DOI:10.28974/idojaras.2025.2.2

IDŐJÁRÁS

Quarterly Journal of the HungaroMet Hungarian Meteorological Service
Vol. 129, No. 2, April – June, 2025, pp. 133–160

Five-parameter log-normal distribution and its modification

Marie Budíková[1,*], **Jan Holub**[2], **Ladislav Budík**[3], **Lenka Přibylová**[1], and **Ivana Horová**[1]

[1]*Department of Mathematics and Statistics, Faculty of Science, Masaryk University*
Kotlářská 2, Brno, 611 37, Czech Republic

[2]*Global Change Research Institute of the Czech Academy of Sciences*
Bělidla 986/4a, Brno, 603 00, Czech Republic

[3]*Czech Hydrometeorological Institute*
Kroftova 2578/43, Brno, 616 67, Czech Republic

* *Corresponding Author e-mail: budikova@math.muni.cz*

(Manuscript received in final form May 6, 2024)

Abstract— The introduction of the LN5 and mLN5 distributions extends the commonly used three-parameter log-normal distribution (LN3) by enhancing tail modeling, which is critical for accurate representation of extreme values in hydrology and climatology. This paper details two methods for parameter estimation: the established local maximum likelihood method and the newly developed triangular method, an adaptation of the relative least squares approach. The effectiveness of these distributions is demonstrated through their application to datasets from the Czech Hydrometeorological Institute, encompassing average daily flow, precipitation, atmospheric pressure, and air temperature. Results show significant improvements in modeling extreme events with LN5 and mLN5 over LN3, as well as over other compared distributions such as generalized Gamma and generalized Weibull, particularly in tail behavior, underscoring their potential for advancing environmental studies. Appendices include comprehensive derivations of the functional characteristics of LN5 and mLN5 and introduce an alternative parametrization for LN5.

Key-words: exceedance curve, five-parameter log-normal distribution, maximum likelihood estimate, modified five-parameter log-normal distribution, triangular method

Highlights:
Introduces LN5 distribution and its modification and derives their characteristics.
Applies LN5 to data, compares them with existing models with superior performance.
Enhances approximation of distribution tails to better represent extreme values.

1. Introduction

Exceedance curves are essential in analyzing hydrological and climatological data, offering insights into the likelihood of surpassing specific values, crucial for developing empirical or theoretical models. These curves, essentially inverse survival functions, often employ the three-parameter log-normal distribution (LN3) (*Sangal* and *Biswas*, 1970), Weibull or gamma distributions. Also more flexible alternatives to these classical choices as generalized Weibull distribution (*Mudholkar et al*, 1996), generalized gamma distribution (*Cox et al.*, 2007), or two-piece distributions (*Rubio* and *Hong*, 2016) were suggested in literature. However, these distributions may not accurately represent all data ranges, especially in small drainage basins, and can lead to unrealistic extrapolations at extreme probabilities (*Budík*, 2018). We propose the five-parameter log-normal distribution (LN5) and its modified version (mLN5) as superior alternatives, providing better fits and more accurate extrapolations (*Budík*, 2018, 2019). Contributions, briefly suggested LN5 and mLN5 distributions, omitting the derivation of their functional characteristics. This paper details the functional characteristics of LN5 and mLN5 and their application to real-world data and shows the advantages of LN5 and mLN5 distributions compared to the aforementioned distributions used in hydrological and climatological practice.

The analysis of hydrological and precipitation data shows discrepancies between the theoretical curves and actual measurements for the different distributions mentioned above, especially for the LN3 distribution, which is commonly used in hydrological and climatological practice. Accurate midrange estimation is crucial, but with climate change causing shifts towards extreme events, estimating extremes becomes equally important. The LN5 distribution, based on our experience, effectively addresses these issues, offering several advantages:

a) Near-accurate modeling of exceedance curves and quality extrapolations for large datasets across a range of probabilities, confirmed by a simulation study (*Budík* and *Budíková*, 2020).
b) Enhanced modeling of extremely small and large values, crucial for estimating probabilities of significant climatic and hydrological events in the context of climate change.
c) Greater flexibility in modeling exceedance curves, allowing for precise differentiation of regional climatic or hydrological characteristics.
d) Ability to detect some primary data processing inaccuracies in hydrological data application.
e) When applied to long-term climatological and hydrological data, LN5 or mLN5 parameters are interpretable and can reveal changes in these quantities, correlating with climate trends and landscape drying.

This paper presents the LN5 and mLN5 distributions' density and distribution functions, parameter estimation methods, and applications to real data. We also describe the triangulation method with inverse transformation, a robust estimation approach for natural process-generated data. Previously, LN5, mLN5, and the triangulation method were only suggested for one dataset in conference proceedings; this paper provides a theoretical foundation and shows its usefulness in broader context, specifically on applications on hydrology and climatology data.

The paper is structured as follows: The Methods section discusses the exceedance curve and its relationship to the survival function. The Theory section introduces the LN5 and mLN5 distributions and covers methods for parameter estimation. The Results section applies these findings to real data. The paper concludes with a discussion on the approach's advantages and limitations in the final section. The appendices include theoretical derivations, an explanation of alternative parameterizations, graphs of PDF and CDF functions, a comparison of LMLE estimates for the Morava and the Dyje Rivers, and results from a prior simulation study.

2. Methods

In practice, the empirical exceedance curve, theoretical exceedance curve (*Lane*, 2002), and survival function (see *Fig. 1*) are key concepts for analyzing the probability of an observed variable surpassing a certain threshold, commonly used in studying extreme events like floods, earthquakes, or financial market crashes. The empirical exceedance curve plots the descending values of a variable (on the vertical axis) against the estimated probabilities of exceeding these values (on the horizontal axis). The probability of exceeding a given threshold is estimated as the relative frequency of data points above each threshold. The theoretical exceedance curve, derived from a probability distribution model, arranges the quantiles of the chosen probability distribution in descending order on the vertical axis, and the corresponding exceedance probabilities are plotted on the horizontal axis. The theoretical exceedance curve can be used to estimate the probability of outliers beyond the observed data range. The survival function, inversely related to the theoretical exceedance curve, indicates the probability that the variable's realization will exceed a specific value. It is crucial in assessing survival probabilities or durations across various fields.

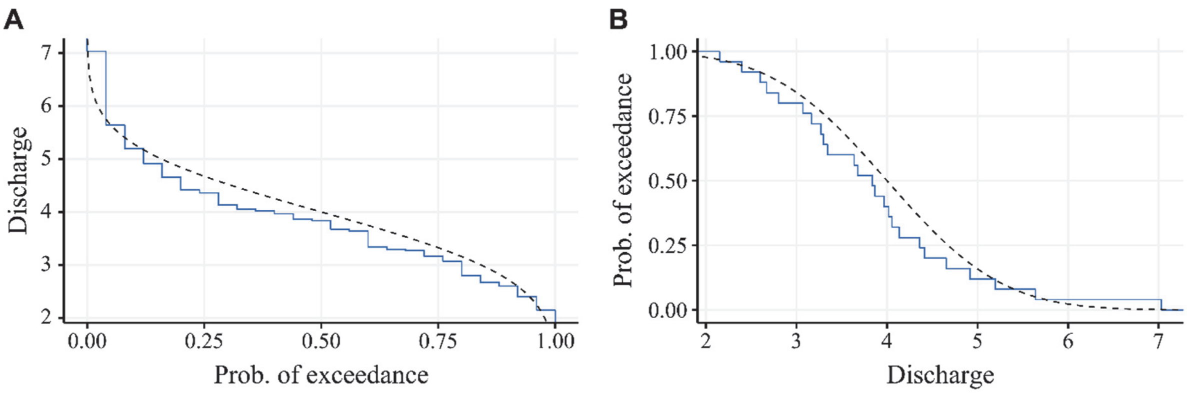

Fig. 1. **A**. The empirical exceedance curve (blue solid line) and the theoretical exceedance curve (black dashed line). **B**. The empirical survival function (blue solid line) and the theoretical survival function (black dashed line).

In summary, while the empirical exceedance curve is based on the measured data, the theoretical exceedance curve is based on a model of the underlying probability distribution, and the survival function is the inverse of the exceedance curve. Obtaining the best theoretical exceedance curve is necessary to estimate the magnitude of extreme events at a given probability of exceedance. To advance this modeling, we introduce the LN5 and mLN5 distributions for constructing exceedance curves. The upcoming Theory section will detail new results regarding the characteristics and parameter estimation methods of these distributions. Subsequently, in the Results section, we will demonstrate the application of these distributions to hydrology and climatology data.

3. Theory

In this section, we derive new LN5 distributions as generalizations of the three-parameter log-normal distribution LN3 (*Sangal and Biswas*, 1970).

3.1. Fundamental five-parameter log-normal distribution

Let $X \sim \mathrm{N}(\mu, \sigma^2)$ be a normally distributed random variable with mean μ and variance σ^2. Let a, $b \in \mathbb{R}^+$, $y_0 \in \mathbb{R}$. Random variable Y defined by the transformation

$$Y = a \exp(\mathrm{sgn}X \cdot |X|^b) + y_0, \tag{1}$$

which follows the five-parameter log-normal distribution, i.e., $Y \sim \mathrm{LN5}(a, b, \mu, \sigma^2, y_0)$, and the parameter vector is denoted by $\boldsymbol{\theta} = (a, b, \mu, \sigma^2, y_0)$.

It is important to recognize that the parameters influence the shape of the distribution function's graph. The location parameter y_0 shifts the distribution. Parameter a, shape parameter in general, is a parameter of scale if $y_0 = 0$. The remaining parameters b, μ, σ^2 are shape parameters. In particular, μ and σ^2 correspond to the mean and variance of an inversely transformed random variable

$$\mathrm{sgn}(Y - a - y_0)\ln^{\frac{1}{b}}\left(((Y - y_0)/a)^{\mathrm{sgn}(Y-a-y_0)}\right).$$

The probability density, cumulative distribution, and quantile functions of the five-parameter log-normal distribution take the following form.

Probability density function

$$f(y, \boldsymbol{\theta}) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \cdot \dfrac{\ln^{\frac{1-b}{b}}\left(\frac{a}{y-y_0}\right)}{b(y-y_0)} \cdot \exp\left\{-\dfrac{\left(-\ln^{\frac{1}{b}}\left(\frac{a}{y-y_0}\right) - \mu\right)^2}{2\sigma^2}\right\}, & y \in (y_0, y_0 + a), \\ \dfrac{1}{\sqrt{2\pi}\sigma} \cdot \dfrac{\ln^{\frac{1-b}{b}}\left(\frac{y-y_0}{a}\right)}{b(y-y_0)} \cdot \exp\left\{-\dfrac{\left(\ln^{\frac{1}{b}}\left(\frac{y-y_0}{a}\right) - \mu\right)^2}{2\sigma^2}\right\}, & y \in (y_0 + a, \infty), \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Cumulative distribution function

$$F(y, \boldsymbol{\theta}) = \begin{cases} 0, & y \in (-\infty,\ y_0), \\ \dfrac{1}{2}\left[1 + \mathrm{erf}\left(\dfrac{-\ln^{\frac{1}{b}}\left(\frac{a}{y-y_0}\right) - \mu}{\sqrt{2}\sigma}\right)\right], & y \in [y_0\,, y_0 + a), \\ \dfrac{1}{2}\left[1 + \mathrm{erf}\left(\dfrac{\ln^{\frac{1}{b}}\left(\frac{y-y_0}{a}\right) - \mu}{\sqrt{2}\sigma}\right)\right], & y \in\ [y_0 + a, \infty). \end{cases}$$

Quantile function

$$F^{-1}(\alpha,\boldsymbol{\theta}) = \begin{cases} a\exp\{-[-\mu-\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)]^{b}\}+y_0, & \alpha\in I_1,\\ a\exp\{[\mu+\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)]^{b}\}+y_0, & \alpha\in I_2,\end{cases}$$

where

$$I_1 = \left(0,\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)\right]\right),$$
$$I_2 = \left(\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)\right],1\right),$$

erf is the error function and derivation of Eqs.(2)–(4) and further details on $f(y,\boldsymbol{\theta})$, $F(y,\boldsymbol{\theta})$, and $F^{-1}(\alpha,\boldsymbol{\theta})$ are listed in Appendix A of the Supplementary material, while probability density functions and cumulative distribution functions for various parameters are depicted in Appendix D of the Supplementary material.

3.2. Modified five-parameter log-normal distribution (mLN5)

The density of the LN5 distribution with parameter $b \neq 1$ is not a smooth function, which often does not align with the nature of the data. To address this issue, we have introduced a novel version of the LN5 distribution, the mLN5, which is achieved by modifying the transformation Eq.(1). Additional details can be found in Appendix B of the Supplementary material, while probability density functions and cumulative distribution functions for various parameters are depicted in Appendix D of the Supplementary material.

Let $X \sim \mathrm{N}(\mu,\sigma^2)$ be a normally distributed random variable. Let $a, b \in \mathbb{R}^+$, and $y_0 \in \mathbb{R}$. Random variable Y, a transformation of X in the form

$$Y = \begin{cases} a\exp\{\mathrm{sgn}X\cdot|X|^{b}\}+y_0, & |X|\geq 1,\\ a\exp\{\mathrm{sgn}X\cdot|X|^{b+(1-b)(1-|X|)}\}+y_0, & |X|<1,\end{cases} \tag{3}$$

follows the modified five-parameter log-normal distribution, i.e. $\mathrm{mLN5}(a,b,\mu,\sigma^2,y_0)$, and the parameter vector is denoted by $\boldsymbol{\theta} = (a,b,\mu,\sigma^2,y_0)$.

We define function t that describes transformation Eq.(3) on open sets

$$
t(x) = \begin{cases} t_1(x) = a\exp\{-(-x)^b\} + y_0, & x \in G_1 = (-\infty, -1), \\ t_2(x) = a\exp\{-(-x)^{b+(1-b)(1+x)}\} + y_0, & x \in G_2 = (-1,0), \\ t_3(x) = a\exp\{x^{b+(1-b)(1-x)}\} + y_0, & x \in G_3 = (0,1), \\ t_4(x) = a\exp\{x^b\} + y_0, & x \in G_4 = (1,\infty). \end{cases}
$$

Let's denote the image of function t on a given set G_j as $H_j = t(G_j), j = 1, \dots, 4$ and define a function τ_j as an inverse function to the t_j on H_j for $j = 1, \dots, 4$,

$$
\tau_1(y) = -\ln^{\frac{1}{b}}\left(\frac{a}{y - y_0}\right), \quad y \in H_1 = (y_0,\ a\mathrm{e}^{-1} + y_0)
$$

$$
\tau_4(y) = \ln^{\frac{1}{b}}\left(\frac{y - y_0}{a}\right), \quad y \in H_4 = (a\mathrm{e} + y_0,\ \infty),
$$

where $x = \tau_2(y)$ is the solution of equation $y = t_2(x)$ for $y \in H_2 = (a\mathrm{e}^{-1} + y_0,\ a + y_0)$. Similarly, $x = \tau_3(y)$ is the solution of equation $y = t_3(x)$ for $y \in H_3 = (a + y_0,\ a\mathrm{e} + y_0)$. Let $\tau'_j(y)$ denote derivatives of functions τ_j with respect to y for $j = 1, \dots, 4$.

The probability density, cumulative distribution, and quantile functions of the five-parameter log-normal distribution take the following form.

Probability density function

$$
f(y, \boldsymbol{\theta}) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\ln^{\frac{1}{b}}\left(\frac{a}{y - y_0}\right) + \mu\right]^2\right\}|\tau'_1(y)|, & y \in (y_0,\ a\mathrm{e}^{-1} + y_0) \\ \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}[\tau_2(y) - \mu]^2\right\}|\tau'_2(y)|, & y \in (\ a\mathrm{e}^{-1} + y_0, a + y_0), \\ \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}[\tau_3(y) - \mu]^2\right\}|\tau'_3(y)|, & y \in (a + y_0, a\mathrm{e} + y_0), \\ \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\ln^{\frac{1}{b}}\left(\frac{y - y_0}{a}\right) - \mu\right]^2\right\}|\tau'_4(y)|, & y \in (a\mathrm{e} + y_0, \infty), \\ 0, & \text{otherwise.} \end{cases} \tag{4}
$$

Cumulative distribution function

$$F(y,\boldsymbol{\theta}) = \begin{cases} 0, & y \in (-\infty,\ y_0), \\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\ln^{\frac{1}{b}}\left(\frac{a}{y-y_0}\right)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [y_0,\ a\mathrm{e}^{-1}+y_0), \\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\tau_2(y)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [a\mathrm{e}^{-1}+y_0, a+y_0), \\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\tau_3(y)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [a+y_0, a\mathrm{e}+y_0), \\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\ln^{\frac{1}{b}}\left(\frac{y-y_0}{a}\right)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [a\mathrm{e}+y_0, \infty). \end{cases} \tag{5}$$

Quantile function

$$F^{-1}(\alpha,\boldsymbol{\theta}) = \begin{cases} a\exp\{-[-\mu-\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)]^b\}+y_0, & \alpha \in I_1, \\ t_2\left(\mu+\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)\right), & \alpha \in I_2, \\ t_3\left(\mu+\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)\right), & \alpha \in I_3, \\ a\exp\{[\mu+\sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)]^b\}+y_0, & \alpha \in I_4, \end{cases}$$

where

$$I_1 = \left(0, \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-1-\mu}{\sqrt{2}\sigma}\right)\right]\right),$$
$$I_2 = \left(\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-1-\mu}{\sqrt{2}\sigma}\right)\right], \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)\right]\right),$$
$$I_3 = \left(\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)\right], \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{1-\mu}{\sqrt{2}\sigma}\right)\right]\right),$$
$$I_4 = \left(\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{1-\mu}{\sqrt{2}\sigma}\right)\right], 1\right).$$

3.3. Parameter estimation

Estimates of the unknown parameters of the chosen probability distribution are commonly obtained by the method of moments, the method of maximum likelihood, or the method of relative least squares (*Cohen*, 1951; *Johnson et al*, 1994). Here, we describe two possible methods for five-parameter log-normal distributions. Similar transformations define both discussed distributions. Hence, we describe methods of parameter estimation simultaneously. The idea of the first method is to minimize a specific loss function. The second method is based on maximum likelihood estimation.

Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be a random sample from the LN5 distribution or the mLN5 distribution and $\boldsymbol{y} = (y_1, \dots, y_n)$ be the realization of this random sample. The LN5 and mLN5 distributions have parameters given by vector $\boldsymbol{\theta} = (a, b, \mu, \sigma^2, y_0)$, and $\widehat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{\mu}, \hat{\sigma}^2, \widehat{y_0})$ denotes the vector of estimated parameters.

3.3.1. Triangular method

The triangular method (see *Fig. 2*) is motivated by minimizing the difference between theoretical and empirical cumulative distribution functions over both probability and observed values. To accommodate for the non-symmetry of log-normal distribution, observed values are inversely transformed to normal distribution.

Estimated parameters $\widehat{\boldsymbol{\theta}}$ minimize statistic K, i.e. $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} K(\mathbf{y}, \boldsymbol{\theta})$,

$$K(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{u_i - u_{theor,i}}{u_{theor,i}} \right)^2 + \sum_{i=1}^{n} \left(p_i - p_{theor,i} \right)^2. \tag{6}$$

Here, $p_i = \frac{i}{n+1}$ is a empirical probability; $p_{theor,i} = F(y_i, \boldsymbol{\theta})$ is a transformation of y_i to range (0,1), where $F(y, \boldsymbol{\theta})$ is the probability distribution function of the LN5 or the mLN5 distribution; any function $u_i = \frac{1}{\sigma}(\tau(y_i, a, b, y_0) - \mu)$ is an inverse transformation of observed values y_i to standardized normal distribution, where τ is an inverse function to the function t; $u_{theor,i} = \Phi^{-1}(p_i)$ is transformation of p_i, where Φ is the cumulative distribution function of standardized normal distribution. The first term in K statistic Eq.(6) belongs to the relative least squares method, and the second term belongs to the probability optimization method. The triangular method described here differs from the one proposed in (*Budík*, 2019). In practice, it has been shown that using the value $u_{theor,i}$ in the denominator instead of u_i (see the method of relative least squares) leads to a more accurate estimation of the parameters.

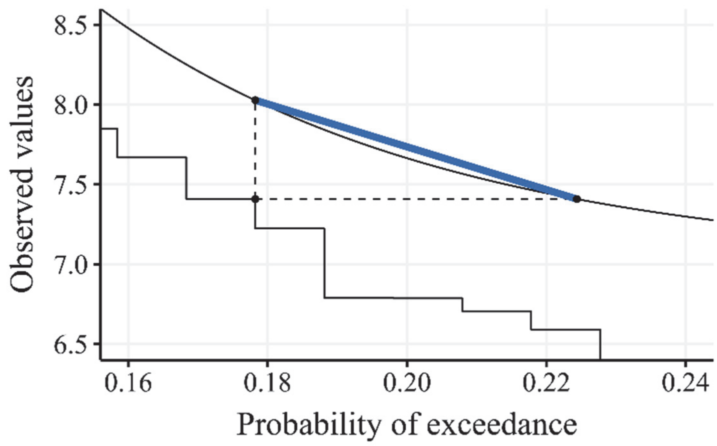

Fig. 2. Illustration of the triangular method.

3.3.2. Local maximum likelihood method

Maximum likelihood estimate of parameters $\boldsymbol{\theta}$ is a vector $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \left(\hat{a}, \hat{b}, \hat{\mu}, \hat{\sigma}^2, \widehat{y_0}\right)$, that satisfies $\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}|\mathbf{Y}\right) \geq \mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}), \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, for a given random sample $\mathbf{Y}$ and the likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{i=1}^{n} f(Y_i, \boldsymbol{\theta})$. We refer to $l(\boldsymbol{\theta}|\mathbf{Y}) = ln\, \mathcal{L}(\boldsymbol{\theta}|\mathbf{Y})$ as the log-likelihood function.

Likelihood function for a observation of a random sample $\boldsymbol{y} = (y_1, \dots, y_n)^{\top}$ is given by

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{\substack{i=1 \\ y_i \in H}}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[\tau_j(y_i) - \mu\right]^2\right\} |\tau'_j(y_i)|, \qquad j = 1, \dots, J,$$

where $H = \bigcup_{j=1}^{J} H_j$, H_j are supports of the density function f from Eqs.(2) and (4), $J = 2$ for the LN5 distribution and $J = 4$ for the mLN5 distribution.

It was shown that under certain conditions log-likelihood function of LN3 distribution approaches infinity (*Cohen*, 1951). One of the possible parameter estimation methods is the local maximum likelihood estimate (LMLE). Convergence to a local maximum of the likelihood function has been shown for LN3 distribution (*Wingo*, 1975; *Griffiths*, 1980). As the LN5 and mLN5 distributions are generalizations of the LN3 distribution, the log-likelihood function $l(\boldsymbol{\theta}|\boldsymbol{y})$ approaches infinity when $y_0 \to \min(y_1, \dots, y_n)$ for both LN5 and mLN5 distributions. To find a reasonable parameter estimate, it is necessary to assume that $y_0 < \min(y_1, \dots, y_n)$. Then log-likelihood function takes the form

$$l(\boldsymbol{\theta}|\boldsymbol{y}) = -n\, ln\sqrt{2\pi} - n\, ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\tau_j(y_i) - \mu\right]^2 + \sum_{i=1}^{n} ln\left|\tau'_j(y_i)\right|, \quad (7)$$

for $j = 1, \dots, J$. The local maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{LMLE}}$ is the solution of a system of equations

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_p} = 0, \qquad p = 1, \dots, 5, \quad \text{where} \quad \theta_p \in \boldsymbol{\theta} = (a, b, \mu, \sigma^2, y_0).$$

This system does not yield any analytical solution for all parameters. However, the estimates of parameters μ and σ^2 can be expressed as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\tau_j\,(y_i), \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\tau_j(y_i) - \mu\right]^2, \quad j = 1, \dots, J, \qquad (8)$$

which imply a possibility to obtain profile log-likelihood (*Sprott*, 2000) from Eqs. (7) and (8) as follows:

$$l\big((a, b, \mu)|\boldsymbol{y}\big) = -n\, ln\sqrt{2\pi} - n\, ln\hat{\sigma} - \tfrac{n}{2} + \textstyle\sum_{i=1}^{n} ln\left|\tau'_j(y_i)\right|, j = 1, \dots, J, \quad (9)$$

and estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{LMLE}}$ as a maximization of expression Eq.(9).

4. Results

4.1. Application in hydrology

To illustrate the capabilities of the LN5 and mLN5 distributions, we apply the procedure described above to the data the Czech Hydrometeorological Institute provided. They are two datasets; the first consists of $n = 36160$ observations of the average daily flow ($m^3 s^{-1}$) of the Morava River at the Kroměříž station, Czech Republic, the second set contains $n = 29190$ observations of the average daily flow ($m^3 s^{-1}$) of the Dyje River at the Podhradí station, Czech Republic. The underlying process is time-based; however, long-term behavior is often studied in hydrological practice, and it is commonly assumed that data are independent (*Sangal* and *Biswas*, 1970). Hence, we will consider provided data as a random sample. Discharge measurements are not exact, and provided data are rounded to two decimal places.

We compare the five-parameter log-normal distribution LN5 and its modification mLN5 with generalized gamma distribution (*Cox et al.*, 2007), generalized Weibull distribution (*Mudholkar et al*, 1996), Cauchy two-piece distributions (*Rubio* and *Hong*, 2016) and the three-parameter variant LN3 (*Sangal* and *Biswas*, 1970). We estimate the parameters using the local maximum likelihood method with optimization by Nelder-Mead method, (*Millard*, 2013). Estimates of parameters for log-normal distributions for the Morava and the Dyje Rivers are given in *Table 1*.

The log-likelihood for the distribution of mLN5 is the highest, and the log-likelihood for the distribution of LN3 is the lowest. Moreover, all distributions are compared using Akaike information criteria listed in *Table 2*.

Table 1. LMLE parameter estimates and log-likelihood for the random sample of daily average discharge of the Morava and the Dyje Rivers for LN3, LN5, and mLN5 distributions.

		$\hat{a}$	$\hat{b}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\widehat{y_0}$	l
Morava	LN3	–	–	3.4550	0.8581	2.3797	-173 479.0
	LN5	98.4527	0.9610	-1.1396	0.8878	2.0099	-173 454.5
	mLN5	91.2712	0.9360	-1.0630	0.8983	1.7339	-173 439.9
Dyje	LN3	–	–	1.6497	0.8599	0.0928	-87 366.5
	LN5	2.5375	1.1975	0.7245	0.6311	-0.0874	-86 671.71
	mLN5	2.3037	1.2605	0.7673	0.6108	0.0199	-86 655.36

Table 2. Akaike information criteria (AIC) of the generalized gamma, generalized Weibull, Cauchy two-piece, LN3, LN5 and mLN5 distributions for the random sample of daily average discharge of the Morava and the Dyje Rivers. The lowest value is in bold.

	GenGamma	GenWeibull	Two-Piece	LN3	LN5	mLN5
Morava	346 985.0	348 146.1	350 879.4	346 964.0	346 919.0	**346 889.8**
Dyje	173 738.3	173 588.5	175 270.2	174 739.1	173 353.4	**173 320.7**

Figs. 3 and *4* show the relative differences between empirical and estimated exceedance curves. See Appendix E of the Supplementary material for a comparison of histograms with density estimates. The analysis of the relative errors of the exceedance shows that, of the tested distributions, the mLN5 distribution is able to most accurately model both the middle part and the tails of the exceedance curve, which is very important when describing extreme

hydrological and climatological events. Deviations on the left side of the curve are apparently caused by extreme values of flow rates, which do not correspond to the expected course of the curve due to the length of the observed period. At the right end of the curve, deviations are probably associated with measurement errors of low flows.

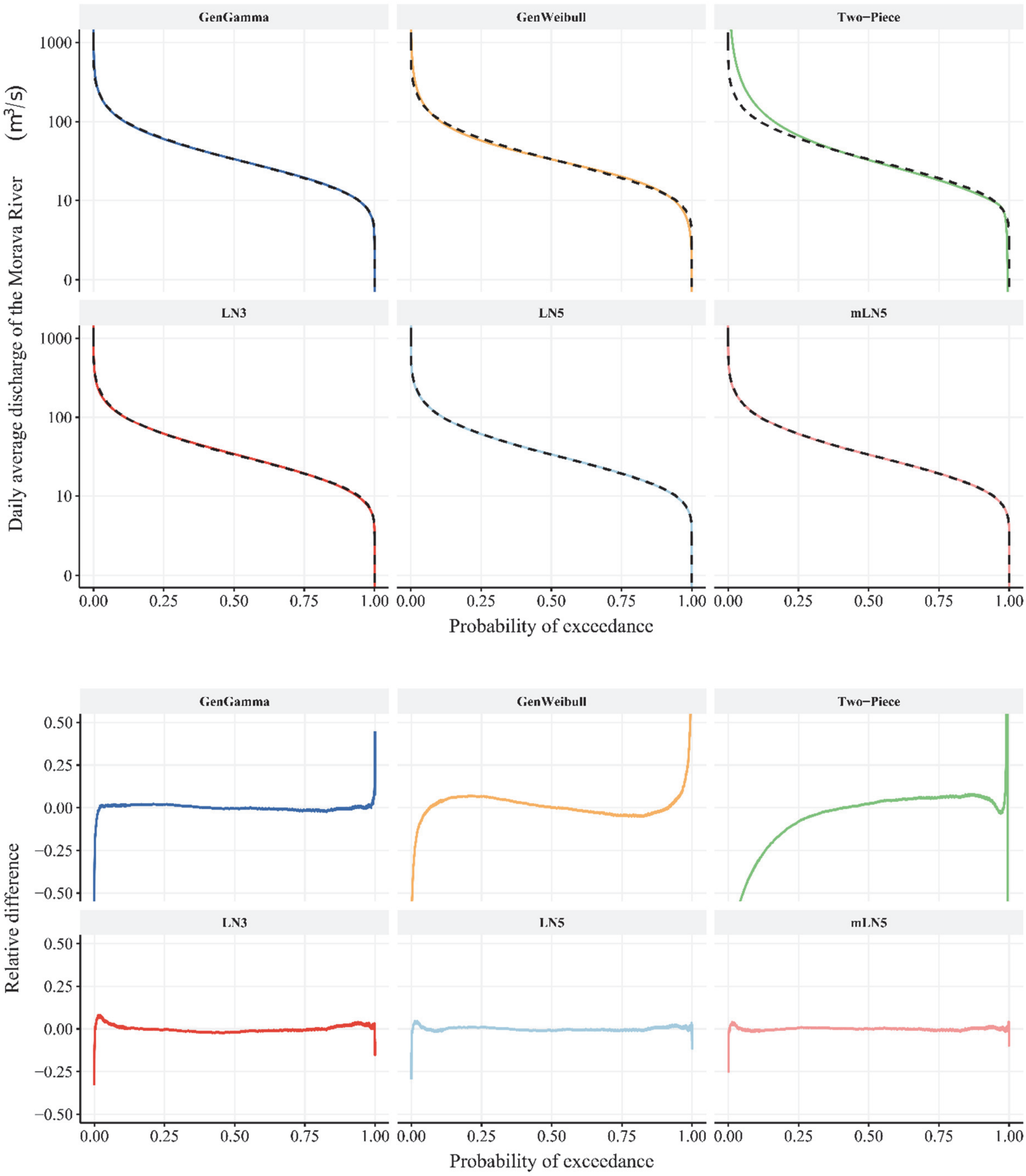

Fig. 3. Exceedance curves for the Morava River. Upper: The empirical exceedance curve (black dashed line) and fitted exceedance curves for the generalized gamma (blue line), generalized Weibull (orange line), Cauchy two-piece (green line), LN3 (red line), LN5 (light blue line) and mLN5 (pink line) distributions. Lower: Relative difference of estimated and empirical exceedance curves.

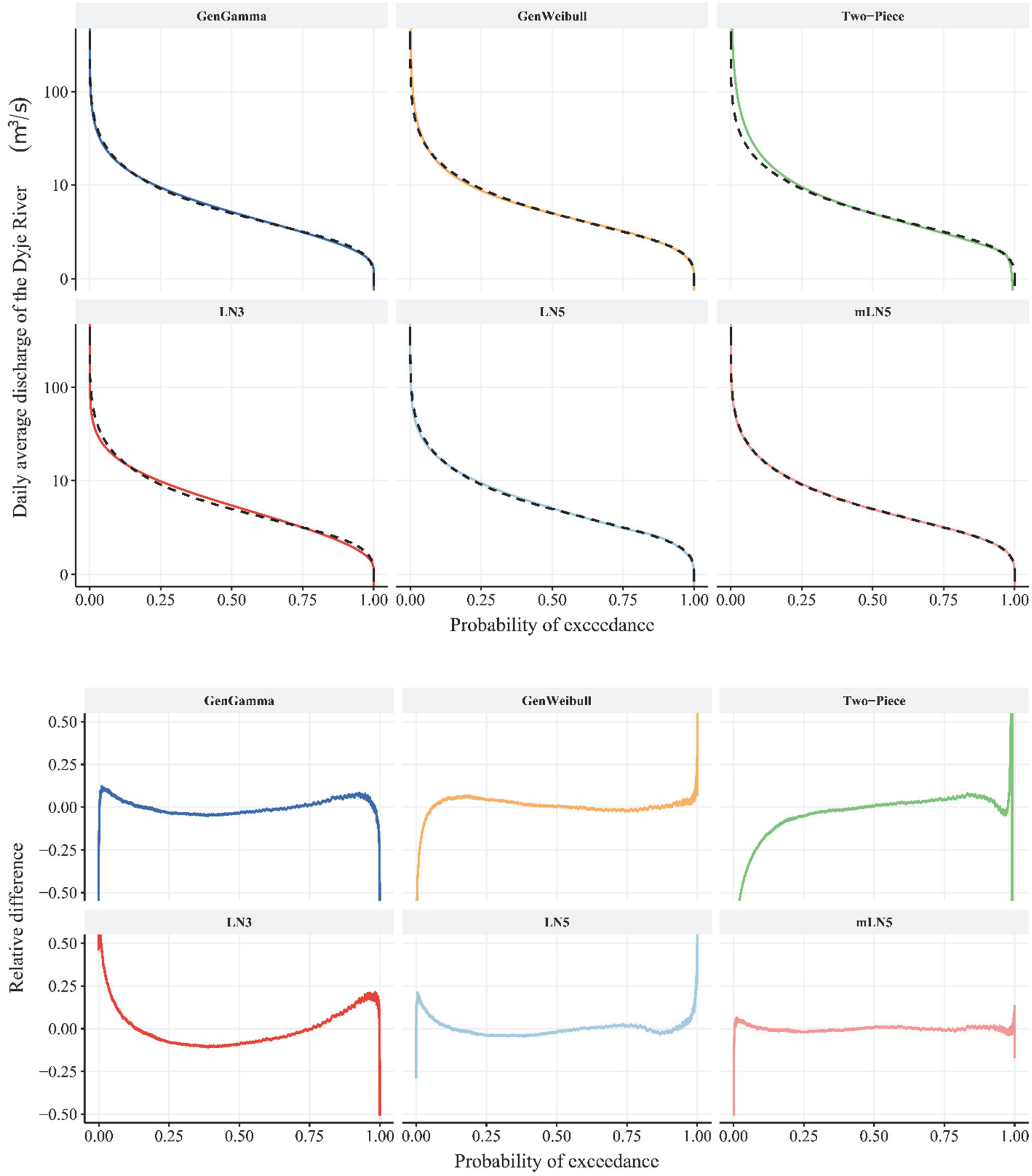

Fig. 4. Exceedance curves for the Dyje River. Upper: The empirical exceedance curve (black dashed line) and fitted exceedance curves for the generalized gamma (blue line), generalized Weibull (orange line), Cauchy two-piece (green line), LN3 (red line), LN5 (light blue line) and mLN5 (pink line) distributions. Lower: Relative difference of estimated and empirical exceedance curves.

4.2. Application in climatology

To illustrate the mLN5 distribution's capabilities, we applied it to daily total precipitation, average daily atmospheric pressure, and average daily air temperature datasets provided by the Czech Hydrometeorological Institute. The precipitation dataset (mm) from the Žatec station includes 44739 observations, with 17110 days of recorded precipitation. The atmospheric pressure dataset (hPa) from the Dukovany station comprises 13514 observations, and the temperature dataset (°C) from the Lysá hora station contains 21915 observations. For temperature data, an alternative parameterization (Appendix C of the Supplementary material) is necessary for the mLN5 distribution to prevent overflow during parameter estimation.

Fig. 5 displays the empirical and estimated exceedance curves for the Žatec station. Traditionally, exceedance curve estimation focused solely on days with measurable rainfall. However, the application of LN5 and mLN5 distributions enables parameter estimation for exceedance curves on both precipitation and non-precipitation days. While significant differences are evident between empirical and estimated curves using the LN3 distribution, such discrepancies are notably absent with the LN5 and mLN5 distributions.

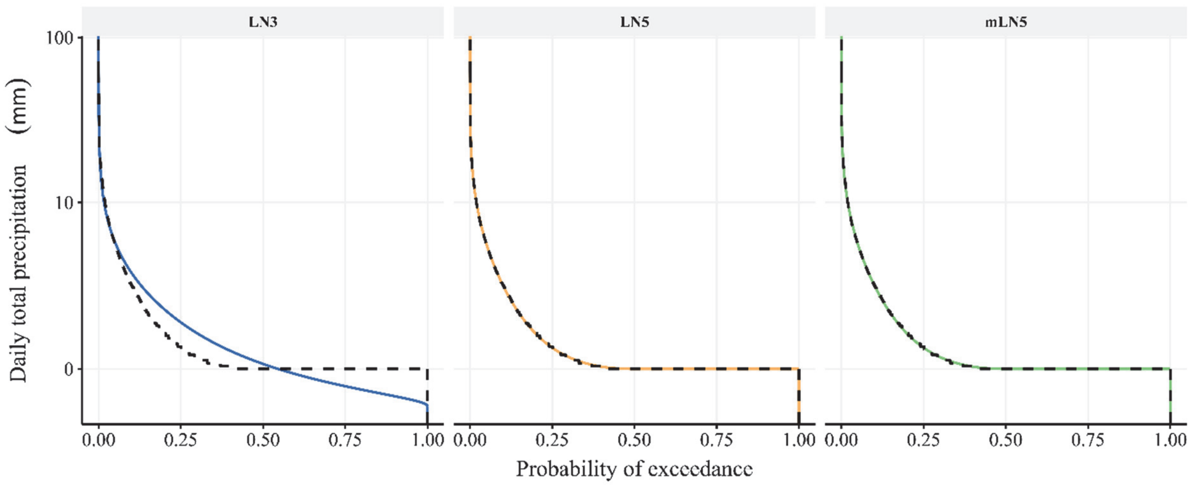

Fig. 5. The empirical exceedance curve (rain) for the Žatec station (black dashed line) and fitted exceedance curves for the LN3 (blue line), LN5 (orange line) and mLN5 (green line) distributions.

For average temperature, we found that using values proportional to potentially radiated energy (according to the Stephan-Boltzmann law) instead of Celsius degrees yields better empirical and estimated curve agreement. *Fig. 6* shows the empirical and estimated exceedance curves for atmospheric pressure at Dukovany and temperature (converted to radiated energy) at Lysá hora. The graphs indicate a near-perfect match in the middle range, though a small sample size may cause imperfections at the margins for pressure data. This mLN5 transformation can be extended to other climatological data, considering factors like air humidity and sunshine duration and intensity.

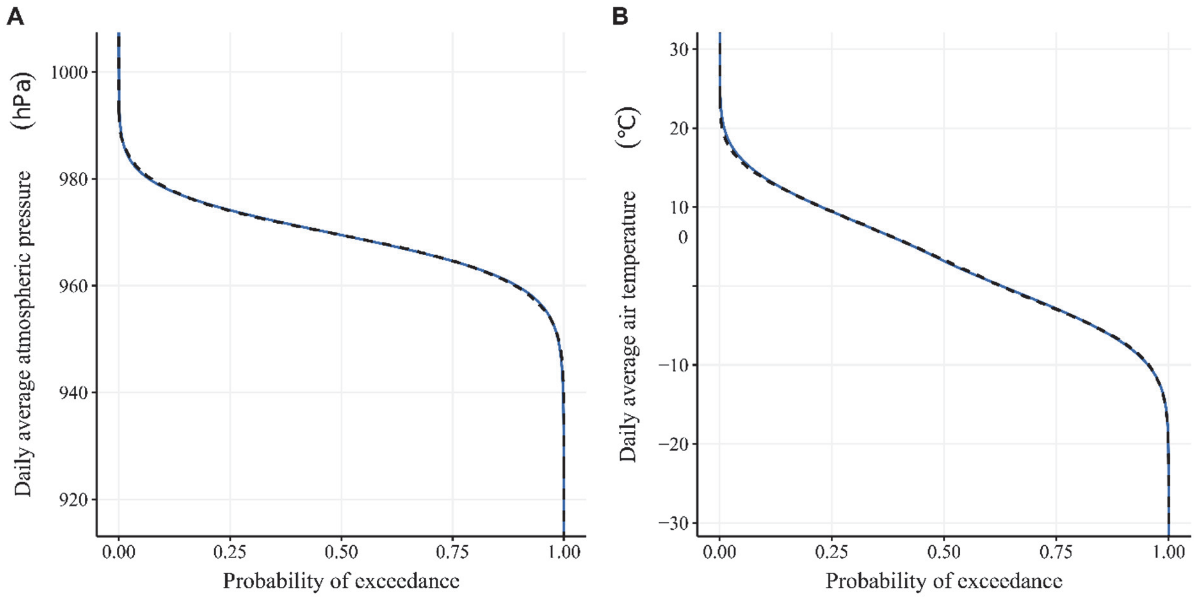

Fig. 6. The empirical exceedance curve (black dashed line) and fitted exceedance curves for the mLN5 (blue line) distribution for atmospheric pressure **A** for Dukovany, and temperature **B** for Lysá hora.

5. Discussion and conclusions

In this paper, we introduced the five-parameter log-normal distribution (LN5) and its modification (mLN5), offering alternatives to the commonly used three-parameter log-normal distribution (LN3) (*Sangal* and *Biswas*, 1970), generalized Weibull distribution (*Mudholkaret al*., 1996), generalized gamma distribution (*Cox et al*., 2007) or Cauchy two-piece distributions (*Rubio* and *Hong*, 2016) for hydrological and climatological data analysis. We provided formulas for their probability density, cumulative distribution, and quantile functions, and outlined parameter estimation methods. Future research could enhance these models, exploring methods like moments estimation and properties of local maximum likelihood estimates.

Hydrological and climatological data have unique characteristics, with uncertainties in both measured values and exceedance probabilities. Traditional least squares methods assume accurate exceedance probabilities but error-prone values, while probability optimization assumes precise values but uncertain probabilities. Our triangular method, minimizing deviations in both dimensions, emerges as particularly suitable for such data. It is also effective for asymmetric exceedance curves, common in hydrology and climatology. As part of exploring the practical applications of the mLN5 distribution, we endeavored to model the exceedance curves for medical data, and once again, we observed a highly favorable agreement between the empirical and theoretical exceedance curves.

The triangular method, particularly with inverse transformation, requires data compatible with normal distribution transformation (e.g., LN2, LN3). While this paper doesn't delve into the triangular method's theoretical aspects, our simulation study (*Budík* and *Budíková*, 2020) (see Appendix F of the Supplementary material) indicates its effectiveness, especially in challenging extrapolations. Its computational efficiency is proven in processing extensive Central European climatological and hydrological datasets, with detailed results to be published separately.

Natural processes often produce data that mixes distributions. For example, flood-induced flow changes affect exceedance curve parameters and may even lead to changes in the distribution itself due to natural causes such as overtopping of reservoirs that may change flow mechanisms from groundwater to surface, etc. We are currently developing a heuristic approach to enhance the accuracy of modeling in such complex scenarios and anticipate publishing the results later. Based on the analysis of hundreds of datasets on average daily flows, we have concluded that there is a certain degree of inverse dependency between the parameters b and σ. As b increases, σ decreases, and vice versa. Furthermore, for streams with large catchment areas (on the order of 10^4 km^2), the parameter b is close to 1. In contrast, for streams with smaller catchment areas, the parameter b can deviate from 1 in both directions. Its values are influenced by the geological characteristics of the catchment, the quantity and quality of vegetation, and the precipitation regime.

This study has several limitations. The LN5 and mLN5 distributions mark a significant step in modeling exceedance curves, particularly for extreme event probabilities, aiding in understanding and adapting to climate change. However, the analyzed data may not always meet independence and identical distribution assumptions. Our proposed procedures, considering data heterogeneity, autocorrelation, and seasonality, have shown promising results in modeling exceedance curves, but further research is needed in this direction. We focused on two parameter estimation methods: the triangular method and local maximum likelihood. Other methods like Bayesian estimation or L-moments could be explored, though they assume precise exceedance probabilities, often unmet in real data. We acknowledge that utilizing a five-parameter distribution typically involves significant computational intensity. A specific challenge in parameter estimation arises from the non-smooth nature of the probability densities associated with the LN5 and mLN5 distributions. Nevertheless, these drawbacks are counterbalanced by the fact that both distributions effectively capture not only the central tendency but also the entire exceedance curve. While analyzing the average daily discharge on the Morava River and the Dyje River, it was revealed that the mLN5 distribution yielded the lowest AIC value among the six distributions investigated (refer to

Table 2). Furthermore, it exhibited the most favorable trajectory of relative estimation errors, as depicted in Figures 3b and 4b. It's important to note that unlike earlier distributions used for exceedance curves, the LN5 and mLN5 distributions enable concurrent analysis of days with precipitation and those without precipitation. Further research should aim to refine the mLN5 distribution and investigate other potential models for better approximating hydrological and climatological data.

Average daily precipitation totals, average daily temperature, and average daily air pressure from the CHMI stations can be found at the following address: https://www.chmi.cz/historicka-data/pocasi/denni-data/Denni-data-dle-z.-123-1998-Sb. Average daily flow rates are available for download at https://isvs.chmi.cz/ords/f?p=11002:HOME:9046927352185:::::. The source code related to this research is available on GitLab at https://gitlab.ics.muni.cz/9607/ln5.

Aknowledgements: This research was funded by the Ministry of Education, Youth, and Sports of the Czech Republic for SustES—Adaptation strategies for sustainable ecosystem services and food security under adverse environmental conditions project [grant number CZ.02.1.01/0.0/0.0/16 019/0000797] and Mathematical and Statistical Modeling project [grant number MUNI/A/1132/2022].

MB: conceptualization, methodology, writing – original draft, validation, project administration. JH: conceptualization, formal analysis, investigation, writing – original draft, software, computations, visualization. LB: conceptualization, investigation, methodology, data curation, writing – original draft, computations, validation. LP: conceptualization, writing – review & editing. IH: conceptualization, validation, supervision.

The authors thank Gejza Wimmer for constructive criticism of the manuscript and Ondřej Pokora for careful checking of the mathematical part of the text.

References

Budík, L., 2018: Hydrological and meteorological data and modifications of statistical distributions - heuristic approach. 17th Conference on Applied Mathematics, APLIMAT 2018 Proceedings*,* Bratislava: STU, 165–172.

Budík, L., 2019: Progress in heuristic approach to stochastic modeling of hydrological and meteorological data. 18th Conference on Applied Mathematics, APLIMAT 2019 Proceedings*,* Bratislava: STU, 109–119.

Budík, L., and Budíková, M., 2020: Comparison of various methods of parameters estimation of mLN5 distribution using simulation study. 19th Conference on Applied Mathematics, APLIMAT 2020 Proceedings, Bratislava: STU, 170–179.

Cohen, A., 1951: Estimating parameters of logarithmic-normal distributions by maximum likelihood. *J. Amer. Stat. 46*(1), 206–212. https://doi.org/10.1080/01621459.1951.10500781

Cox, C., Chu, H., Schneider, M. F., and *Munoz, A.* (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized. *Stat. medicine 26*(23), 4352–4374. https://doi.org/10.1002/sim.2836

Griffiths, D.A., 1980: Interval estimation for the three-parameter lognormal distribution via the likelihood function. *J. Roy. Stat. Soc.. Series C (Applied Statistics) 29*(1) , 58–68. https://doi.org/10.2307/2346411

Johnson, N.L., Kotz, S., and *Balakrishnan, N.,* 1994: Continuous Univariate Distributions, vol. 1. New York: Wiley-Interscience.

Lane, B., 2002: Statistical methods in hydrology, vol. 496. Iowa: Iowa State Press.

Millard, S.P., 2013: EnvStats: An R Package for Environmental Statistics. New York: Springer. https://doi.org/10.1007/978-1-4614-8456-1

Mudholkar, G.S., Srivastava, D.K., and *Kollia, G.D.,* 1996: A generalization of the weibull distribution with application to the analysis of. *J. Amer. Stat. Assoc. 91*(436), 1575–1583. https://doi.org/10.1080/01621459.1996.10476725
Rubio, F. J., and *Hong, Y.,* 2016: Survival and lifetime data analysis with a flexible class of distributions. *J. Appl. Stat. 43*(10), 1794–1813. https://doi.org/10.1080/02664763.2015.1120710
Sangal, B., and *Biswas, A.,* 1970: The 3-parameter log normal distribution and its applications in hydrology. *Water Resour. Res. 6*, 505–515. https://doi.org/10.1029/WR006i002p00505
Sprott, D. A., 2000: Statistical Inference in Science. New York: Springer.
Wingo, D. R., 1975: The use of interior penalty functions to overcome lognormal distribution parameter estimation anomalies. *J. Stat. Comput. Simul. 4*(1), 49–61. https://doi.org/10.1080/00949657508810109

Supplementary material
Five-parameter log-normal distribution and its modification

Marie Budíková[a,*], Jan Holub[b], Ladislav Budík[c], Lenka Přibylová[a], Ivana Horová[a]

[a]*Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlarska 2, Brno, 611 37, Czech Republic*
[b]*Global Change Research Institute of the Czech Academy of Sciences, Belidla 986/4a, Brno, 603 00, Czech Republic*
[c]*Czech Hydrometeorological Institute, Kroftova 2578/43, Brno, 616 67, Czech Republic*

Appendices

A Five-parameter log-normal distribution (LN5)

Let t be a real function $t(x) = a\exp\{\operatorname{sgn}x \cdot |x|^b\} + y_0$. Let τ_1 be an inverse function to the function t on $H_1 = t(G_1) = (y_0, a + y_0)$, where $G_1 = (-\infty, 0)$, in the form $\tau_1(y) = -\ln^{\frac{1}{b}}(\frac{a}{y-y_0})$, and its derivative $\tau_1'(y) = \frac{1}{b(y-y_0)}\ln^{\frac{1-b}{b}}(\frac{a}{y-y_0})$. Similarly, let τ_2 be an inverse function to the function t on $H_2 = t(G_2) = (a + y_0, \infty)$, where $G_2 = (0, \infty)$, in the form $\tau_2(y) = \ln^{\frac{1}{b}}(\frac{y-y_0}{a})$, $\tau_2'(y) = \frac{1}{b(y-y_0)}\ln^{\frac{1-b}{b}}(\frac{y-y_0}{a})$.

Let $X \sim \mathrm{N}(\mu, \sigma^2)$ be a normally distributed random variable and $g_X(x)$ be its probability density function. The function $t(x)$ is regular on open and disjoint intervals G_1, G_2, and function $f(y) = \sum_{i=1}^{2} f_j(y)$ is **a probability density function** of transformed random variable $Y = t(X)$ that has **LN5 distribution**, where

$$f_j(y) = \begin{cases} g_X(\tau_j(y))|\tau'(y)| = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{-\frac{1}{2\sigma^2}[\tau_j(y)-\mu]^2\Big\}\big|\tau_j'(y)\big|, & y \in H_j, \\ 0, & \text{otherwise.} \end{cases}$$

Function $f(y)$ can be expressed as

$$f(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{\ln^{\frac{1-b}{b}}\left(\frac{a}{y-y_0}\right)}{b(y-y_0)} \cdot \exp\left\{-\frac{\left(-\ln^{\frac{1}{b}}\left(\frac{a}{y-y_0}\right)-\mu\right)^2}{2\sigma^2}\right\}, & y \in (y_0, y_0 + a), \\ \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{\ln^{\frac{1-b}{b}}\left(\frac{y-y_0}{a}\right)}{b(y-y_0)} \cdot \exp\left\{-\frac{\left(\ln^{\frac{1}{b}}\left(\frac{y-y_0}{a}\right)-\mu\right)^2}{2\sigma^2}\right\}, & y \in (y_0 + a, \infty), \\ 0, & \text{otherwise.} \end{cases}$$

Let $F(y, \boldsymbol{\theta}) = \int_{-\infty}^{y} f(t, \boldsymbol{\theta})\mathrm{d}t$ denote the cumulative distribution function for the probability density function $f(y, \boldsymbol{\theta})$ of LN5 distribution with parameters $\boldsymbol{\theta} = (a, b, \mu, \sigma^2, y_0)^\top$. Let $U \sim \mathrm{N}(0, 1)$ be a random variable with its cumulative distribution function $\Phi(u) = \frac{1}{2}[1 + \operatorname{erf}(\frac{u}{\sqrt{2}})]$ where erf is the error function $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\mathrm{d}t$. For transformations

*Corresponding author
Email address: budikova@math.muni.cz (Marie Budíková)

w_1, w_2, we have

$$w_1 = \frac{-\ln^{\frac{1}{b}}\left(\frac{a}{t-y_0}\right)-\mu}{\sigma}, \qquad w_2 = \frac{\ln^{\frac{1}{b}}\left(\frac{t-y_0}{a}\right)-\mu}{\sigma},$$

$$\mathrm{d}w_1 = \frac{\ln^{\frac{1-b}{b}}\left(\frac{a}{t-y_0}\right)}{\sigma b(t-y_0)}\mathrm{d}t, \qquad \mathrm{d}w_2 = \frac{\ln^{\frac{1-b}{b}}\left(\frac{t-y_0}{a}\right)}{\sigma b(t-y_0)}\mathrm{d}t,$$

$$y_0 \xrightarrow{w_1} -\infty, \qquad a+y_0 \xrightarrow{w_2} -\frac{\mu}{\sigma},$$

$$z_1 \xrightarrow{w_1} v_1 := \frac{-\ln^{\frac{1}{b}}\left(\frac{a}{z_1-y_0}\right)-\mu}{\sigma}, \qquad z_2 \xrightarrow{w_2} v_2 := \frac{\ln^{\frac{1}{b}}\left(\frac{z_2-y_0}{a}\right)-\mu}{\sigma}.$$

Let $z_1 \in H_1$, then

$$F(z_1) = \int_{-\infty}^{y_0} 0\,\mathrm{d}t + \int_{y_0}^{z_1} \frac{1}{\sqrt{2\pi}\sigma}\cdot\frac{\ln^{\frac{1-b}{b}}\left(\frac{a}{t-y_0}\right)}{b(t-y_0)}\cdot\exp\left(-\frac{\left(-\ln^{\frac{1}{b}}\left(\frac{a}{t-y_0}\right)-\mu\right)^2}{2\sigma^2}\right)\mathrm{d}t =$$
$$= \int_{-\infty}^{v_1}\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}w_1^2\right)\mathrm{d}w_1 = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\ln^{\frac{1}{b}}\left(\frac{a}{z_1-y_0}\right)-\mu}{\sqrt{2}\sigma}\right)\right],$$

Let $z_2 \in H_2$, then

$$F(z_2) = \int_{-\infty}^{y_0} 0\,\mathrm{d}t + \int_{y_0}^{a+y_0} \frac{1}{\sqrt{2\pi}\sigma}\cdot\frac{\ln^{\frac{1-b}{b}}\left(\frac{a}{t-y_0}\right)}{b(t-y_0)}\cdot\exp\left(-\frac{\left(-\ln^{\frac{1}{b}}\left(\frac{a}{t-y_0}\right)-\mu\right)^2}{2\sigma^2}\right)\mathrm{d}t+$$
$$+\int_{a+y_0}^{z_2}\frac{1}{\sqrt{2\pi}\sigma}\cdot\frac{\ln^{\frac{1-b}{b}}\left(\frac{t-y_0}{a}\right)}{b(t-y_0)}\cdot\exp\left(-\frac{\left(\ln^{\frac{1}{b}}\left(\frac{t-y_0}{a}\right)-\mu\right)^2}{2\sigma^2}\right)\mathrm{d}t = \Phi\left(-\frac{\mu}{\sigma}\right)-\Phi(-\infty)+$$
$$+\Phi\left(\frac{\ln^{\frac{1}{b}}\left(\frac{z_2-y_0}{a}\right)-\mu}{\sigma}\right)-\Phi\left(-\frac{\mu}{\sigma}\right) = \Phi\left(\frac{\ln^{\frac{1}{b}}\left(\frac{z_2-y_0}{a}\right)-\mu}{\sigma}\right) = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\ln^{\frac{1}{b}}\left(\frac{z_2-y_0}{a}\right)-\mu}{\sqrt{2}\sigma}\right)\right].$$

Combining terms $F(z_1)$, $F(z_2)$, we obtain **cumulative distribution function**

$$F(y,\boldsymbol{\theta}) = \begin{cases} 0, & y \in (-\infty,\ y_0),\\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{-\ln^{\frac{1}{b}}\left(\frac{a}{y-y_0}\right)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [y_0,\ a+y_0),\\ \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\ln^{\frac{1}{b}}\left(\frac{y-y_0}{a}\right)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in [a+y_0,\ \infty),\end{cases}$$

Let $X \sim \mathrm{N}(\mu,\sigma^2)$ and $F_X^{-1}(\alpha)$ its quantile function $F_X^{-1}(\alpha) = \mu + \sqrt{2}\sigma\,\mathrm{erf}^{-1}(2\alpha-1)$, for $\alpha \in (0,1)$. Then the transformed random variable $Y = t(X)$, $Y \sim \mathrm{LN5}(a,b,\mu,\sigma^2,y_0)$, has **quantile function**

$$F^{-1}(\alpha) = t(F_X^{-1}(\alpha)) = a\cdot\exp\left\{\mathrm{sgn}(F_X^{-1}(\alpha))\left|F_X^{-1}(\alpha)\right|^b\right\}+y_0, \quad \alpha \in (0,1).$$

This is equivalent to

$$F^{-1}(\alpha,\boldsymbol{\theta}) = \begin{cases} a\exp\left\{-\left[-\mu-\sqrt{2}\sigma\mathrm{erf}^{-1}(2\alpha-1)\right]^b\right\}+y_0, & \alpha\in I_1,\\ a\exp\left\{\left[\mu+\sqrt{2}\sigma\mathrm{erf}^{-1}(2\alpha-1)\right]^b\right\}+y_0, & \alpha\in I_2,\end{cases}$$

where

$$I_1=\left(0,\ \tfrac{1}{2}\left[1+\mathrm{erf}\left(\tfrac{-\mu}{\sqrt{2}\sigma}\right)\right]\right),\quad I_2=\left[\tfrac{1}{2}\left[1+\mathrm{erf}\left(\tfrac{-\mu}{\sqrt{2}\sigma}\right)\right],\ 1\right).$$

B Modified five-parameter log-normal distribution (mLN5)

Let function t be a transformation

$$t(x)=\begin{cases} t_1(x)=a\exp\{-(-x)^b\}+y_0, & x\in G_1=(-\infty,-1),\\ t_2(x)=a\exp\{-(-x)^{b+(1-b)(1+x)}\}+y_0, & x\in G_2=(-1,0),\\ t_3(x)=a\exp\{x^{b+(1-b)(1-x)}\}+y_0, & x\in G_3=(0,1),\\ t_4(x)=a\exp\{x^b\}+y_0, & x\in G_4=(1,\infty).\end{cases}$$

Now, the following expressions hold for the inverse functions:

Function $\tau_1(y)=-\ln^{\frac{1}{b}}(\frac{a}{y-y_0})$ is an inverse function to t_1 on $H_1=t(G_1)=(y_0,\ ae^{-1}+y_0)$, and its derivative satisfies $\tau_1'(y)=\frac{1}{b(y-y_0)}\ln^{\frac{1-b}{b}}(\frac{a}{y-y_0})$. Function τ_2 is an inverse function to t_2 on $H_2=t(G_2)=(ae^{-1}+y_0,\ a+y_0)$ and $\tau_2(y)=x$ solves equation $y=t_2(x)$ in form of equation

$$\ln[\ln(\tfrac{a}{y-y_0})]=[b+(1-b)(1+x)]\ln(-x).$$

Moreover, its derivative satisfies $\tau_2'(y)=\left[t_2'(x)|_{x=\tau_2(y)}\right]^{-1}$, where

$$t_2'(x)=a\exp\{-(-x)^{1+x-bx}\}(-x)^{x-bx}[1+x(1-b)(1+\ln(-x))].$$

Function τ_3 is as an inverse function to the function t_3 on $H_3=t(G_3)=(a+y_0,\ ae+y_0)$, and $\tau_3(y)=x$ solves equation $y=t_3(x)$ in form of equation

$$\ln[\ln(\tfrac{y-y_0}{a})]=[b+(1-b)(1-x)]\ln(x),$$

with derivative with respect to y given as $\tau_3'(y)=[t_3'(x)|_{x=\tau_3(y)}]^{-1}$, where

$$t_3'(x)=a\exp\{x^{1-x+bx}\}x^{-x+bx}[1-x(1-b)(1+\ln(x))].$$

Function τ_4 is an inverse function to the function t_4 on $H_4=t(G_4)=(ae+y_0,\ \infty)$, that is $\tau_4(y)=\ln^{\frac{1}{b}}(\frac{y-y_0}{a})$, and its derivative is $\tau_4'(y)=\frac{1}{b(y-y_0)}\ln^{\frac{1-b}{b}}(\frac{y-y_0}{a})$.

Let $X\sim\mathrm{N}(\mu,\sigma^2)$ be a normally distributed random variable and $g_X(x)$ be its probability density function. Let function $t(x)$ be regular on open and disjoint intervals $G_1, \ldots, G_4$. Then function $f(y) = \sum_{i=1}^{4} f_i(y)$ is **a probability density function** of transformed random variable $Y = t(X)$ that has **mLN5 distribution**, where

$$f_j(y) = \begin{cases} g_X(\tau_j(y))|\tau'(y)| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}[\tau_j(y) - \mu]^2\right\}\left|\tau_j'(y)\right|, & y \in H_j, \\ 0, & \text{otherwise.} \end{cases}$$

It can be shown that $t_2'(x) > 0, \forall x \in G_2$, and $t_3'(x) > 0, \forall x \in G_3$, if $b < b_{\max} \doteq 8,389$. Hence function t is regular on intervals $G_1, \ldots, G_4$ for $0 < b < b_{\max}$.

To obtain cumulative distribution function $F(y, \boldsymbol{\theta})$ from known probability density function $f(y, \boldsymbol{\theta})$, we evaluate non-negative parts $f_j(y, \boldsymbol{\theta})$. Let $z_j \in H_j$ denote endpoints of interval $H_j = (H_j^L, H_j^U)$, $j = 1, \ldots, 4$, then

$$\int_{H_j^L}^{z_j} f(t)\mathrm{d}t = \int_{H_j^L}^{z_j} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}[\tau_j(t) - \mu]^2\right\}|\tau_j'(t)|\mathrm{d}t.$$

Let $w_j = \frac{\tau_j(t)-\mu}{\sigma}$ be transformations that modify this expression. Then we have for the bounds

$$z_j \xrightarrow{\tau_j} \frac{\tau_j(z_j)-\mu}{\sigma}, y_0 \xrightarrow{\tau_1} -\infty, ae^{-1} + y_0 \xrightarrow{\tau_1,\tau_2} \frac{-1-\mu}{\sigma}, a + y_0 \xrightarrow{\tau_2,\tau_3} \frac{-\mu}{\sigma}, ae + y_0 \xrightarrow{\tau_3,\tau_4} \frac{1-\mu}{\sigma}, \infty \xrightarrow{\tau_4} \infty.$$

Then integrating density $f(t, \boldsymbol{\theta})$ over sets $H_1, \ldots, H_4$ gives

$$\begin{aligned} \int_{H_1} f(t, \boldsymbol{\theta})\mathrm{d}t &= \tfrac{1}{2}[1 + \mathrm{erf}(\tfrac{-1-\mu}{\sqrt{2}\sigma})] - 0, \\ \int_{H_2} f(t, \boldsymbol{\theta})\mathrm{d}t &= \tfrac{1}{2}[1 + \mathrm{erf}(\tfrac{-\mu}{\sqrt{2}\sigma})] - \tfrac{1}{2}[1 + \mathrm{erf}(\tfrac{-1-\mu}{\sqrt{2}\sigma})] \\ \int_{H_3} f(t, \boldsymbol{\theta})\mathrm{d}t &= \tfrac{1}{2}[1 + \mathrm{erf}(\tfrac{1-\mu}{\sqrt{2}\sigma})] - \tfrac{1}{2}[1 + \mathrm{erf}(\tfrac{-\mu}{\sqrt{2}\sigma})] \\ \int_{H_4} f(t, \boldsymbol{\theta})\mathrm{d}t &= 1 - \frac{1}{2}[1 + \mathrm{erf}(\tfrac{1-\mu}{\sqrt{2}\sigma})] \end{aligned}$$

Hence

$$F(z_j, \boldsymbol{\theta}) = \int_{H_j^L}^{z_j} f(t)\,\mathrm{d}t = \int_{-\infty}^{y_0} 0\,\mathrm{d}t + \int_{H_1} f(t)\,\mathrm{d}t + \cdots + \int_{H_j^L}^{z_j} f(t)\,\mathrm{d}t = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z_j - \mu}{\sqrt{2}\sigma}\right)\right],$$

for $j = 1, \ldots, 4$. This yields full form of **cumulative distribution function** $F(y, \boldsymbol{\theta})$

$$F(y, \boldsymbol{\theta}) = \begin{cases} 0, & y \in (-\infty,\ y_0), \\ \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\tau_j(y)-\mu}{\sqrt{2}\sigma}\right)\right], & y \in H_j, j = 1, \ldots, 4. \end{cases}$$

Quantile function $F^{-1}(y, \boldsymbol{\theta})$ is obtained as an inverse function of cumulative distribution function $F(y, \boldsymbol{\theta})$. Let $y \in H_j$, $j = 1, \ldots, 4$, then for $\alpha = \frac{1}{2}[1 + \mathrm{erf}(\frac{\tau_j(y)-\mu}{\sqrt{2}\sigma})] \in I_j$, we have

$$\tau_j(y) = \mu + \sqrt{2}\sigma\mathrm{erf}^{-1}(2\alpha - 1) \text{ and } y = t_j\left(\mu + \sqrt{2}\sigma\mathrm{erf}^{-1}(2\alpha - 1)\right),$$

where

$$
\begin{aligned}
I_1 &= (F(y_0), F(ae^{-1}+y_0)) = \left(0,\ \tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{-1-\mu}{\sqrt{2}\sigma}\right)\right]\right),\\
I_2 &= [F(ae^{-1}+y_0), F(a+y_0)) = \left[\tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{-1-\mu}{\sqrt{2}\sigma}\right)\right],\ \tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{-\mu}{\sqrt{2}\sigma}\right)\right]\right),\\
I_3 &= [F(a+y_0), F(ae+y_0)) = \left[\tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{-\mu}{\sqrt{2}\sigma}\right)\right],\ \tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{1-\mu}{\sqrt{2}\sigma}\right)\right]\right),\\
I_4 &= [F(ae+y_0), F(\infty)) = \left[\tfrac{1}{2}\left[1+\operatorname{erf}\left(\tfrac{1-\mu}{\sqrt{2}\sigma}\right)\right],\ 1\right).
\end{aligned}
$$

C Alternative parametrization

The five-parameter log-normal distribution and its modification can be written with alternative parametrization with the potential to improve the numerical optimization of estimation methods.

Let $\mu' = \frac{\mu}{\sigma}$, $\sigma' = \sigma^b$, then $\operatorname{sgn} X|X|^b$ can be written as $\operatorname{sgn}(U+\mu')|U+\mu'|^b\sigma'$, where $X = \mu + \sigma U$ and U is a random variable with a standardized normal distribution, $U \sim \mathrm{N}(0,1)$. Then random variable Y from the LN5 distribution is in the form of

$$Y = a\exp\{\operatorname{sgn}(U+\mu')\sigma'|U+\mu'|^b\} + y_0.$$

If $|X| < 1$, then $|X|^{b+(1-b)(1-|X|)} = |U+\mu'|^{b+(1-b)(1-|X|)}(\sigma')^{(1+(\frac{1}{b}-1)(1-|X|))}$. Hence a random variable Y from the mLN5 distribution can be written as follows

$$
Y = \begin{cases}
a\exp\{\operatorname{sgn}(U+\mu')\sigma'|U+\mu'|^b\} + y_0, & |X| \geq 1,\\
a\exp\{\operatorname{sgn}(U+\mu')|U+\mu'|^{b+(1-b)(1-|X|)}(\sigma')^{(1+(\frac{1}{b}-1)(1-|X|))}\} + y_0, & |X| < 1.
\end{cases}
$$

D PDF and CDF functions for LN5 and mLN5

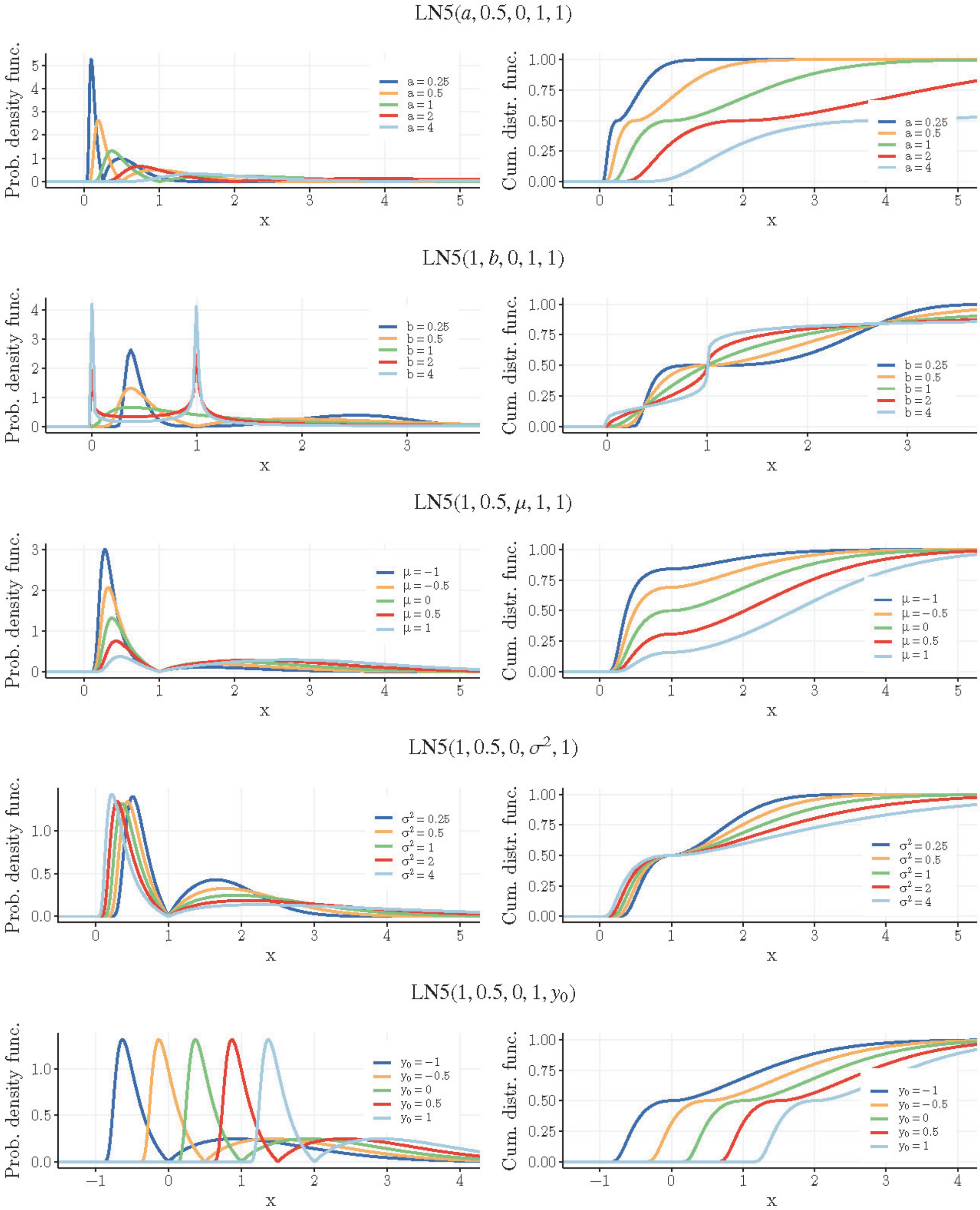

Figure 1: Probability density and cumulative distribution functions of the LN5 distribution.

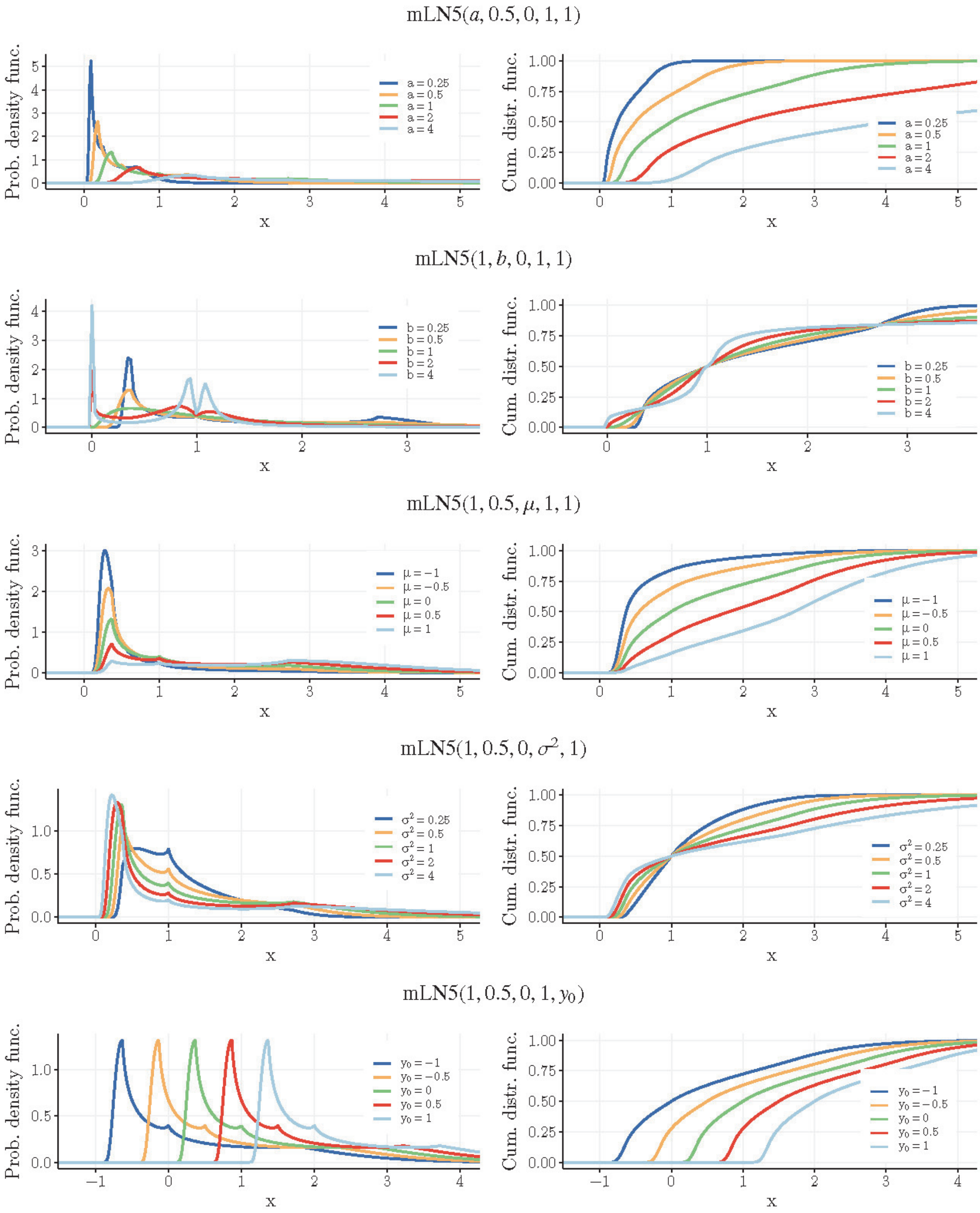

Figure 2: Probability density and cumulative distribution functions of the mLN5 distribution.

E Comparison of LMLE estimates for Morava nad Dyje rivers

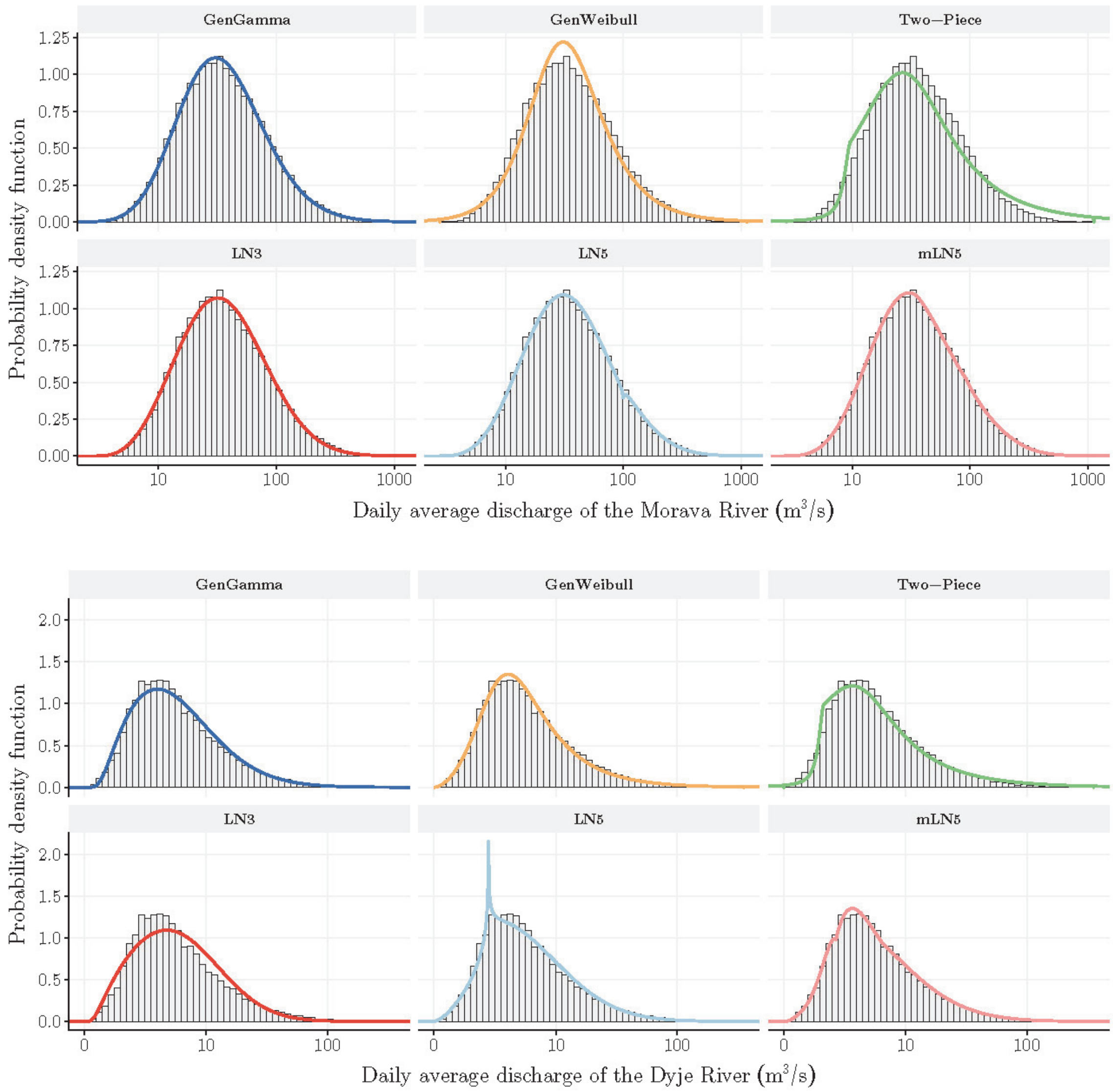

Figure 3: Histograms and superpone fitted probability density functions for the generalized gamma (blue line), generalized Weibull (orange line), Cauchy two-piece (green line), LN3 (red line), LN5 (light blue line) and mLN5 (pink line) distributions.

F Simulation study results from [1]

A set of 100,000 values was generated from a standardized normal distribution $N(0, 1)$. Seven distinct random samples were selected from this foundational dataset, each encompassing 10,000 data points. Values within these samples transformed $Y = a \exp\{\text{sgn}X \cdot |X|^b\} + y_0$ to derive values consistent with the LN5 distribution.

Throughout the sampling, the parameters were kept constant at the following values: $b = 0.8$, $\mu = 1$, $\sigma = 1$, and $y_0 = 0$. These parameters closely emulate the real cumulative frequency curves for daily water discharges. Here, $y_0 = 0$ represents scenarios where the watercourse becomes dry. The parameter a was set as 0.294528, for more details [1]. The simulation analysis revealed that among the methods evaluated in terms of parameter estimates (see 1) and extrapolation (see [1]), Triangular Method, Variant 2, emerged as the most optimal.

References

[1] Budík, L. and Budíková, M. (2020). Comparison of various methods of parameters estimation of mLN5 distribution using simulation study. In *19th Conference on Applied Mathematics, APLIMAT 2020 Proceedings*, pages 170–179, Bratislava. STU.

Table 1: Parameter estimations computed for simulation samples of dataset with parameters $a = 0.294528$, $b = 0.8$, $\mu = 1$, and $\sigma = 1$. The best fit is denoted in bold. For the estimation of parameters, the methods employed were: Relative Least Squares Method (RLSM), Probability Optimization Method (POM), Triangular Method, Variant 1: combination of RLSM and POM (TM1), and Triangular Method, Variant 2: combination of LSM and POM incorporating an inverse transformation (TM2).

Method	$\hat{a}$	$\hat{b}$	$\hat{\mu}$	$\hat{\sigma}$
Sample no. 1				
RLSM	0.243358	0.782592	1.24503	1.057016
POM	0.223280	0.770924	1.384804	1.104462
TM1	0.274504	0.783096	1.094912	1.026296
TM2	**0.275616**	**0.789582**	**1.085200**	**1.015933**
Sample no. 2				
RLSM	0.238844	0.790799	1.269184	1.033957
POM	0.210920	0.777133	1.442264	1.090935
TM1	0.268627	0.792766	1.121960	1.004279
TM2	**0.269994**	**0.798859**	**1.111332**	**0.994402**
Sample no. 3				
RLSM	0.236678	0.789933	1.271358	1.054050
POM	0.210268	0.775730	1.438383	1.110359
TM1	0.270639	0.789515	1.106404	1.022046
TM2	**0.275251**	**0.795143**	**1.081732**	**1.101214**
Sample no. 4				
RLSM	0.238529	0.788103	1.267800	1.041082
POM	0.210255	0.772953	1.446264	1.08990
TM1	0.273279	0.787952	1.103734	1.016186
TM2	**0.275736**	**0.793663**	**1.084771**	**1.050257**
Sample no. 5				
RLSM	0.255029	0.777143	1.193399	1.050257
POM	0.227308	0.770170	1.345661	1.096242
TM1	0.281798	0.779331	1.067624	1.021678
TM2	**0.286431**	**0.784197**	**1.044230**	**1.010946**
Sample no. 6				
RLSM	0.235462	0.786703	1.283621	1.056514
POM	0.205117	0.774641	1.473530	1.117418
TM1	0.259347	0.790325	1.159128	1.028999
TM2	**0.262668**	**0.794762**	**1.139972**	**1.019406**
Sample no. 7				
RLSM	0.245514	0.769837	1.255831	1.059999
POM	0.213389	0.760617	1.448668	1.119533
TM1	0.276621	0.773506	1.102148	1.022328
TM2	**0.283040**	**0.778053**	**1.070653**	**1.009546**