



RESEARCH ARTICLE

A general extreme value-based Gaussian global navigation satellite systems measurement error distribution for mission-critical applications

Ma'mon Saeed Alghananim¹ and Washington Yotto Ochieng

Department of Civil and Environmental Engineering, Imperial College London, London, UK

Corresponding author: Ma'mon Saeed Alghananim; Email: m.alghananim17@imperial.ac.uk

Received: 21 November 2023; **Revised:** 26 February 2025; **Accepted:** 08 April 2025; **First published online:** 26 May 2025

Keywords: American democracy; anarchism; civil liberties; history of capitalism; politics of political economy

Abstract

Global Navigation Satellite Systems (GNSS) positioning and integrity monitoring models and algorithms currently generically assume that measurement errors follow a Gaussian distribution. As this is not always the case, there is a trade-off affecting system safety and availability, emphasising the need for better error characterisation in mission-critical applications. Research to date has shown advantages of Generalised Extreme Value (GEV) distribution for mapping extreme events. However, it is more complex than the Gaussian distribution, especially in the error convolution process. This paper derives a distribution, referred to as the *GEV-based Gaussian distribution*, that benefits from the advantages of both the GEV and Gaussian distributions in mapping extreme events and simplicity, respectively. The proposed distribution is tested against Gaussian, GEV and Generalised t distribution. The results show that the proposed distribution can provide a better bound for extreme events than the tested distribution both for pseudorange and carrier phase errors.

1. Introduction

Global Navigation Satellite Systems (GNSS) offer a wide range of benefits across multiple sectors such as transportation, agriculture, surveying, emergency services and telecommunications. Among GNSS applications, there are mission-critical ones that require a high level of integrity, defined as a navigation system's ability to raise an alarm when it should not be used.

At the user receiver level, integrity is traditionally monitored by a Receiver Autonomous Integrity Monitoring (RAIM) algorithm, which has two main functions: (1) detecting and if possible excluding failures; and (2) raising an alarm when the system should not be used due to the impact of a failure resulting in the estimated accuracy (Protection Level – PL) being larger than a specified the maximum allowable position error referred to as the Alarm Limit (AL). Specifically, the PL is the upper bound that position error must not exceed without being detected with a given probability, called integrity risk. The PL is computed by considering faulty measurements and their impact within the positioning domain while ensuring conservative assumptions in the positioning algorithm.

One of the main assumptions in the positioning algorithm is that the error distribution follows a Gaussian distribution. The advantages of characterising GNSS errors using a Gaussian distribution in system design and analysis (Ober et al., 2001) include: (1) representation by only two parameters, with a mean of zero and only the standard deviation requiring transmission; (2) the convolution of two Gaussian distributions resulting in Gaussian distribution, which is useful in differential GNSS positioning; (3) the widespread availability of probability and cumulative density function computations

for the Gaussian distribution in standard software packages; and (4) reduced computational complexity compared with higher-order moments, as the required parameters are easily derived from a dataset. However, GNSS errors are not always well-characterised by a Gaussian distribution, particularly in the tail regions (Panagiotakopoulos et al., 2014). Consequently, research efforts have been made to replace the Gaussian distribution by alternatives, overbounding the Gaussian distribution or using an overbounded alternative distribution.

In replacing the Gaussian distribution with alternatives, two main types of distributions have been used in GNSS error characterisation: the Generalised Extreme Value (GEV) distribution, which belongs to the Extreme Value Theory (EVT) family, and Pareto distributions. The Pareto distribution, which is based on three parameters, has been used to model GNSS errors in the positioning domain (Ahmad et al., 2014; Li et al., 2023).

The early implantation of EVT in the GNSS field included modelling tropospheric delay to account for extreme events (Collins and Langley, 1998). More recently, EVT has been used to model tropospheric delay and derive residual tropospheric delay error models through extreme value analysis techniques (Rózsa et al., 2020).

In the context of error characterisation, Panagiotakopoulos et al. (2014) proposed the use of the GEV distribution, unifying three types of EVT distributions – the Gumbel, Fréchet and Weibull distributions – demonstrating that replacing the Gaussian distribution with the GEV distribution provides: (1) a more accurate representation of residual errors and a better characterisation of extreme errors in the tail, due to the inclusion of an additional shape parameter that allows for varying rates of tail decay; (2) a less conservative safety threshold than that derived from a Gaussian-based safety threshold for a given missed detection probability; and (3) improved position accuracy due to the adoption of a more appropriate distribution.

In the field of error distribution overbounding, the first mathematical definition of overbounding, known as CDF overbounding, was introduced by DeCleene (2000), followed by paired CDF overbounding, proposed by Rife et al. (2004a). Subsequent research has focused on achieving CDF overbounding and/or paired CDF overbounding, leading to methods such as Excess-Mass CDF (EMC) and Excess-Mass PDF (EMP) (Rife et al., 2004b), Normal Inverse Gaussian (NIG) (Braff and Shively, 2005), symmetric overbounding of correlated errors (Rife and Gebre-Egziabher, 2007), and the combination of CDF bounding and paired overbounding techniques (Blanch et al., 2017; Blanch et al., 2018).

While none of these methods have fully achieved overbounding, Alghananim and Ochieng (2025) developed the first overbounding method, known as the Maximum Non-Bounded Difference (MnBD) method, capable of achieving both CDF and paired overbounding definitions. Their analysis, based on data from 20 Ordnance Survey (OS) National GNSS reference stations in the UK, revealed that the standard deviation of an overbounded Gaussian distribution can be more than 17 times greater than that of the fitted Gaussian distribution. They also applied the MnBD method to both Gaussian and GEV distributions, demonstrating that the use of an overbounded GEV distribution significantly enhances system availability.

Although the GEV distribution has advantages in mapping extreme events, it faces implementation challenges, as the convolution of GEV-distributed errors has not yet been thoroughly explored, which adds complexity to its implementation compared with the Gaussian distribution.

To take advantage of the GEV distribution in mapping extreme events, and the simplicity of the Gaussian distribution both in terms of representation and error convolution, this paper derives a distribution from the two, referred to as the *GEV-based Gaussian distribution*. The proposed distribution is evaluated using Kolmogorov–Smirnov test (KS) and graphical assessment against the traditional Gaussian, GEV and Generalised t distribution.

This paper is structured as follows. Section 2 summarises the mathematical models for the Probability Density Function (PDF) and Cumulative Density Function (CDF) for the Gaussian, GEV and Generalised t distributions, along with Maximum Likelihood estimation and KS assessment.

Section 3 details the proposed GEV-based Gaussian distribution, including its mathematical derivation. Section 4 presents the results from graphical and KS assessments.

2. Error distributions, maximum likelihood and Kolmogorov–Smirnov assessment

2.1. Error distributions

The Gaussian distribution is a symmetric distribution which is a function of two parameters: mean (μ) and standard deviation (σ). The Gaussian distribution PDF ($f(x)$) and CDF ($F(x)$) are given by

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

$$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{t-\mu}{\sigma}\right)^2} dt \quad (2)$$

The Gaussian distribution's mean, median, mode and variance are represented by

$$\text{mean} = \text{mode} = \text{median} = \mu \quad (3)$$

$$\text{variance} = \sigma^2 \quad (4)$$

The Generalised t distribution is a function of three parameters: location (μ_t), scale (σ_t) and degree of freedom (v). The Generalised t distribution PDF ($f(x)$) and CDF ($F(x)$) are expressed as

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sigma \sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} \left[\frac{v + \left(\frac{x_i - \mu}{\sigma_t}\right)^2}{v} \right]^{-\frac{v+1}{2}} \quad (5)$$

$$F(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sigma \sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} \int_{-\infty}^x \left[\frac{v + \left(\frac{t - \mu_t}{\sigma_t}\right)^2}{v} \right]^{-\frac{v+1}{2}} dt \quad (6)$$

The Generalised t distribution's mean, median, mode and variance are given by

$$\text{mean} = \mu_t, \quad v > 1 \quad (7)$$

$$\text{variance} = \sigma_t^2 \frac{v}{v-2}, \quad v > 2 \quad (8)$$

$$\text{mode} = \text{median} = \mu_t \quad (9)$$

The GEV distribution (Jenkinson, 1955) is a function of three parameters: location (μ), scale (σ) and extreme value index or shape (ξ). GEV combines three types of extreme value theory distributions: Gumbel, Fréchet and Weibull – also known as type (I), type (II) and type (III), respectively. For $\xi = 0$, the distribution is of the type (I); $\xi > 0$ is of type (II) and $\xi < 0$ is of type (III). The GEV distribution has a finite left boundary and has a positive skewness for $\xi > 0$, a finite right boundary and a positive skewness for $\xi < 0$, and extends to infinity in both directions for $\xi = 0$. GEV PDF($f(x)$) and CDF($F(x)$) are expressed as

$$f(x) = \frac{1}{\sigma} Q(x)^{\xi+1} e^{-Q(x)} \quad (10)$$

$$F(x) = e^{-Q(x)} \quad (11)$$

$$Q(x) = \begin{cases} (1 + \xi(\frac{x-\mu}{\sigma}))^{-1/\xi} & \xi \neq 0 \\ e^{-\frac{x-\mu}{\sigma}} & \xi = 0 \end{cases} \quad (12)$$

The GEV distribution's mean, median, mode and variance are expressed as

$$mean = \begin{cases} \mu + \sigma \frac{\Gamma(1-\xi)-1}{\xi} & \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & \xi = 0 \\ \infty & \xi \geq 1 \end{cases} \quad (13)$$

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \frac{1}{k} - \log n \right) \quad (14)$$

$$Variance = \begin{cases} \sigma^2 \frac{g_2 - g_1^2}{\xi^2} & \xi \neq 0, \xi < \frac{1}{2} \\ \frac{\sigma\pi^2}{6} & \xi = 0 \\ \infty & \xi \geq \frac{1}{2} \end{cases} \quad (15)$$

$$g_k = \Gamma(1 - k\xi) \quad (16)$$

$$median = \begin{cases} m + \sigma \frac{(\ln 2)^{-\xi}-1}{\xi} & \xi \neq 0 \\ \mu - \sigma \ln(\ln(2)) & \xi = 0 \end{cases} \quad (17)$$

2.2. Maximum likelihood

Maximum likelihood estimates PDF parameters (θ) by maximising the likelihood function, and is given by

$$L_n(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f(\theta|x_i) \quad (18)$$

where $(x_1, \dots, x_n)$ is the data sample, $f(x|\theta)$ the probability density function with (θ) parameters and n the sample size.

Maximum likelihood estimation is based on maximising the log-likelihood function, formed to simplify the original function. The function is given by

$$l(\theta; x_i) = \ln(L(\theta, x_i)) \quad (19)$$

The partial derivative with respect to θ can be set to zero to estimate the parameter, since the goal is to maximise the log-likelihood function.

$$\frac{\partial l}{\partial \theta_i} = 0 \quad (20)$$

Then, an iterative numerical procedure (e.g. Newton–Raphson) can be employed to solve the equations returned from the partial derivative of the log-likelihood function.

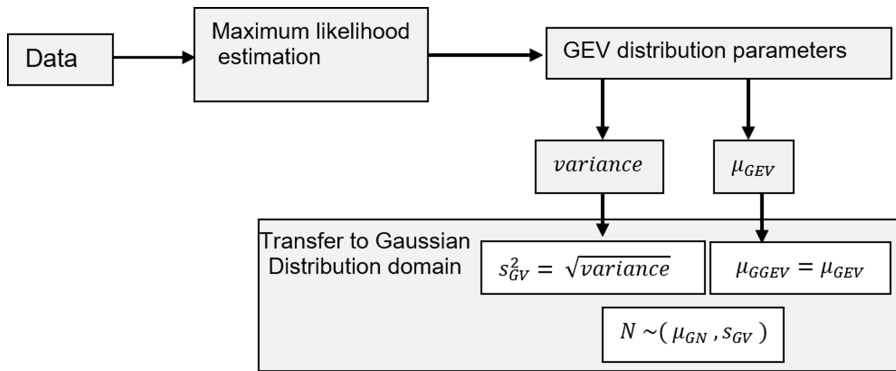


Figure 1. Derivation of GEV-based Gaussian distribution.

2.3. The Kolmogorov–Smirnov (KS) test

KS tests whether a sample of data came from a population with a specific distribution. This test is based on computing the maximum difference between the empirical CDF (eCDF) and CDF of the hypothesised distribution. The KS test is given by

$$D = \max_n(D_1, D_2) \quad (21)$$

$$D_1 = \max_n(F(x_I) - eCDF(x_I)) \quad (22)$$

$$D_2 = \max_n(eCDF(x_I) - F(x_I)) \quad (23)$$

The hypothesis that the error sample follows a given error distribution is given by

- *Null-Hypothesis* H_0 : the error sample $(x_1, \dots, x_n)$ follows a hypothesised error distribution

$$\chi^2 > \chi^2_{1-\alpha, k-s} \quad (24)$$

- *Alternate Hypothesis* H_1 : the error sample $(x_1, \dots, x_n)$ does not follow a hypothesised error distribution

$$\chi^2 \geq \chi^2_{1-\alpha, k-c} \quad (25)$$

The KS test values are employed in this paper to evaluate the overall fitting performance of the candidate distributions.

3. The derived GEV-based Gaussian distribution

GEV-based Gaussian is a Gaussian distribution generated from the GEV distribution (Figure 1). The first step in creating this distribution is to estimate the GEV distribution parameters (mean, scale parameter, shape parameter) using the maximum likelihood estimation. The GEV distribution's mean and variance are then computed from the GEV distribution's estimated parameters.

The mathematical derivation of the GEV-based Gaussian distribution starts from the likelihood function for the GEV distribution, which is represented by

$$f(x_1, \dots, x_n | \mu_{GEV}, \sigma_{GEV}, \xi) = \prod_{i=1}^n f(x_i | \mu_{GEV}, \sigma_{GEV}, \xi) \quad (26)$$

where $m_{GEV}, \sigma_{GEV}, \xi$ are the GEV distribution parameters.

Then, the log-likelihood function of the GEV distribution is represented by

$$l(\mu_{GEV}, \sigma_{GEV}, \xi; x_i) = \ln(L(\mu_{GEV}, \sigma_{GEV}, \xi | x_i)) \quad (27)$$

$$l(\mu_{GEV}, \sigma_{GEV}, \xi; x_i) = -n \ln(\sigma_{GEV}) + \sum_{i=1}^n \left[\left(\frac{1}{\xi} - 1 \right) \ln(y_i) - (y_i)^{\frac{1}{\xi}} \right] \quad (28)$$

$$y_i = 1 - \frac{\xi}{\sigma_{GEV}} (x_i - \mu_{GEV}) \quad (29)$$

To maximise the likelihood function $l(\mu_{GEV}, \sigma_{GEV}, \xi; x_i)$, let the partial derivative of the function with respect to the parameters equal zero.

The partial derivative of the function with respect to the mean is represented, step by step, by

$$\frac{\partial l}{\partial \mu_{GEV}} = 0 = \sum \left(\frac{1}{y_i} \left(\frac{1}{\xi} - 1 \right) - \frac{1}{\xi} y_i^{\left(\frac{1}{\xi} - 1 \right)} \right) \cdot \frac{\xi}{\sigma_{EV}} \quad (30)$$

$$\frac{\partial l}{\partial \mu_{GEV}} = 0 = \frac{1}{\sigma_{GEV}} \sum \frac{1}{y_i} \left(\left(\frac{1}{\xi} - 1 \right) - \frac{1}{\xi} y_i^{\left(\frac{1}{\xi} \right)} \right) \cdot \xi \quad (31)$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma_{GEV}} \sum \frac{1 - \xi - y_i^{\frac{1}{\xi}}}{y_i} = 0 \quad (32)$$

The partial derivative of the function with respect to the shape parameter is represented by

$$\frac{\partial l}{\partial \xi} = \frac{\partial l}{\partial y} \cdot \frac{\partial y}{\partial \xi} \quad (33)$$

$$\frac{\partial l}{\partial \xi} = \frac{-1}{\xi^2} \sum \left(\frac{1}{y_i} \ln(y_i) \left(1 - \xi - y_i^{\frac{1}{\xi}} \right) \right) + \frac{1 - \xi - y_i^{\frac{1}{\xi}}}{y_i} \xi \left(\frac{x_i - \mu_{GEV}}{\sigma_{GEV}} \right) = 0 \quad (34)$$

To simplify the partial derivative of the function with respect to the sigma, the likelihood function $l(\mu_{GEV}, \sigma_{GEV}, \xi; x_i)$ is rewritten as

$$l(\mu_{GEV}, \sigma_{GEV}, \xi; x_i) = l_1(n, \sigma_{GEV}) + l_2(y_i, \xi) \quad (35)$$

From Equation (36), the partial derivative of the likelihood function with respect to the sigma is derived step by step as follows:

$$\frac{\partial l}{\partial \sigma_{GEV}} = \frac{\partial l_1}{\partial \sigma_{GEV}} + \frac{\partial l_2}{\partial y_i} \cdot \frac{\partial y_i}{\partial \sigma_{GEV}} \quad (36)$$

$$\frac{\partial l_1}{\partial \sigma_{GEV}} = \frac{-n}{\sigma_{GEV}} \quad (37)$$

$$\frac{\partial l_2}{\partial y_i} \cdot \frac{\partial y_i}{\partial \sigma_{GEV}} = \sum \left(\left(\frac{1}{\xi} - 1 \right) \times \frac{1}{y_i} - \frac{1}{\xi} y_i^{\frac{1}{\xi} - 1} \right) \cdot \frac{\xi (x_i - \mu_{GEV})}{\sigma_{GEV}^2} \quad (38)$$

$$\frac{\partial l_2}{\partial y_i} \cdot \frac{\partial y_i}{\partial \sigma_{GEV}} = \sum \frac{1 - \xi - y_i^{\frac{1}{\xi}}}{y_i} \cdot \left(\frac{x_i - \mu_{GEV}}{\sigma_{GEV}^2} \right) \quad (39)$$

$$\frac{\partial l}{\partial \sigma_{GEV}} = \frac{\partial l}{\partial \sigma_{GEV}} + \frac{\partial l}{\partial y} \cdot \frac{\partial y}{\partial \sigma_{GEV}} = 0 \quad (40)$$

$$\frac{\partial l}{\partial \sigma_{GEV}} = \frac{-n}{\sigma_{GEV}} + \sum \frac{1 - \xi - y_i^{\frac{1}{\xi}}}{y_i} \cdot \left(\frac{x_i - \mu_{GEV}}{\sigma_{GEV}^2} \right) = 0 \quad (41)$$

From Equations (32), (35) and (41), the three GEV parameters are estimated using the Newton–Raphson method. Then, the GEV distribution variance is computed using the estimated parameters, represented by

$$\text{Variance} = \begin{cases} \frac{\sigma_{GEV}(g_2 - g_1^2)}{\xi^2} & \xi \neq 0, \xi < \frac{1}{2} \\ \sigma_{GEV}^2 \times \frac{\pi^2}{6} & \xi = 0 \\ \infty & \xi \geq \frac{1}{2} \end{cases} \quad (42)$$

$$g_k = \Gamma(1 - k\xi) \quad (43)$$

Then, the GEV-based Gaussian distribution parameters $(\mu_{GbG}, \sigma_{GbG})$ are computed by

$$\mu_{GbG} = \mu_{GEV} \quad (44)$$

$$\sigma_{GbG} = \begin{cases} \sqrt{\frac{\sigma_{GEV}(g_2 - g_1^2)}{\xi^2}} & \xi \neq 0, \xi < \frac{1}{2} \\ \sqrt{\sigma_{GEV}^2 \times \frac{\pi^2}{6}} & \xi = 0 \\ \infty & \xi \geq \frac{1}{2} \end{cases} \quad (45)$$

4. Results

This section provides a summary of the results from Graphical and KS assessments. The GEV-based Gaussian distribution is tested against the Gaussian, GEV and Generalised t distributions.

In total, 26 datasets were collected over 26 different days in dynamic mode and in mixed environments (urban, semi-urban and open sky). The data were collected using a vehicle-mounted multi-constellation GNSS receiver (Ublox's F9 receiver) and Trimble's Applanix reference system. Data collection was carried out in Switzerland, spanning a route of approximately 3,455 km through various cities and regions, as illustrated in the map presented in Figure 2. The data encompass a variety of environments: highways represent open-sky conditions, sometimes bordering on semi-urban areas, while cities like Zürich and Thalwil exemplify urban environments. The vehicle maintained an average speed of approximately 70 km/h, with a minimum of 42 km/h and a maximum of 142 km/h.

The position, velocity, heading and accuracy data were generated using an Applanix reference system and Applanix post-processing software, with an average among all datasets with horizontal accuracy of 4 cm and vertical accuracy of 6 cm. Given the high level of accuracy from the reference system, the range computed from this system was considered the true value. Subsequently, measurement errors were estimated by comparing data from the GNSS receiver with the reference system resulting in more than 2 million rows of pseudorange and carrier phase errors. The GNSS receiver employed divergence-free carrier smoothing techniques, and the truth system's results were used as a baseline to estimate the residuals.

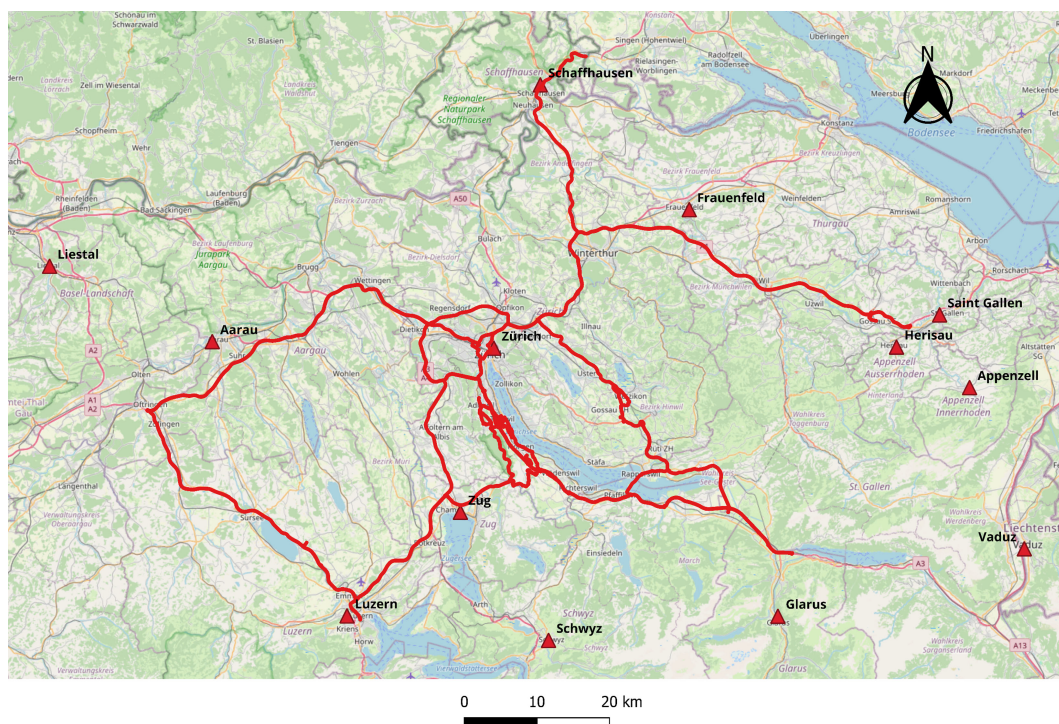


Figure 2. Map of the data collection route (~3,455 km) used in this study.

4.1. Kolmogorov–Smirnov test assessment

KS assessment tests the GEV-based Gaussian distribution for overall fitting, which reflect the positioning reliability. Since the test is based on the maximum difference between CDF and ECDF, it gives an indication of the overall fitting, which might not be sufficient to evaluate the goodness-of-fit in some cases.

By comparing the KS test values, the results show that the Generalised t distribution is the best of the three candidate distributions in terms of fitting pseudorange and carrier phase error, as seen in all datasets, as shown in Table 1.

Mixed results are seen when the remaining three distributions are compared: Gaussian, GEV and GEV-based Gaussian distribution. Table 2 shows the number of times that the value of the test is the lowest between the three distributions for carrier phase and pseudorange error datasets.

In more detail, for carrier phase error, the KS test values show that the GEV-based Gaussian distribution was the best in terms of fitting between the three distributions in 12 datasets, followed by the GEV distribution, which was the best in 11 datasets, while Gaussian was best in three datasets. For pseudorange error, the results show that the Gaussian distribution is the best with 20 datasets, followed by the GEV-based Gaussian distribution with 6 datasets.

Overall, as shown in Tables 1 and 2, the results overall show that Generalised t is the best in fitting the error data, as the average value of KS test in Generalised t is the lowest. The table also shows that GEV-based Gaussian shows better performance than Gaussian and GEV in fitting carrier phase error (complex data sets), while the Gaussian distribution provides better fitting than GEV and GEV-based Gaussian in pseudorange datasets.

Table 1. Average, minimum and maximum KS test values for the 26 datasets.

	Statistic	Generalised t	Gaussian	GEV	GEV-based Gaussian
Pseudorange error	Average	0.044	0.320	0.375	0.345
	Maximum	0.078	0.369	0.417	0.414
	Minimum	0.024	0.269	0.316	0.320
Carrier phase error	Average	0.076	0.481	0.479	0.474
	Maximum	0.123	0.522	0.546	0.527
	Minimum	0.046	0.438	0.425	0.440

Table 2. Number of datasets so that the value of the test is the lowest between three distributions for carrier phase and pseudorange error datasets.

	Gaussian	GEV	GEV-based Gaussian
Pseudorange error	20	0	6
Carrier phase error	3	11	12

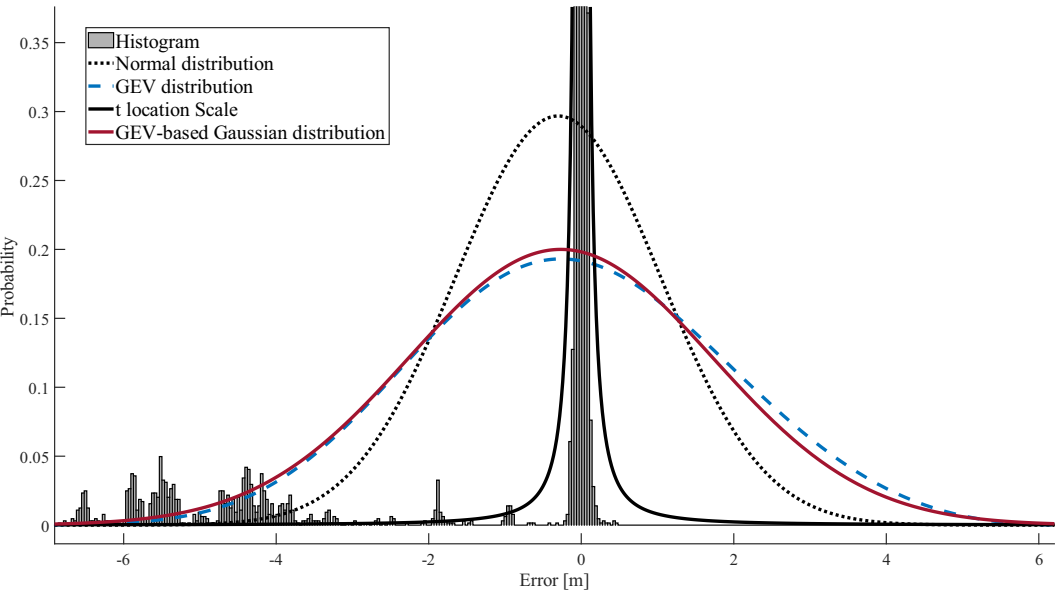


Figure 3. Carrier phase error characterisation, in the PDF domain, for dataset 1.

4.2. Graphical assessment

The graphical assessment in this section evaluates the GEV-based Gaussian distribution against the three distributions in terms of tail fitting and overbounding, both of which are of special importance for mission-critical applications. The results show that none of the distributions overbound the empirical distribution. However, the GEV-based Gaussian and GEV distributions provide a better bounding at the tails than the Gaussian and Generalised t distributions.

In more detail, for carrier phase error, the results show that the GEV-based Gaussian distribution is the best in terms of overall overbounding, as seen in almost all datasets. This is followed by the GEV distribution, which also shows in many cases slightly better performance in terms of overbounding than the GEV-based Gaussian distribution at the right tails. Figures 3–8 show that the GEV family

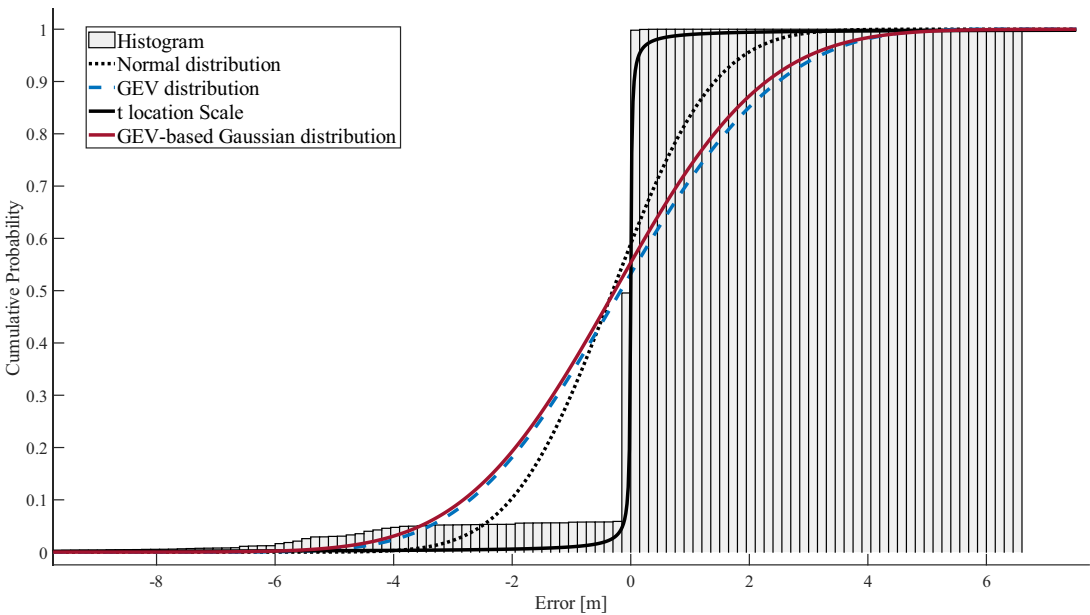


Figure 4. Carrier phase error characterisation, in the CDF domain, for dataset 1.

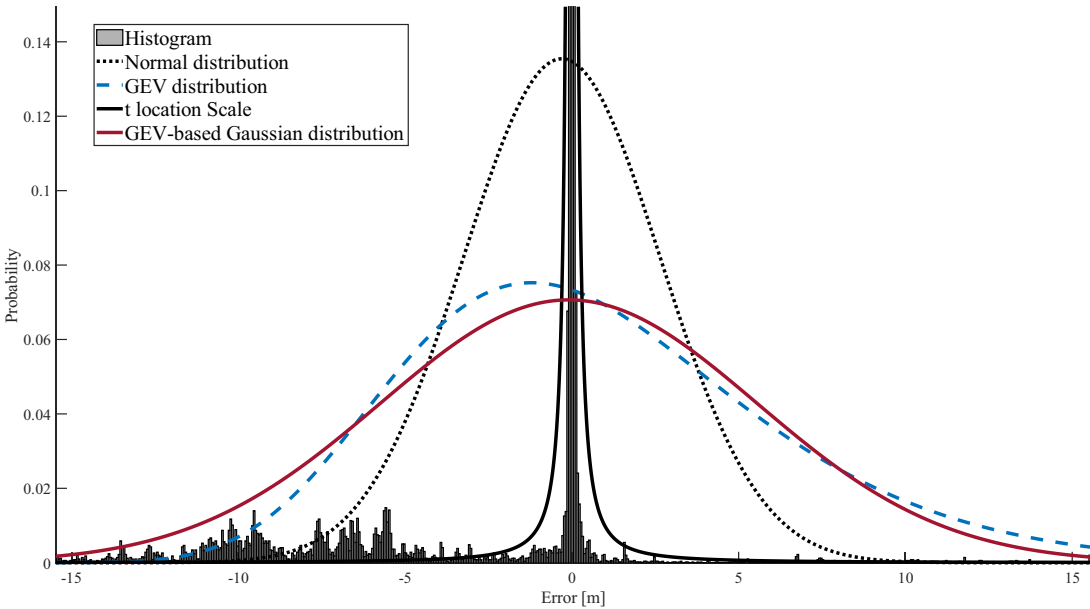


Figure 5. Carrier phase error characterisation, in the PDF domain, for dataset 2.

distribution overbounds more extreme events. This is clearly seen in the CDF left tail, where more extreme events are covered under the distribution tails.

For pseudorange error, the results show that the GEV and GEV-based Gaussian distributions are the best for bounding. The GEV distribution shows a slightly better performance than the GEV-based Gaussian distribution for overall overbounding, although the latter is best in terms of bounding in the left tail.

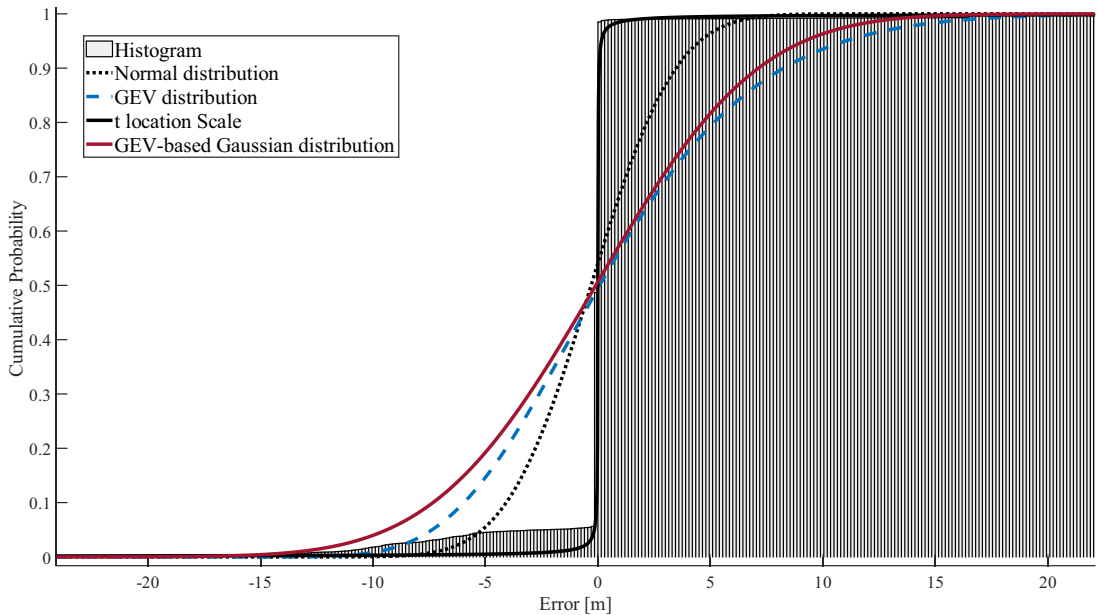


Figure 6. Carrier phase error characterisation, in the CDF domain, for dataset 2.

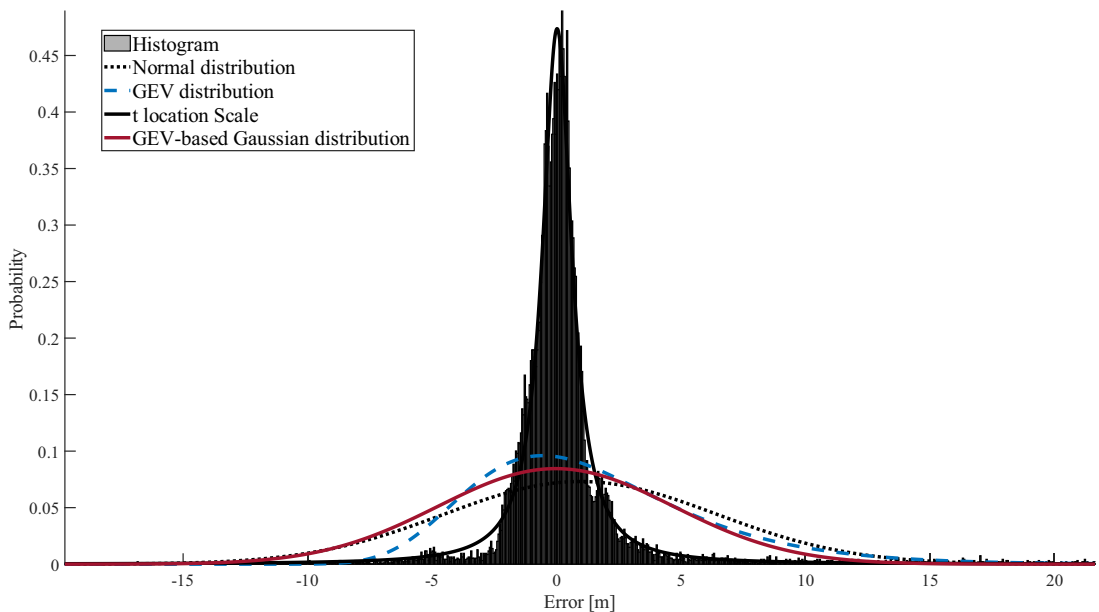


Figure 7. Pseudorange error characterisation, in the PDF domain, for dataset 1.

In addition, the results also show that the Generalised t distribution provides the best fitting for the majority of the datasets. Mixed results are observed when comparing the remaining three distributions: Gaussian, GEV and GEV-based Gaussian distribution. For tail fitting, the results show that GEV and GEV-based Gaussian distributions are the best of the four distributions.

Overall, the results show that the GEV-based Gaussian distribution is the best among tested distributions in terms of overbounding carrier phase error, followed by the GEV distribution. The

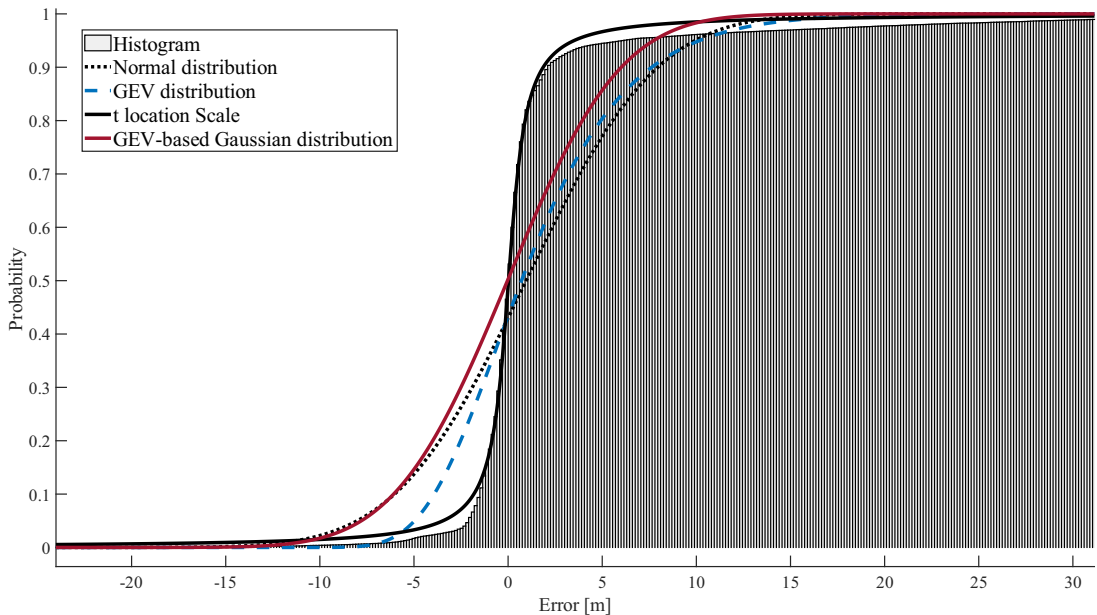


Figure 8. Pseudorange error characterisation, in the CDF domain, for dataset 1.

corresponding distributions for pseudorange error are GEV and GEV-based Gaussian. Furthermore, both the GEV distribution and the GEV-based Gaussian distribution exhibit superior tail fitting compared with the other two distributions, while Generalised t provides the best fitting for the core.

4.3. Summary

The proposed distribution has been tested against three distributions (Gaussian, GEV and Generalised t) in two aspects: fitting (overall and tail) and overbounding.

The KS and graphical assessment show that Generalised t provides the best overall fitting for the pseudorange and carrier phase error among the four distributions, followed by GEV-based Gaussian for carrier phase error and Gaussian for pseudorange error. Furthermore, the graphical assessments show that the GEV family (GEV and GEV-based Gaussian) provides the best fitting for the tail.

In addition, the graphical assessment shows that the GEV family provides better overall bounding for pseudorange and carrier phase error. Specifically, the GEV-based Gaussian distribution exhibits better overbounding for the carrier phase error, while the GEV distribution shows better overbounding results for the pseudorange error.

5. Conclusions

This paper proposes a derived distribution generated from the GEV and Gaussian distributions, referred to as the GEV-based Gaussian distribution, to take advantage of both distributions. The results have shown that the distribution combines the GEV distribution's ability in mapping extreme events, with Gaussian distribution simplicity in computation and error propagation.

The GEV-based Gaussian distribution has been tested using Kolmogorov–Smirnov test (KS), and graphical assessment against Gaussian, GEV distribution and Generalised t .

The assessments have shown that none of the tested distributions overbounds the empirical distribution. However, the results show that the novel GEV-based Gaussian distribution is the best in terms of bounding extreme events in the carrier phase error datasets, followed by the GEV distribution.

Furthermore, the GEV distribution is the best in terms of bounding the pseudorange error, followed by the GEV-based Gaussian distribution.

Due to the GEV distribution's complexity compared with the Gaussian distribution in the error convolution process, this paper proposes using GEV-based Gaussian distribution for monitoring GNSS integrity for mission-critical applications, while accounting for the needs for overbounding.

The GEV-based Gaussian distribution can be applied to both system-level and user-level integrity monitoring. At the system level, errors can be characterised using this distribution through integrity monitoring system reference stations and Continuously Operating Reference Station (CORS) networks, with the distribution's parameters transmitted to users via satellite-based or ground-based augmentation systems. Users can then use this information to enhance positioning and navigation safety compared with conventional Gaussian-based approaches. Research is ongoing to evaluate the impact of using the GEV-based Gaussian distribution in the positioning domain.

Availability of data and materials. The data used in this study were provided by u-blox.

Acknowledgements. We extend our sincere gratitude to u-blox for their invaluable contribution to our research. Their provision of high-quality GNSS data was instrumental in the successful completion of our study.

Authors' contributions. Ma'mon spearheaded all research activities for this study. He meticulously documented the research results and was instrumental during the development phase. Ma'mon also took the lead in executing the prototypes, ensuring they were in harmony with the research insights.

Washington served as a guiding force, supervising Ma'mon throughout the research journey. His invaluable guidance, consistent feedback and vigilant oversight were pivotal in ensuring the research adhered to the highest standards and objectives, solidifying the integrity and calibre of the work.

Funding statement. No external funding was secured for this research.

Competing interests. Ma'mon Alghananim and Washington Ochieng declare that there are no competing interests related to the content or publication of this work.

Ethics approval and consent to participate. Given the nature of this research, it does not involve human participants, human data or human tissue. Consequently, no ethics approval was required as per applicable institutional and national guidelines and regulations. Similarly, the concept of 'consent to participate' is not relevant to the context of this study since it involves the analysis of technical data that does not contain any personal or sensitive information.

Consent for publication. All authors have reviewed the manuscript and consent to its publication.

References

- Ahmad, K. A. B., Sahmoudi, M. and Macabiau, C. (2014). Characterisation of GNSS Receiver Position Errors for User Integrity Monitoring in Urban Environments. In *ENC-GNSS 2014, European Navigation Conference*.
- Alghananim, M. S. and Ochieng, W. (2025). Maximum non-bounded difference method for overbounding Global Navigation Satellite System errors. *GPS Solutions*, **29**(22), article number 22. <https://doi.org/10.1007/s10291-024-01772-x>.
- Blanch, J., Walter, T. and Enge, P. (2017). A MATLAB toolset to determine strict Gaussian bounding distributions of a sample distribution. In *Proceedings of the 30th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*, 4236–4247.
- Blanch, J., Walter, T. and Enge, P. (2018). Gaussian bounds of sample distributions for integrity analysis. *IEEE Transactions on Aerospace and Electronic Systems*, **55**(4), 1806–1815. <https://doi.org/10.1109/TAES.2018.2801785>.
- Braff, R. and Shively, C. (2005). A method of over bounding ground based augmentation system (GBAS) heavy tail error distributions. *The Journal of Navigation*, **58**, 83–103.
- Collins, J. P. and Langley, R. B. (1998). The residual tropospheric propagation delay: How bad can it get? In *Proceedings of ION GPS 1998, 11th International Technical Meeting of the Satellite Division of the Institute of Navigation*, 729–738.
- Declene, B. (2000). Defining pseudorange integrity-overbounding. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2000)*, 1916–1924.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, **81**(348), 158–171.
- Li, L., Li, R., Wang, L., Wang, R., Li, M. and Li, M. (2023). GNSS integrity risk evaluation in position domain based on the Generalised Pareto distribution. *Measurement Science and Technology*, **34**(9), 095010.
- Ober, P.B., Farnsworth, R., Breeuwer, E. and Van Willigen, D. (2001). SBAS integrity verification. In *Proceedings of ION GPS 2001*, 1805–1812.

- Panagiotakopoulos, D., Majumdar, A. and Ochieng, W.Y.** (2014). Extreme value theory-based integrity monitoring of global navigation satellite systems. *GPS Solutions*, **18**(1), 133–145.
- Rife, J. and Gebre-Egziabher, D.** (2007). Symmetric overbounding of correlated errors. *Navigation*, **54**, 109–124.
- Rife, J., Pullen, S., Pervan, B. and Enge, P.** (2004a). Paired overbounding and application to GPS augmentation. Position Location and Navigation Symposium. In *PLANS 2004, Position Location and Navigation Symposium (IEEE Cat. No. 04CH37556)*. IEEE, 439–446.
- Rife, J., Walter, T. & Blanch, J.** (2004b). Overbounding SBAS and GBAS error distributions with excess-mass functions. In *Proceedings of the International Symposium on GNSS/GPS, Sydney, Australia*.
- Rózsa, S., Ambrus, B., Juni, I., Ober, P. B. and Mile, M.** (2020). An advanced residual error model for tropospheric delay estimation. *GPS Solutions*, **24**(4), 103. <https://doi.org/10.1007/s10291-020-01014-5>.