | Distribution (method) | Value | Non-exceedance probability F and return periods (years) | 0.8 | 0.9 | 0.96 | 0.98 | 0.99 | 0.995 | 0.998 | 0.9999 |
|--------------|-----------------------|-------|--------------------------------------------------------|-----|-----|------|------|-----|-------|-------|--------|
|              |                       |       | 5 years | 10 years | 25 years | 50 years | 100 years | 200 years | 500 years | 1000 years |
| 1734         | GEV (LM)              | Upper | 1792.68  | 2053.44  | 2401.40  | 2679.17  | 2974.69   | 3287.25   | 3726.79   | 4090.87   |
|              |                       | Fit   | 1672.52  | 1894.84  | 2159.10  | 2343.82  | 2518.14   | 2683.33   | 2888.98   | 3035.64   |
|              |                       | Lower | 1550.01  | 1739.41  | 1934.07  | 2051.44  | 2146.71   | 2225.71   | 2313.85   | 2366.97   |
|              |                       | σ_x   | 73.18    | 93.85    | 140.17   | 190.16   | 252.20    | 325.72    | 440.24    | 540.19    |
| 1733         | LN3 (LM)              | Upper | 767.79   | 1013.46  | 1451.72  | 1884.85  | 2418.09   | 3072.27   | 4154.99   | 5147.50   |
|              |                       | Fit   | 676.20   | 858.63   | 1136.91  | 1379.62  | 1653.23   | 1960.57   | 2423.55   | 2820.60   |
|              |                       | Lower | 592.66   | 722.10   | 890.29   | 1020.29  | 1155.07   | 1292.01   | 1481.33   | 1633.02   |
|              |                       | σ_x   | 53.10    | 89.72    | 173.71   | 269.67   | 398.93    | 567.49    | 862.95    | 1152.48   |
| 20,011       | LN3 (LM)              | Upper | 448.91   | 564.06   | 796.03   | 1048.47  | 1381.59   | 1825.85   | 2608.65   | 3394.75   |
|              |                       | Fit   | 409.21   | 490.172  | 627.79   | 759.01   | 917.33    | 1106.48   | 1410.85   | 1688.03   |
|              |                       | Lower | 373.16   | 424.98   | 498.58   | 558.64   | 624.19    | 694.04    | 799.05    | 880.90    |
|              |                       | σ_x   | 23.40    | 43.34    | 94.03    | 158.35   | 253.43    | 389.47    | 655.63    | 946.25    |
| 436          | GEV (LM)              | Upper | 279.72   | 308.77   | 342.26   | 365.56   | 386.92    | 406.88    | 431.07    | 448.43    |
|              |                       | Fit   | 268.1    | 295.46   | 324.16   | 341.90   | 356.99    | 369.89    | 384.15    | 393.17    |
|              |                       | Lower | 256.69   | 282.18   | 306.28   | 319.37   | 329.39    | 337.53    | 345.47    | 349.89    |
|              |                       | σ_x   | 7.04     | 8.14     | 10.99    | 14.01    | 17.48     | 21.22     | 26.34     | 30.25     |
| 1856         | LN3 (LM)              | Upper | 114.50   | 149.52   | 208.35   | 262.73   | 329.02    | 405.82    | 530.33    | 638.20    |
|              |                       | Fit   | 100.30   | 126.74   | 165.28   | 197.64   | 233.10    | 271.90    | 328.75    | 376.28    |
|              |                       | Lower | 87.08    | 105.70   | 129.14   | 146.87   | 163.90    | 181.27    | 205.40    | 223.71    |
|              |                       | σ_x   | 8.38     | 13.39    | 24.29    | 36.22    | 51.71     | 71.22     | 104.12    | 135.19    |

Note: The Fit values (in bold) are estimated flood quantiles for different return periods while σ_x represents their SE. The Lower and Upper indicate the lower and upper limits of 90% confidence interval of flood quantiles.
best-suited combination of probability distribution and estimation method are given in Table 7 for each gauging site. The uncertainty measure ($\sigma_s$) of quantile estimates and 90% confidence interval of flood magnitude for different return periods are given in Table 7 and graphically represented in Figure 3. The SE of quantile estimates indicates that the longer return periods have more uncertainty in the estimates of flood magnitude.

5 | CONCLUSIONS

Flood frequency analysis of five gauging sites of the Ume River is carried out by using LN3, GEV, GLO, and GUM distributions. The ML and LM estimation methods are used to estimate the parameters of the distributions. To identify the best-suited combination of probability distribution and estimation method, we have assigned the rank to each combination in each GOF test and accuracy measure. We have selected the best suitable distribution at each gauging site that has the highest combined rank score. For gauging sites Stornorrfors Krv and Solberg, the GEV estimated by the LM method is found the best-fitted distribution. For gauging sites Gardiken, Harrsele Krv and Överuman Nedre, the LN3 with the LM estimation method is found to be the most suitable distribution. The results of flood frequency analysis of the Ume River could be used for the study of flood, water resource management and designing of hydraulic structure in the same basin and catchment area. Because of the close distributional fit at each site, the GEV and LN3 distributions with the LM method are better choices for flood forecasting and could be the best candidate distributions for a regional flood frequency analysis of the Ume River. The findings from this study would be useful for at-site flood frequency analyses on other rivers of Sweden.

ACKNOWLEDGEMENTS

Dr. Shahid Mehmood (University of Oxford, UK) is acknowledged with thanks for proofreading the manuscript.
DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Swedish Meteorological and Hydrological Institute at https://vattenwebb.smhi.se/station/

ORCID
Mahmood Ul Hassan https://orcid.org/0000-0003-2889-0263
Sajjad Haider Bhatti https://orcid.org/0000-0002-3206-4565

REFERENCES

Afreen, S., & Muhammad, F. (2012). Flood frequency analysis of various dams and barrages in Pakistan. Irrigation and Drainage, 61(1), 116–128.

Ahmad, I., Fawad, M., & Mahmood, I. (2015). At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. Polish Journal of Environmental Studies, 24(6), 2345–2353.

Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., … MacDonald, N. (2012). Review of applied statistical methods for flood frequency analysis in Europe, Milestone Report for WG2 of COST Action ES0901. Tech. rep., Centre for Ecology & Hydrology, Wallingford, UK, on behalf of COST, Available from http://www.cost.eu/media/publications.

Cunnane, C. (1989). Statistical distributions for flood frequency analysis. Operational Hydrology Report (WMO). World Meteorological Organization Operational Hydrology Report No. 33.

Guru, N., & Jha, R. (2015). Flood frequency analysis of Tel Basin of Mahanadi River system, India using annual maximum and POT flood data. Aquatic Procedia, 4, 427–434.

Hamed, K., & Rao, A. R. (1999). Flood frequency analysis, Boca Raton, FL: CRC press.

Hassan, M. U., Hayat, O., & Noreen, Z. (2019). Selecting the best probability distribution for at-site flood frequency analysis: A study of Torne River. SN Applied Sciences, 1(12), 1629–1638.

Hosking, J R M. (1986). The theory of probability weighted moments. IBM Research Division, TJ Watson Research Center.

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 B: Methodological, 52, 105–124.

Hosking, J. R. M., & Wallis, J. R. (2005). Regional frequency analysis: An approach based on L-moments, Cambridge, UK: Cambridge University Press.

Kite, G. W. (1977). Frequency and risk analyses in hydrology. Water Resources Publications, 303, https://books.google.se/books?id=WZt-AAAAIAAJ

Meylan, P., Favre, A.-C., & Musy, A. (2012). Predictive hydrology: A frequency analysis approach, Boca Raton, FL: CRC Press.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. The Computer Journal, 7(4), 308–313.

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Alsans, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. Natural Hazards, 69(3), 1803–1813.

Saf, B. (2009). Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. Water Resources Management, 23(3), 531–551.

The Swedish Meteorological and Hydrological Institute (SMHI). Data files retrieved from SMHI webpage: https://vattenwebb.smhi.se/station/

Vogel, R. M., & Wilson, I. (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the United States. Journal of Hydrologic Engineering, 1(2), 69–76.

How to cite this article: Kousar S, Khan AR, Ul Hassan M, Noreen Z, Bhatti SH. Some best-fit probability distributions for at-site flood frequency analysis of the Ume River. J Flood Risk Management. 2020;13:e12640. https://doi.org/10.1111/jfr3.12640