# Statistical Characterisation of GNSS Data for a Stationary Receiver using Non-Gaussian Distributions

Abu Bantu*[iD], Józef Wiora[iD]

*Department of Measurements and Control Systems, Faculty of Automatic Control, Electronics, and Computer Science, Silesian University of Technology, ul. Akademicka 16, 44-100, Gliwice, Poland, abu.feyo.bantu@polsl.pl, jozef.wiora@polsl.pl*

Abstract: Accurately characterising datasets is crucial for effective statistical modelling, particularly when analysing Global Navigation Satellite System (GNSS) data. While traditional approaches often assume a Gaussian distribution, real-world GNSS datasets frequently exhibit heavy-tailed and skewed properties, prompting the need to explore alternative statistical models. The study examines the suitability of non-Gaussian distributions, specifically the Laplace, skew-normal, skew-$t$, and generalised hyperbolic (GH) distributions, for modelling GNSS data obtained from a stationary receiver. Using empirical GNSS datasets, we estimate parameters within confidence intervals (CIs) through weighted maximum likelihood estimation (WMLE). Model performance is assessed using log-likelihood analysis, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and root mean squared error (RMSE). Our comparative analysis shows that heavy-tailed and skewed distributions, particularly those offering greater flexibility in capturing extreme deviations, consistently outperform the conventional normal distribution. Among the non-Gaussian models considered, the GH distribution provides the best overall performance. These results emphasise the importance of selecting appropriate statistical models to improve uncertainty quantification in GNSS-based measurements.

Keywords: non-Gaussian, weighted maximum likelihood estimation, generalised hyperbolic distribution, goodness-of-fit test, uncertainty quantification, confidence interval

## 1. INTRODUCTION

Accurately characterising datasets is essential for effective statistical modelling, particularly when dealing with asymmetric and heavy-tailed data samples. Global Navigation Satellite System (GNSS) latitude data are examples of such samples [1]. GNSS is a satellite constellation that provides signals from space, transmitting positioning and timing data to receivers on Earth [2]. This system includes GPS (USA), GLONASS (Russia), Galileo (Europe), and BeiDou (China) [3]. It is essential for applications such as navigation, surveying, mapping, and scientific research, where precise location data are required. However, GNSS positioning errors arise from multiple sources, including ionospheric and tropospheric delays, satellite clock drift, multipath interference, and receiver noise [4]. To account for these errors, classical statistical models often assume a Gaussian distribution of positioning errors. However, empirical studies have revealed deviations from normality, suggesting that errors exhibit heavy-tailed and skewed characteristics [5]. These deviations challenge conventional error modelling approaches and highlight the need for alternative probability distributions that better capture the statistical properties of positioning errors.

Residual pseudorange errors in single-frequency differential GNSS systems arise from sources such as signal multipath, increased receiver noise due to interference, residual differential troposphere error, and ionospheric gradients [6]. To address these errors, the study proposed modelling the total pseudo-range error distribution by combining the probability density functions of individual error sources, moving beyond classical Gaussian assumptions. As GNSS errors originate from various sources, modelling their total impact requires alternative statistical approaches. The application of non-classical error theory of measurements (NETM) was proposed to better handle these deviations, particularly for large datasets ($n > 500$) [7]. In urban environments, GNSS positioning performance is affected by outlier measurements, including multipath effects and non-line-of-sight (NLOS) receptions.

These outliers present a significant challenge in GNSS modelling. To address these effects, Wen *et al.* [8] proposed a graduated non-convexity factor graph optimisation technique, which effectively reduced the impact of outliers and improved positioning accuracy. In addition to outlier mitigation, enhancing state estimation under non-Gaussian noise conditions has also been a key research focus. Raitoharju *et al.* [9] introduced an algorithm that integrates empirical noise models into Kalman filtering, enabling improved estimation accuracy in scenarios where Gaussian assumptions fail. This approach improved estimation accuracy by incorporating non-Gaussian noise characteristics, making it particularly useful when measurement errors deviate from Gaussian assumptions.

*Corresponding author: abu.feyo.bantu@polsl.pl (Abu Bantu)*

Although methods such as Kalman filtering can improve state estimation, selecting an appropriate statistical distribution remains a critical challenge. Several studies have examined the suitability of alternative distributions, including Cauchy, Student's t, Laplace, lognormal, and skew-normal, for data fitting and modelling applications. For example, Sun-Yong Choi *et al.* [10] applied twelve different distributions, including Cauchy, Laplace, normal, and Student's t, to model stock index returns, using information criteria and goodness-of-fit tests to identify the best-fit distribution. Alzaatreh *et al.* [11] proposes a family of generalised Cauchy distributions, denoted as T-Cauchy(Y), using the T-(RY) framework. This family is generated using the quantile functions of uniform, exponential, log-logistic, logistic, extreme value, and Fréchet distributions. It provides a flexible approach for modelling heavy-tailed data, demonstrating greater adaptability than classical Gaussian models. Aryal *et al.* [12] investigated the Laplace probability distribution and its derivatives, focusing on their applicability in modelling real-world problems. The study highlights the growing interest in constructing flexible parametric classes of probability distributions and examines how the Laplace distribution and its variants have been widely applied in statistical modelling.

Building on these findings, this study systematically evaluates the suitability of non-Gaussian distributions, including the Laplace, skew-normal, skew-t, and generalised hyperbolic (GH) distributions, for modelling GNSS latitude data from a stationary receiver. These models are applied using weighted maximum likelihood estimation (WMLE) to achieve robust parameter estimation and confidence interval (CI) construction.

The novelty of this work lies in developing a unified statistical framework that combines empirical GNSS data with non-Gaussian modelling and WMLE-based estimation to quantify measurement uncertainty more accurately under real-world, heavy-tailed conditions. This approach moves beyond the classical Gaussian assumption, providing a comparative evaluation of flexible asymmetric distributions using multiple goodness-of-fit criteria such as log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and root mean squared error (RMSE).

The added value of this study lies in its direct application of metrological principles from the Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Metrology (VIM) to non-Gaussian error modelling. This enhances the statistical reliability and interpretability of GNSS uncertainty assessment, bridging the gap between theoretical distribution fitting and practical measurement evaluation.

To achieve this, we analyse empirical GNSS latitude datasets and employ rigorous statistical methodologies for model evaluation. The rest of this paper is structured as follows: Section 2 describes the methods and materials used in our research. Section 3 details the selected synthetic data distribution; Section 4 presents goodness-of-fit tests to compare the selected models; Section 5 presents the results and discussion; and Section 6 provides the conclusion.

## 2. SYNTHETIC DATA DISTRIBUTIONS

### A. Laplace distribution

The Laplace distribution is a continuous probability distribution characterised by a sharp peak at the mean and heavier tails than the normal distribution [13]. It can be derived as the difference between two independent exponential distributions:

$$X = E_1 - E_2, \quad E_1, E_2 \sim \exp(\lambda).$$

The probability density function (PDF) of the Laplace distribution is [14]

$$f(x_i; \mu_0, \sigma) = \frac{1}{2\sigma} e^{-\frac{|x_i - \mu_0|}{\sigma}}, \quad (1)$$

where $\mu_0$ is the location parameter and $\sigma$ is the scale parameter, which must be strictly positive ($\sigma > 0$), and $x_i$ is the data point. The Laplace distribution has a variance of $2\sigma^2$ and exhibits exponentially decaying tails, which are heavier than those of the normal distribution but lighter than those of the Cauchy distribution. The Cauchy distribution is a continuous probability distribution characterised by its location parameter (mean) $\mu_0$ and scale parameter $\sigma$ [15].

### B. Skew-normal distribution

The skew-normal distribution extends the normal distribution by introducing a shape parameter ($\alpha$) to control skewness. It is defined by three parameters: location ($\mu_0$), which controls the centre of distributions; scale ($\sigma$), which determines the spread; and shape ($\alpha$) [16]. This distribution is constructed by multiplying the normal PDF by a transformation of its cumulative distribution function (CDF). A random variable $X$ follows a skew-normal distribution if $X = \mu_0 + \sigma Z$, where $Z$ follows a standard skew-normal distribution. Thus, the skew-normal PDF is given by

$$f(x_i) = 2\phi(x_i)\Phi(\alpha x_i), \quad (2)$$

where the standard normal PDF is

$$\phi(x_i) = \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \quad (3)$$

and the standard normal CDF is

$$\Phi(x_i) = \int_{-\infty}^{x} \phi(t)dt = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x_i}{\sqrt{2}}\right)\right], \quad (4)$$

where erf (error function) is defined in [6]. For a general skew-normal distribution with $\mu_0$ and $\sigma$, we standardise the variable $x_i$ as

$$z = \frac{x_i - \mu_0}{\sigma}. \quad (5)$$

Substituting (5) into (2), the general skew-normal PDF is

$$f(x_i \mid \mu_0, \sigma, \alpha) = \frac{2}{\sigma}\phi\left(\frac{x_i - \mu_0}{\sigma}\right)\Phi\left(\alpha\frac{x_i - \mu_0}{\sigma}\right). \quad (6)$$

### C. Generalised hyperbolic distribution

It is a flexible class of continuous probability distributions capable of modelling heavy tails and skewness. It includes many other distributions as special cases, such as the normal, Laplace, hyperbolic, variance gamma, and normal-inverse Gaussian (NIG) distributions [17]. The key distinction between the GH distribution and the standard hyperbolic distribution is the inclusion of the index parameters $\alpha$, resulting in a five-parameter family ($\alpha$, $\lambda$, $\beta$, $\sigma$, and $\mu$). The PDF of the GH distribution arises from the expression of $X$ as a normal mean-variance mixture, where the mixing distribution is a generalised inverse Gaussian (GIG)

$$X \mid W = w \sim \mathcal{N}(\mu + \beta w, w\sigma^2), \tag{7}$$

where $W$ is a positive mixing variable distributed as the GIG, and where $\gamma > 0$ is another scale-like parameter

$$W \sim \text{GIG}(\alpha, \sigma^2, \gamma^2) \tag{8}$$

with the PDF

$$f_W(w) = \frac{(\gamma/\sigma)^\alpha}{2K_\alpha(\sigma\gamma)} w^{\alpha-1} \exp\left(-\frac{1}{2}\left(\frac{\sigma^2}{w} + \gamma^2 w\right)\right), \quad w > 0 \tag{9}$$

$$X \sim \text{GH}(\alpha, \lambda, \beta, \sigma, \mu),$$

where the parameters are: $\alpha \in \mathbb{R}$ (shape), $\lambda > 0$ (tail heaviness), the $|\beta| < \lambda$ controls skewness, and it satisfies the constraint, $\sigma > 0$ (scale), and $\mu \in \mathbb{R}$ (location). In the one-dimensional case, the GH distribution has the following PDF

$$f_X(x_i; \alpha, \lambda, \beta, \sigma, \mu) = a(\alpha, \lambda, \beta, \sigma) \cdot \exp^{\beta(x_i-\mu)}$$
$$\cdot \frac{K_{\alpha-\frac{1}{2}}\left(\lambda\sqrt{\sigma^2+(x_i-\mu)^2}\right)}{\left(\sqrt{\sigma^2+(x_i-\mu)^2}\right)^{(1/2-\alpha)}}, \quad (x_i \in \mathbb{R}) \tag{10}$$

where $a(\alpha, \lambda, \beta, \sigma)$ is a norming constant, $K_\alpha(\cdot)$ is the modified Bessel function of the third kind (see [18]), and $a(\alpha, \lambda, \beta, \sigma)$ is a norming constant given by

$$a(\alpha, \lambda, \beta, \sigma) = \frac{(\lambda^2 - \beta^2)^{\alpha/2}}{\sqrt{2\pi}\,\lambda^{\alpha-1/2} K_\alpha\left(\sigma\sqrt{\lambda^2-\beta^2}\right)}. \tag{11}$$

Overall, the GH has the following PDF after substituting (11) into (10)

$$f_X(x_i; \alpha, \lambda, \beta, \sigma, \mu) = \frac{(\lambda^2-\beta^2)^{\alpha/2}}{\sqrt{2\pi}\,\lambda^{\alpha-1/2} K_\alpha\left(\sigma\sqrt{\lambda^2-\beta^2}\right)}$$
$$\cdot \frac{K_{\alpha-\frac{1}{2}}\left(\lambda\sqrt{\sigma^2+(x_i-\mu)^2}\right)}{\left(\sqrt{\sigma^2+(x_i-\mu)^2}\right)^{(1/2-\alpha)}} \cdot e^{\beta(x_i-\mu)}. \tag{12}$$

### D. Skew-t distribution

It is an extension of the Student's t-distribution that allows for skewness as well as heavy tails. The Student's t-distribution models the sample mean of a normally distributed

population when the sample size is small and the standard deviation is unknown [19]. The skew-*t* distribution is useful for modelling asymmetric data with outliers or heavy-tailed behaviour [20]. Its distribution arises from a normal mean-variance mixture. The random variables $Y$ and $X$ are then defined as follows

$$Y = \beta|Z| + \frac{1-\beta}{2} \cdot Z \tag{13}$$

and

$$X = \mu + \sigma \cdot \frac{Y}{\sqrt{V/\nu}}, \tag{14}$$

where: $Z \sim \mathcal{N}(0,1)$ is a standard normal variable, $V \sim \chi^2_\nu$ is a chi-squared variable with $\nu$ degrees of freedom, $\beta \in (-1,1)$ is the skewness parameter, and $\mu$ and $\sigma > 0$ are the location and scale parameters. The distribution of $X$ is skewed and heavy-tailed and is called the skew-*t* distribution. Its overall PDF is

$$f(x_i; \mu, \sigma, \alpha, \nu) = \frac{2}{\sigma} \cdot t_\nu\left(\frac{x_i-\mu}{\sigma}\right) \cdot T_{\nu+1}$$
$$\left(\alpha \cdot \frac{x_i-\mu}{\sigma} \cdot \sqrt{\frac{\nu+1}{\nu+\left(\frac{x_i-\mu}{\sigma}\right)^2}}\right). \tag{15}$$

### E. Justification for the choice distributions

The Laplace, skew-normal, skew-*t*, and GH distributions were selected for their complementary ability to represent key non-Gaussian characteristics commonly observed in GNSS data, namely asymmetry, heavy tails, and excess kurtosis. The Laplace model serves as a simple reference with moderate tail behaviour, while the skew-normal introduces a shape parameter to accommodate data asymmetry. The skew-t extends this flexibility by modelling both skewness and pronounced tail heaviness. The GH distribution provides the most general framework, encompassing several of these models as special cases and offering superior adaptability for complex GNSS error structures.

Together, these distributions form a progressive hierarchy from simple to highly flexible models, enabling a systematic evaluation of non-Gaussian behaviour in GNSS measurements. Broader generalisations, such as the generalised alpha-skew-t, variance gamma, or NIG distributions, could further enhance modelling accuracy by offering additional control over tail behaviour and asymmetry, and are recommended for future investigation.

### 3. GOODNESS-OF-FIT TESTS

A goodness-of-fit test is a statistical method used to determine how well a set of observed data matches a specific theoretical distribution or model [21].

To evaluate the suitability of various statistical models for GNSS data, it is essential to consider distributions capable of capturing asymmetry and heavy tails. Among the candidate models are the Laplace, skew-normal, GH, and skew-*t* distributions, each offering different levels of flexibility in representing real-world data characteristics. Parameter estimation for these distributions is typically performed using

the WMLE method. This is a modified form of maximum likelihood estimation (MLE) that applies weights to reduce bias, particularly for individuals with extreme response patterns [22]. The estimated parameter vector $\theta^*$ is defined as

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{n} w_i \log \mathscr{L}(\theta \mid x_i), \qquad (16)$$

where:
- $\theta$ is the vector of parameters to be estimated,
- $x_i$ is the i-th data point,
- $w_i$ is the weight assigned to $x_i$, satisfying $\sum_{i=1}^{n} w_i = 1$,
- $\mathscr{L}(\theta \mid x_i)$ is the likelihood of observing $x_i$ given parameters of $\theta$.

This approach provides robust parameter estimates and enables the construction of 95 % CIs for the fitted models. CIs are determined using the percentile method, which relies on the quantiles of the fitted distribution. The general formula for the CI is given by [23]

$$\text{CI} = \left[ F^{-1}(\alpha, \text{params}), F^{-1}(1-\alpha, \text{params}) \right], \qquad (17)$$

where $F^{-1}(\alpha, \text{params})$ and $F^{-1}(1-\alpha, \text{params})$ correspond to the lower and upper quantile functions (percent-point function, PPF) of the fitted distribution. The confidence level is set at 95 %, leading to the following values for $\alpha$, which represents the significance level used to determine the probability of the CI not capturing the true parameter. In a two-tailed CI, the total significance level $\alpha$ is symmetrically divided between the two tails of the distribution. This means each tail contains a probability of $\frac{\alpha}{2}$

$$\frac{\alpha}{2} = \frac{1 - \text{confidence level}}{2} = 0.025 \quad \text{(for a 95 \% CI)},$$
$$1 - \frac{\alpha}{2} = 0.975.$$

In this study, the constructed CIs correspond to the expected range of the measured quantity (GNSS latitude) as represented by each fitted distribution. Thus, the intervals indicate the range within which the true value of the measurand is expected to lie with a 95 % confidence level.

To evaluate the adequacy of statistical models in representing GNSS error distributions, various goodness-of-fit metrics are employed. These include the log-likelihood, AIC, BIC, and RMSE. Together, these measures provide a rigorous framework for identifying the most appropriate distributional model.

### A. Log-likelihood function ($\mathscr{L}$)

The log-likelihood function is a tool for evaluating how well a probability distribution fits the observed data. It quantifies the data compatibility with the assumed model by summing the logarithms of the PDF values at each observation. Higher values indicate a better fit.

For $n$ independent observations $X = \{x_1, x_2, \ldots, x_n\}$ and a PDF $f(x \mid \theta)$, the log-likelihood is defined as [24]

$$\mathscr{L}(\theta \mid X) = \sum_{i=1}^{n} \ln f(x_i \mid \theta), \qquad (18)$$

where $\theta$ represents the estimated parameters of the distribution, $n$ is the number of data points, and $f(x_i \mid \theta)$ is the PDF evaluated at $x_i$. Equations (19–23) are obtained by substituting the respective model distribution expressions into (18) to derive the log-likelihood function for each distribution. Starting with the Laplace distribution (1) substituted into (18), the log-likelihood function is

$$\mathscr{L}(\mu_0, \sigma) = \sum_{i=1}^{n} \ln \left[ \frac{1}{2\sigma} e^{-\frac{|x_i - \mu_0|}{\sigma}} \right]$$
$$= -n\ln(2\sigma) - \sum_{i=1}^{n} \frac{|x_i - \mu_0|}{\sigma}. \qquad (19)$$

The log-likelihood function of the skew-normal distribution is

$$\mathscr{L}(\mu_0, \sigma, \alpha) = \prod_{i=1}^{n} f(x_i \mid \mu_0, \sigma, \alpha). \qquad (20)$$

After substituting (6) into (18), the log-likelihood function is

$$\mathscr{L}(\mu_0, \sigma, \alpha) = \prod_{i=1}^{n} \frac{2}{\sigma} \phi\left( \frac{x_i - \mu_0}{\sigma} \right) \Phi\left( \alpha \frac{x_i - \mu_0}{\sigma} \right) \qquad (21)$$

taking the logarithm of (21) yields the log-likelihood function

$$\log \mathscr{L}(\mu_0, \sigma, \alpha) = \sum_{i=1}^{n} \log \left[ \frac{2}{\sigma} \phi\left( \frac{x_i - \mu_0}{\sigma} \right) \Phi\left( \alpha \frac{x_i - \mu_0}{\sigma} \right) \right]$$
$$= n\log 2 - n\log \sigma + \sum_{i=1}^{n} \log \phi\left( \frac{x_i - \mu_0}{\sigma} \right)$$
$$+ \sum_{i=1}^{n} \log \Phi\left( \alpha \frac{x_i - \mu_0}{\sigma} \right). \qquad (22)$$

By substituting (12) into (18), the log-likelihood function of the GH distribution is as follows

$$\log \mathscr{L}(\alpha, \lambda, \beta, \sigma, \mu) = \sum_{i=1}^{n} \left[ \frac{\alpha}{2} \log(\lambda^2 - \beta^2) - \left( \alpha - \frac{1}{2} \right) \log \lambda \right.$$
$$- \log\left( \sqrt{2\pi} \right) - \log K_\alpha \left( \sigma \sqrt{\lambda^2 - \beta^2} \right)$$
$$+ \log K_{\alpha - \frac{1}{2}} \left( \lambda \sqrt{\sigma^2 + (x_i - \mu)^2} \right)$$
$$- \left( \frac{1}{2} - \alpha \right) \log \left( \sqrt{\sigma^2 + (x_i - \mu)^2} \right)$$
$$\left. + \beta(x_i - \mu) \right]. \qquad (23)$$

Substituting the skew-$t$ PDF (15) into (18), the for log-likelihood is

$$\ln \mathscr{L}(\mu, \sigma, \alpha, \nu) = \sum_{i=1}^{n} \left[ \ln\left( \frac{2}{\sigma} \right) + \ln t_\nu \left( \frac{x_i - \mu}{\sigma} \right) + \ln T_{\nu+1} \right.$$
$$\left. \left( \alpha \cdot \frac{x_i - \mu}{\sigma} \cdot \sqrt{\frac{\nu + 1}{\nu + \left( \frac{x_i - \mu}{\sigma} \right)^2}} \right) \right]. \qquad (24)$$

## B. Akaike Information Criterion (AIC)

The AIC is a widely used metric for model selection, balancing goodness-of-fit with model complexity and fit quality. It is defined as [25]:

$$\text{AIC} = -2\mathscr{L}(\theta) + 2k, \tag{25}$$

where $k$ denotes the number of parameters in the model. For the Laplace distribution, which has ($k = 2$), the AIC is given by $\text{AIC} = -2\mathscr{L}(\theta) + 2 \cdot 2$, making it directly comparable based on log-likelihood values. Similarly, the skew-normal and skew-$t$ distributions, each with three parameters ($k = 3$), have AIC values expressed as $\text{AIC} = -2\mathscr{L}(\theta) + 2 \cdot 3$. The AIC for the GH distribution is calculated as $\text{AIC} = -2\mathscr{L}(\theta) + 2 \cdot 5$, where 5 represents the number of estimated parameters. A lower AIC indicates a better-fitting model with minimal complexity.

## C. Bayesian Information Complexity (BIC)

The BIC is a model selection tool that imposes a stronger penalty on model complexity than the AIC. The BIC is defined as [26]

$$\text{BIC} = -2\mathscr{L}(\theta) + k\ln(n). \tag{26}$$

The penalty term $k\ln(n)$ ensures that the BIC strongly discourages overfitting, especially for large datasets.

The Laplace distribution is characterised by two parameters ($k = 2$). Based on (25), the corresponding BIC values are calculated using the formula: $\text{BIC} = -2\mathscr{L}(\theta) + 2 \cdot \ln(n)$.

Similarly, the skew-normal and skew-$t$ distributions each involve three parameters ($k = 3$). According to (25), their BIC values are computed as: $\text{BIC} = -2\mathscr{L}(\theta) + 3 \cdot \ln(n)$.

The GH distribution involves five parameters ($k = 5$) and BIC is expressed as: $\text{BIC} = -2\mathscr{L}(\theta) + 5 \cdot \ln(n)$. A lower BIC indicates a better model fit while imposing a stronger penalty on model complexity than the AIC, thereby helping to prevent overfitting.

## D. Root Mean Squared Error (RMSE)

RMSE is a widely used metric for quantifying the accuracy of a model or estimator by measuring the average deviation between observed and predicted values. By squaring the differences, RMSE penalises larger errors more heavily, providing a sensitive measure of overall fit [27]. It is particularly useful for comparing multiple models or candidate distributions in terms of their ability to represent the observed data. In our GPS latitude data analysis, we use RMSE to evaluate how closely each fitted probability distribution aligns with the observed measurements, thereby establishing a quantitative basis for comparing model performance. Formally, for a set of $n$ observed values $y_i$ and corresponding fitted values $\hat{y}_i$ derived from a candidate distribution, RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{27}$$

where $y_i$ denotes the observed GPS latitude values, $(\hat{y}_i)$ represents the predicted or sampled values from the fitted distribution, and $(y_i - \hat{y}_i)$ indicates the residual (error) for each observation. A lower RMSE indicates better model accuracy (small average error), while a higher value suggests larger deviations between the predictions and the actual values.

The RMSE for the Laplace distribution with mean $\mu_0$ and variance $2\sigma^2$ can be directly computed as

$$\text{RMSE} = \sqrt{2\sigma^2 + (\hat{y} - \mu_0)^2}. \tag{28}$$

For the skew-normal distribution with mean $\frac{\alpha\sigma}{\alpha-1}$ and variance $\frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}$, it is

$$\text{RMSE} = \sqrt{\frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)} + \left(\hat{y} - \frac{\alpha\sigma}{\alpha-1}\right)^2}, \alpha > 2 \tag{29}$$

and undefined for $\alpha \leq 2$.

For the GH distribution with parameters $\alpha$, $\lambda$, $\beta$, $\sigma$, and $\mu$, the mean and variance are

$$\mathbb{E}[Y] = \mu + \sigma \cdot \frac{K_{\alpha+1}(\sigma\sqrt{\lambda^2 - \beta^2})}{K_\alpha(\sigma\sqrt{\lambda^2 - \beta^2})} \cdot \frac{\beta}{\sqrt{\lambda^2 - \beta^2}}, \tag{30}$$

where $K_\alpha(\cdot)$ is the modified Bessel function of the third kind. The variance of the GH distribution is given by

$$\text{Var}(Y) = \sigma^2\left[\frac{K_{\alpha+1}(\sigma\sqrt{\lambda^2 - \beta^2})}{K_\alpha(\sigma\sqrt{\lambda^2 - \beta^2})} + \frac{K_{\alpha+2}(\sigma\sqrt{\lambda^2 - \beta^2})}{K_\alpha(\sigma\sqrt{\lambda^2 - \beta^2})} - \left(\frac{K_{\alpha+1}(\sigma\sqrt{\lambda^2 - \beta^2})}{K_\alpha(\sigma\sqrt{\lambda^2 - \beta^2})}\right)^2\right]. \tag{31}$$

Then, the RMSE is

$$\text{RMSE} = \sqrt{\text{Var}(Y) + (\hat{y} - \mathbb{E}[Y])^2}. \tag{32}$$

For a skew-$t$ distribution with location parameter $\mu$, scale $\sigma$, skewness $\lambda$, and degrees of freedom $\nu$, the mean and variance (for $\nu > 2$ and $\nu > 4$, respectively) are given by

$$\mathbb{E}[Y] = \mu + \sigma\delta\sqrt{\frac{\nu}{\pi}} \cdot \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{33}$$

$$\text{Var}(Y) = \sigma^2\left(\frac{\nu}{\nu-2} - \delta^2 \cdot \frac{\nu}{\pi} \cdot \left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2\right), \tag{34}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. The RMSE is defined as

$$\text{RMSE} = \sqrt{\text{Var}(Y) + (\hat{y} - \mathbb{E}[Y])^2}. \tag{35}$$

The presented formulas [(28), (29), (32), and (35)] explicitly define how the RMSE values were computed for each fitted distribution. Each expression corresponds to the RMSE calculation applied to a specific model, quantifying the average squared deviation between the observed GPS latitude data and the values predicted by the fitted distribution. These computed RMSE values are reported in Table 1, allowing direct comparison of the fitting accuracy among the candidate distributions.

## 4. MATERIALS AND METHODS

To collect GNSS sample data, we deployed an L76k GPS module at a fixed outdoor location. This module, known for its compact design and high positioning accuracy, was connected to a Raspberry Pi, which served as the control system. The Raspberry Pi handled module configuration, data logging, and real-time monitoring, ensuring continuous and reliable data collection throughout the experiment.

During 72 hours of continuous operation, the GPS module recorded 1 303 376 rows of National Marine Electronics Association (NMEA) sentences. These included various sentence types such as:

- `$GPRMC`: Recommended Minimum Specific GNSS data,
- `$GPGGA`: Global Positioning System fix data,
- `$GPGLL`: Geographic Position Latitude / Longitude,
- `$GPVTG`: Track made good and ground speed,
- `$GPGSA`: GNSS DOP and Active Satellites,
- `$GPGSV`: GNSS Satellites in View.

From this dataset, 162 922 `$GPRMC` sentences were extracted for further analysis. These sentences were selected because they include essential positional and signal quality parameters such as fix quality, satellite count, horizontal dilution of precision (HDOP), and latitude / longitude coordinates. Python scripts were used to efficiently filter, parse, and structure the extracted data into a log file.

The GNSS latitude data distribution was first characterised before fitting multiple non-Gaussian distributions (Laplace, skew-normal, skew-$t$, and GH). Parameter estimation and CIs for each distribution were determined using WMLE to evaluate measurement reliability and quantify uncertainty. In addition, model performance was assessed using log-likelihood, AIC, BIC, and RMSE.

## 5. RESULTS AND DISCUSSION

Fig. 1 presents the PDF and CDF for the latitude data, compared with four fitted distributions using WMLE: Laplace, skew-normal, GH, and skew-$t$. For numerical consistency, the latitude values were normalised by subtracting the reference latitude of $50.28\,000°$ N, so the values reported in Ta-

ble 1 correspond to the least significant digits of the measured coordinates. For example, the Laplace model mean value $864 \times 10^{-5}$ represents an actual latitude of approximately $50.28\,864°$ N, with a 95 % CI of $[50.28\,805°, 50.28\,924°]$. The same interpretation applies to the other models listed in Table 1. In Fig. 1(a), the empirical PDF (blue histogram) exhibits a right-skewed and heavy-tailed structure. The Laplace distribution fits the peak well but fails to capture the tail behaviour, while the skew-normal improves on the asymmetry but still under-represents the tail. In contrast, the skew-$t$ distribution closely matches the empirical density across the entire range, accurately modelling the sharp peak and heavy right tail. The GH distribution exhibits the best alignment with the empirical data, particularly capturing the skewness and heavy tail behaviour, which aligns with the goodness-of-fit metrics in Table 1. Fig. 1(b) shows the empirical CDF along with the fitted CDFs. The Laplace distribution deviates significantly around the central region, while the skew-normal offers better central alignment but lacks precision in the tails. GH again provides the closest match to the empirical CDF, with skew-$t$ exhibiting slightly superior tail conformity. Overall, both PDF and CDF analyses confirm that the GH distribution offers the best fit for the observed GNSS latitude data, effectively capturing its skewed and heavy-tailed nature.

Table 1 presents the estimated parameters, CIs, and performance metrics for four fitted models applied to GPS latitude data. Among these, the GH distribution demonstrates the best overall performance, with the highest log-likelihood ($\mathcal{L} = 1.20 \times 10^6$) and the lowest AIC and BIC values ($-2.40 \times 10^6$) for both, indicating an optimal balance between goodness-of-fit and model complexity. It also achieves a low RMSE ($2.8 \times 10^{-4}$), confirming its accuracy in approximating the observed data. The skew-$t$ model also performs competitively, with a slightly lower RMSE ($2.6 \times 10^{-4}$) and a similar log-likelihood, although its AIC and BIC values are marginally less favourable than those of the GH. The GH distribution incorporates additional parameters, $\alpha$, $\lambda$, and $\beta$, to model heavy tails, kurtosis, and skewness, respectively. Notably, the small value of $\alpha = 0.05$ suggests mild asymmetry, highlighting the distribution's suitability for capturing the non-Gaussian characteristics of the data. The skew-normal
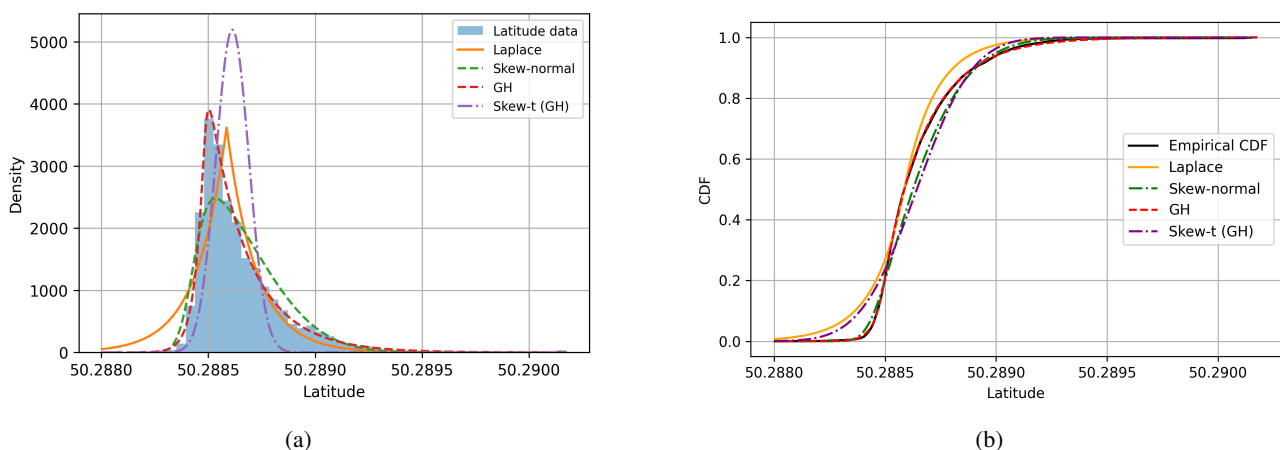


Fig. 1. Probability density function (a) and cumulative distribution function (b) for latitude data with fitted distributions using weighted MLE.

Table 1. Estimated parameters with 95 % CIs and performance metrics for the fitted models applied to GNSS latitude data.

| Model | Parameters | | | | | | CI* | $\mathscr{L}$ | AIC | BIC | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu^*$ $\times 10^{-5}$ | $\sigma$ $\times 10^{-5}$ | $\nu$ $\times 10^3$ | $\alpha$ | $\lambda$ | $\beta$ | | $\times 10^6$ | $\times 10^6$ | $\times 10^6$ | |
| Laplace | 864 | 20 | – | – | – | – | [805; 924] | 1.16 | −2.33 | −2.33 | $3.5 \cdot 10^{-4}$ |
| Skew-normal | 857 | 21 | – | 0.50 | – | – | [825; 904] | 1.13 | −2.27 | −2.27 | $2.5 \cdot 10^{-4}$ |
| GH | 845 | 0.06 | – | 0.05 | 1.39 | 0.43 | [841; 917] | 1.20 | −2.40 | −2.40 | $2.8 \cdot 10^{-4}$ |
| Skew-t | 861 | 14 | 5.0 | – | – | – | [825; 897] | 1.16 | −2.32 | −2.32 | $2.6 \cdot 10^{-4}$ |

*Note: The values marked with an asterisk ($\mu$ and CI) are expressed as least significant digits relative to a base latitude of 50.28000° N and scaled by $10^{-5}$.



(a) Laplace



(b) Skew-normal
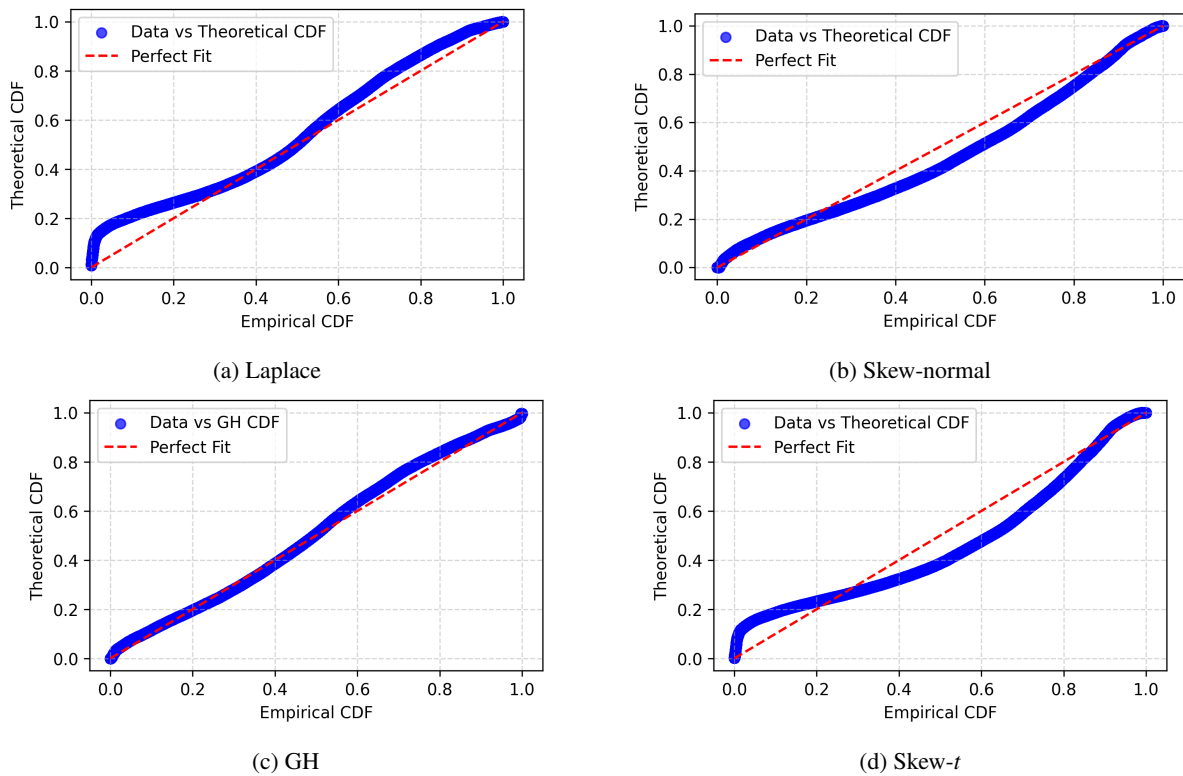


(c) GH



(d) Skew-*t*

Fig. 2. Residual analysis of the selected distributions using P–P plot.

model performs moderately well, capturing skewness through its shape parameter ($\alpha = 0.50$), but its RMSE ($2.5 \times 10^{-4}$) and lower likelihood suggest it is less flexible in modelling heavy tails. In contrast, the Laplace model performs worse, with the highest RMSE ($3.5 \times 10^{-4}$) and the lowest likelihood, failing to capture the data's asymmetry and heavy tails. Overall, the GH distribution provides the most reliable representations of the latitude data based on both fit quality and parameter efficiency. Fig. 1 and Table 1 together evaluate the statistical models' fit to GPS latitude data, supporting both qualitative and quantitative analysis.

This study uses probability–probability (P–P) plots to visually assess the agreement between empirical latitude data and theoretical distributions. These plots compare the CDF of the observed data with that of a specified theoretical model by plotting empirical probabilities against theoretical ones, making them a valuable tool for evaluating goodness-of-fit [28].

Fig. 2 shows the P–P plots for the residual analysis of four fitted distributions: Laplace, skew-normal, GH, and skew-*t*- against the empirical CDF of the latitude data. The closer the blue curve is to the red dashed line (representing a perfect fit), the better the distribution matches the empirical data. In Fig. 2(c), the GH distribution aligns most closely with the diagonal, indicating the best overall fit and demonstrating its effectiveness in capturing both skewness and kurtosis in the data. The skew-normal and skew-*t* distributions in Fig. 2(b) and (d) show a generally good fit, though slight deviations at the lower tails and centre suggest minor limitations in capturing the data's asymmetry or kurtosis. In contrast, the Laplace distribution in Fig. 2(a) deviates more noticeably from the diagonal, especially in the lower and upper tails, indicating a poor fit to the empirical data. This is consistent with Laplace's limited ability to capture asymmetry and heavy tails. Overall, the GH distribution demonstrates supe-

rior performance in residual behaviour, confirming its suitability for modelling the GPS latitude data compared to the other distributions analysed.

In summary, the P–P plots confirm that the GH distribution provides the best empirical fit, followed by the skew-*t* and skew-normal models, with Laplace performing the worst. These visual results are consistent with the quantitative metrics from the likelihood and RMSE analyses.

## 6. CONCLUSIONS

This study critically assessed four probability distributions: Laplace, skew-normal, GH, and skew-*t*- for modelling skewed and heavy-tailed data using GNSS latitude data as an example. Using the WMLE, we estimated parameters and CIs to quantify uncertainty. We also evaluated model performance using log-likelihood, AIC, BIC, and RMSE.

The analysis showed that, although the Laplace model offers a simple symmetric representation, it performed the weakest overall. Its limited ability to capture heavy tails and asymmetry resulted in the highest RMSE, lowest likelihood, and the widest CIs. The skew-normal model improved fit by capturing asymmetry and producing tighter intervals, yet it struggled with heavy tails.

The skew-*t* distribution provided a more balanced fit, effectively modelling both skewness and moderate tail behaviour. It performed well across all metrics, with only minor deviations in the lower tails and centre. The GH distribution proved most robust, closely aligning with empirical data and excelling in both statistical and visual evaluations. Its additional shape, skewness, and kurtosis parameters enabled it to capture complex distributional features, outperforming all other models. Both GH and skew-*t* showed tighter CIs, indicating more reliable parameter estimates.

The novelty of this work lies in establishing a unified statistical framework that integrates WMLE-based parameter estimation with multiple non-Gaussian models to evaluate GNSS measurement uncertainty under real-world, non-Gaussian noise conditions. Unlike previous research, which focused mainly on Gaussian assumptions, this study demonstrates that heavy-tailed and asymmetric models, particularly GH and skew-t, significantly improve distributional fit and the accuracy of uncertainty quantification.

These results highlight the importance of flexible, asymmetric models for accurate GNSS error characterisation. While RMSE quantifies overall deviation, it overlooks distributional nuances evident in models such as skew-normal, which achieve low RMSE but misrepresent tail structure. In contrast, GH and skew-*t* models deliver superior performance across all criteria. AIC and BIC further support their suitability by balancing fit and complexity, ensuring robustness and generalisability to new datasets.

This work has practical implications for GNSS-based positioning, calibration, and metrological applications, particularly in environments where impulsive or non-Gaussian noise dominates. Future work will extend this framework to dynamic GNSS scenarios and explore its integration with advanced filtering and real-time uncertainty estimation techniques.

## REFERENCES

[1] Bantu, A., Wiora, J. (2025). Analysis of fitting GNSS data provided by stationary receiver to non-gaussian distributions. In *2025 15th International Conference on Measurement*. IEEE, 266–269. https://doi.org/10.23919/MEASUREMENT66999.2025.11078743.

[2] Madry, S. (2024). *Global Navigation Satellite Systems and Their Applications, (2nd Edition)*. Springer. https://doi.org/10.1007/978-3-031-74488-4.

[3] Teunissen, P. J., Montenbruck, O. (eds.) (2017). *Springer Handbook of Global Navigation Satellite Systems*. Springer. https://doi.org/10.1007/978-3-319-42928-1.

[4] Yap, B. W., Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation,* 81 (12), 2141–2155. https://doi.org/10.1080/00949655.2010.520163.

[5] Kanhere, A. V., Gupta, S., Shetty, A., Gao, G. (2022). Improving GNSSS positioning using neural-network-based corrections. *NAVIGATION: Journal of the Institute of Navigation,* 69 (4), navi.548. https://doi.org/https://doi.org/10.33012/navi.548.

[6] Dautermann, T., Mayer, C., Antreich, F., Konovaltsev, A., Belabbas, B., Kalberer, U. (2012). Non-Gaussian error modeling for GBAS integrity assessment. *IEEE Transactions on Aerospace and Electronic Systems,* 48 (1), 693–706. https://doi.org/10.1109/TAES.2012.6129664.

[7] Dvulit, P., Savchuk, S., Sosonka, I. (2020). The processing of GNSS observation by non-classical error theory of measurements. *Geodynamics* 1 (28), 19–28. https://doi.org/10.23939/jgd2020.01.019.

[8] Wen, W., Zhang, G., Hsu, L.-T. (2022). GNSS outlier mitigation via graduated non-convexity factor graph optimization. *IEEE Transactions on Vehicular Technology,* 71 (1), 297–310. https://doi.org/10.1109/TVT.2021.3130909.

[9] Raitoharju, M., Nurminen, H., Cilden-GulerNSS, D., Särkkä, S. (2021). Kalman filtering with empirical noise models. In *2021 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE. 1–7. https://doi.org/10.1109/ICL-GNSS51451.2021.9452318.

[10] Choi, S.-Y., Yoon, J.-H. (2020). Modeling and risk analysis using parametric distributions with an application in equity-linked securities. *Mathematical Problems in Engineering,* 2020, 9763065. https://doi.org/10.1155/2020/9763065.

[11] Alzaatreh, A., Lee, C., Famoye, F., Ghosh, I. (2016). The generalized cauchy family of distributions with

applications. *Journal of Statistical Distributions and Applications,* 3 (12). https://doi.org/10.1186/s40488-016-0050-3.

[12] Aryal, G. R. (2006). *Study of laplace and related probability distributions and their applications.* Dissertation, University of South Florida, Florida, US.

[13] Puig, P., Stephens, M. A. (2000). Tests of fit for the laplace distribution, with applications. *Technometrics,* 42 (4), 417–424. https://doi.org/10.1080/00401706.2000.10485715.

[14] Agu, F., Onwukwe, C. (2019). Modified laplace distribution, its statistical properties and applications. *Asian Journal of Probability and Statistics,* 4 (1), 1–14. https://doi.org/10.9734/ajpas/2019/v4i130104.

[15] Jaber, A., Rashid, T., Rasheed, M., Mahmood, R. S., Maalej, O. (2024). Analysis of cauchy distribution and its applications. *Journal of Positive Sciences,* 4 (4). https://doi.org/10.52688/ASP54542.

[16] González-Estrada, E., Villaseñor, J. A., Acosta-Pech, R. (2022). Shapiro-Wilk test for multivariate skew-normality. *Computational Statistics,* 37 (4), 1985–2001. https://doi.org/10.1007/s00180-021-01188-y.

[17] van Wyk, A., Azzalini, A., Bekker, A. (2024). Computational aspects of likelihood-based inference for the univariate generalized hyperbolic distribution. *Communications in Statistics-Simulation and Computation,* 1–21. https://doi.org/10.1080/03610918.2024.2427225.

[18] Abramowitz, M., Stegun, I. A. (eds.) (1964). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Vol. 55. U.S. Department of Commerce, National Bureau of Standards.

[19] Ahsanullah, M., Kibria, B. G., Shakil, M. (2014). *Normal and Student's t Distributions and Their Applications.* Atlantis Press Paris, ISBN 978-94-6239-060-7.

[20] Dias-Domingues, T., Mouriño, H., Sepúlveda, N. (2024). Classification methods for the serological status based on mixtures of skew-normal and skew-t distributions. *Mathematics,* 12 (2), 217. https://doi.org/10.3390/math12020217.

[21] Borrajo, M. I., González-Manteiga, W., Martínez-Miranda, M. D. (2024). Goodness-of-fit test for point processes first-order intensity. *Computational Statistics & Data Analysis,* 194, 107929. https://doi.org/10.1016/j.csda.2024.107929.

[22] Jacinto, G., Filipe, P. A., Braumann, C. A. (2022). Weighted maximum likelihood estimation for individual growth models. *Optimization,* 71 (11), 3295–3311. https://doi.org/10.1080/02331934.2022.2075745.

[23] Ialongo, C. (2019). Confidence interval for quantiles and percentiles. *Biochemia Medica,* 29 (1), 5–17. https://doi.org/10.11613/BM.2019.010101.

[24] Langat, P. K., Kumar, L., Koech, R. (2019). Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. *Water,* 11 (4), 734. https://doi.org/10.3390/w11040734.

[25] Chakrabarti, A., Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. *Philosophy of Statistics,* 7, 583–605. https://doi.org/10.1016/B978-0-444-51862-0.50018-6.

[26] Kono, H., Kubokawa, T. (2023). Consistent Bayesian information criterion based on a mixture prior for possibly high-dimensional multivariate linear regression models. *Scandinavian Journal of Statistics,* 50 (3), 1022–1047. https://doi.org/10.1111/sjos.12617.

[27] Bantu, A., Wiora, J. (2025). Impact of non-Gaussian noise on position accuracy provided by global navigation satellite systems (GNSS). *IEEE Access,* 13, 161750–161761. https://doi.org/10.1109/ACCESS.2025.3610048.

[28] Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences, (3rd Edition).* Academic Press, ISBN 9780123850225.