

Extreme Value Theory based Integrity Monitoring of Global Navigation Satellite Systems

Dr Dimitri Panagiotakopoulos
Dr Arnab Majumdar
Professor Washington Ochieng

Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London

ABSTRACT

Measurements consistency based Receiver Autonomous Integrity Monitoring (RAIM) is the main technique for monitoring the integrity of Global Satellite Navigation Systems (GNSS) at the user level. Existing RAIM algorithms utilize two tests, in the position domain a test for RAIM availability and in the measurement domain a test for failure detection. These tests involve the computation of three parameters: test statistic, decision threshold and protection level. The test statistic is based on the actual measurements in the form of the Sum of the Squared Errors (SSE). The decision threshold is chosen on the basis of the statistical characteristics of the SSE including the assumption that the errors are normally distributed. However, in practice residual error distributions exhibit heavier tails than predicted by the Gaussian model. Therefore, this paper challenges the normality assumption of the residual navigation errors in three ways. First, real data are used to assess its impact on the traditional RAIM algorithm. Second, Extreme Value Theory (EVT) is applied to the tails and the Generalized Extreme Value (GEV) distribution is derived to capture residual navigation errors. Third, the performance of the traditional RAIM approach is compared with that employing the GEV distribution. The results demonstrate that the GEV model is a more accurate representation of the distribution of residual navigation errors than the conventional Gaussian model and should be used in the development of integrity monitoring algorithms.

1. INTRODUCTION

Mission critical applications such as aviation require that the highest level of safety and other performance standards are satisfied. In the case of navigation, these standards are known as the Required Navigation Performance (RNP) parameters. The RNP is specified by four parameters: *accuracy*, *integrity*, *continuity* and *availability*, with integrity being the parameter directly linked to safety. Integrity is a measure of the trust that can be placed in the correctness of the information provided by a navigation system. It includes the ability of the system to provide timely and valid warnings to users when it must not be used for the intended operation. In particular, a navigation system is required to deliver an alert or warning of any malfunction as a result of a set alert limit being exceeded, to users within a specified time-to-alert. Integrity risk, also known as the probability of misleading information, is defined as the probability that the positioning error exceeds the alert limit and that the event is undetected. Integrity can be provided at system level through an independent network of monitoring stations and a dedicated integrity channel and autonomously at the sensor/receiver level (Feng et al. 2006). The former exploits augmentation systems, which relay integrity information to the users from Ground-Based Augmentation Systems (GBAS) or Space-Based Augmentation Systems (SBAS). Receiver or user level integrity monitoring is accomplished by Receiver Autonomous Integrity Monitoring (RAIM) and its variations.

RAIM is based on statistical consistency checks using redundant measurements. It involves two tests; the first determines if the conditions exist to execute a RAIM calculation, i.e. RAIM availability. If RAIM is available, the second test detects the presence of a failure. Literature on RAIM algorithms is dominated by statistics on failure detection by least squares or parity methods, with the threshold determined by using a constant probability of false alert, P_{FA} (Brown 1996). The test for RAIM availability is formulated by projecting the Minimum Detectable Bias (MDB) to the position domain to realize a Protection Level (PL), which is compared with the Alert Limit (AL). The MDB is determined by the threshold of the failure detection test and the probability of missed detection (P_{MD}). It is this threshold which is the key factor that links the two tests and plays an important role in a trade-off analysis between integrity risk and continuity risk. Fig. 1 shows the general RAIM architecture.

The decision threshold is chosen on the basis of statistical characteristics of the test statistic and is dependent either on P_{FA} or P_{MD} . The most common test statistic used in RAIM is the Sum of the Squared Errors (SSE) from the computed residual estimation errors. These errors are assumed to be normally distributed, and the corresponding probabilities P_{FA} and P_{MD} are assumed to follow a Chi-squared and noncentral Chi-Squared distribution respectively (Ober 2003). Whilst the normality assumption simplifies the calculations, there is evidence that in reality the residual error distributions exhibit heavier tails than predicted by the Gaussian model (DeCleene 2000; Walter et al. 2002). Therefore, the conventional RAIM algorithm fails to efficiently compute the decision threshold as both the chi-square

and noncentral chi-square distributions are invalid. In these situations, a better characterization of the tails of the residual distributions is required. This, in turn, requires consideration of robust statistical theory applied to the modeling of the tails of distributions.

Section 2 briefly reviews RAIM algorithms. This is followed in Section 3 by a discussion of the Gaussian assumption. Section 4 introduces Extreme Value Theory (EVT). Section 5 presents the data and statistical methodology used. The main results are given in Section 6. The implications of these results are discussed in Section 7, and conclusions drawn in Section 8.

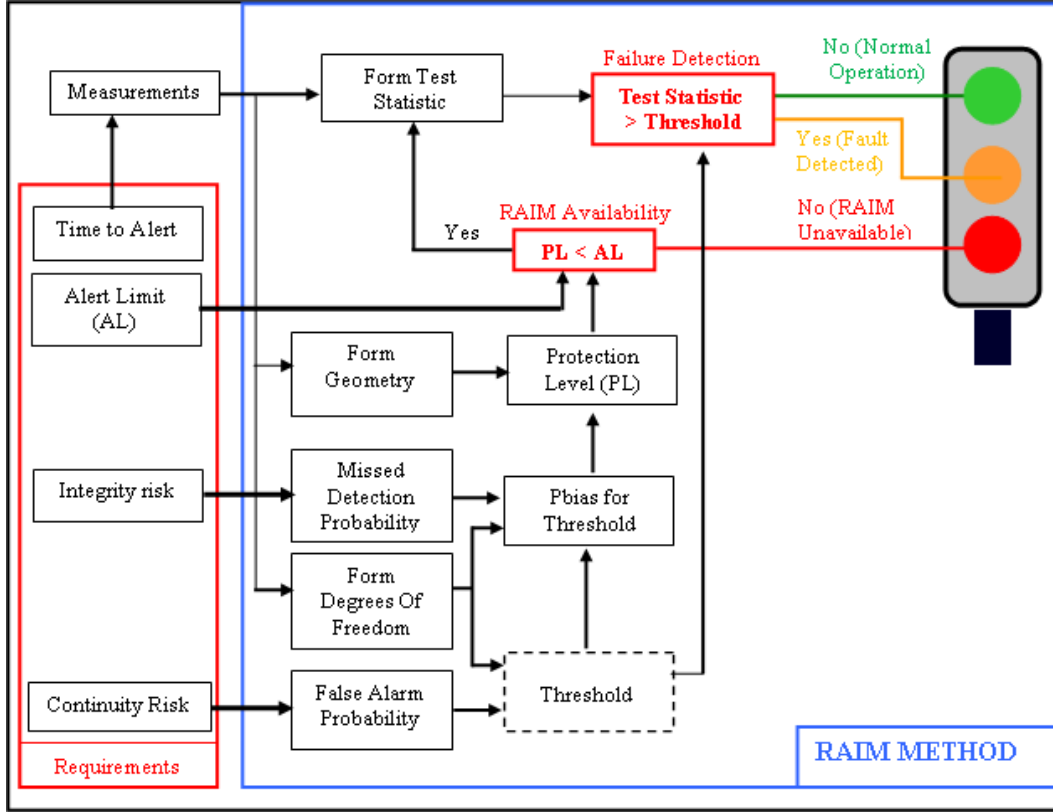


Fig. 1: General RAIM Architecture (adapted from Feng et al., 2006)

2. RAIM ALGORITHMS

The two main classes of RAIM algorithms are *snapshot* and *filtering*. The former uses only the current measurements while the latter uses both current and historical data, together with *a priori* assumptions on dynamics. The major drawback of filtering methods is their insensitivity to slowly growing errors (Bhatti, 2008). The snapshot scheme is the most widespread method for RAIM due to its faster response to sudden failures compared to the filtering scheme. It is the most commonly used method in basic stand-alone GNSS and has no requirement to make any questionable assumptions about how a system arrives at its current state.

Snapshot RAIM algorithms include range comparison, least squares residuals, parity and Marginally Detectable Error (MDE) methods (Ochieng et al. 2002). These methods are equivalent, the major issues being the test statistics and decision thresholds, and computational complexity (Brown 1992). Extensions have been developed for integration of GPS with other sensors.

Given the basic linearized GNSS standalone positioning model:

$$y_{nx1} = G_{nx4} x_{4x1}^{true} + \varepsilon_{nx1} \quad (1)$$

Where:

y is the difference between the pseudorange measurement and the predicted parameter based on a nominal user position and clock bias; it is a $nx1$ vector with n being the number of measurements,

G is the observation or design matrix;

x^{true} is the vector of the unknown parameters; and $\varepsilon = [\varepsilon_1 \varepsilon_2 \dots \varepsilon_n]^T$ is the measurement error vector caused by receiver noise, wave propagation, ephemeris etc.

The least squares solution for the estimation of x is given by:

$$\hat{x} = (G^T G)^{-1} G^T y = Ay \quad (2)$$

The residuals can then be obtained as:

$$w = y - \hat{y} = [I - G(G^T G)^{-1} G^T] \varepsilon = S \varepsilon \quad (3)$$

where $A = (G^T G)^{-1} G^T$ and $S = I - G(G^T G)^{-1} G^T$.

Then the SSE can be obtained from the residuals by:

$$SSE = w^T w = \varepsilon^T S^2 \varepsilon = \varepsilon^T S \varepsilon \quad (4)$$

The SSE is the basic observation in RAIM algorithms. It has a chi-square distribution with $n - 4$ degrees of freedom assuming that the measurement errors are independent and normally distribution with zero mean. Two variations of the SSE are used as test statistics (Brown1996).

$$test\ statistic = \sqrt{\frac{SSE}{n-4}} \quad (5)$$

$$test\ statistic = \sqrt{SSE} \quad (6)$$

The two test statistics are equivalent except that expression (5) takes account of the need for redundancy by ensuring that there are at least five satellites visible. Expression (5) is the baseline RAIM algorithm used as the reference for the comparative assessments with the new algorithms developed in this paper.

The probability of false alarm P_{FA} and the probability of missed detection P_{MD} are (Brown, 1996):

$$P_{FA} = P(\text{test statistic} \geq T / H_0) \quad (7)$$

$$P_{MD} = P(\text{test statistic} < T / H_1) \quad (8)$$

The decision threshold (T) can be computed deterministically using the RNP parameters. The corresponding thresholds are thus:

$$T_{FA} = \left(\text{inv} \left[\chi_{n-4}^2 (1 - P_{FA}) \right] \cdot \frac{\sqrt{n-4}}{\sigma} \right)^2 \quad (9)$$

$$T_{MD} = \left(\text{inv} \left[\chi_{n-4}^2 (f \cdot \text{SLOPE}_{HMAX} / \sigma)^2 \right] (P_{MD}) \right) \cdot \frac{\sqrt{n-4}}{\sigma} \quad (10)$$

where $\text{SLOPE}_{HMAX} = \text{Max}_i \left(\sqrt{(\mathbf{A}_{1i}^2 + \mathbf{A}_{2i}^2)(n-4) / \mathbf{S}_{ii}} \right)$ is the geometry factor representing the ratio of the horizontal position error to the test statistic used to project the MDE from the measurement domain to the position, \mathbf{A}_{ji} denotes the j^{th} row and i^{th} column element of the \mathbf{A} matrix, \mathbf{S}_{ii} denotes the i^{th} element of the diagonal matrix \mathbf{S} of (2), xAL is the alert limit, and f ($f \leq 1$) is a conservative factor introduced due to the uncertainty in SLOPE_{HMAX} .

Ideally, both the P_{FA} and P_{MD} should be minimized, i.e. keep both continuity and integrity risks at a minimum. However, this is not possible as these risks work against each other, in the sense that when decreasing integrity risk for example, continuity risk increases and vice versa. This is because each risk is tightly linked to a different type of error of the hypothesis test. Thus the threshold should ideally be between the T_{FA} and T_{MD} . An extension to the baseline algorithm is the Weighted Least Squares RAIM (WLS-RAIM) technique

3. THE GAUSSIAN ASSUMPTION: LIMITATIONS AND ALTERNATIVES

The Gaussian error distribution assumption is used widely in navigation and only two parameters are needed to characterize the errors: *mean* and *variance*. The *Central Limit Theorem* (CLT), which has a powerful influence in the acceptance of the Gaussian distribution, states that the mean of a large number of independent random variables in this case the residual errors associated with different sources including ionosphere, troposphere and multipath, tends to a Gaussian distribution, regardless of their original distribution. This leads to various other common distributions e.g. Chi-Square, Fisher and Student, which were derived as functions of Gaussian variables. However, the use of theoretical distributions is limited by concerns about their tails, which characterize gross errors or blunders, i.e. infrequent but large

errors. Both the size and shape of the tails are important in risk estimation. For example, it has been shown that collision risk is heavily influenced by the values at the very edge of the tails (Brooker 2004). The issue with efficiently characterizing the tails stem from the fact that more data is required to make useful and confident statements about the error distributions.

Many researchers have noted that the Gaussian model fails to efficiently characterize large errors that occur more frequently than predicted by the model for residual navigation errors (Parker 1972; Hsu 1979). Alternative probability distributions have been suggested to describe these residual errors, including *exponential* and *double exponential* (Laplace) (Anderson and Ellis 1971; Shively 2000). These are more pessimistic in the tails than the Gaussian model and therefore, provide a safer assumption. Such distributions characterize altimetry system errors and horizontal errors encountered in the navigation of aircraft (Moek et al. 2001). However, the lack of data makes it impossible to rigorously validate these models in practice and the data requirements to fully describe the probability distribution of residual navigation errors with a high degree of confidence is a major issue in navigation research. Therefore, instead of characterizing the error distribution itself, the navigation community has explored the provision of integrity by the method of overbounding. This consists of defining a distribution more conservative at the tails than the actual one, i.e. with more mass at the tails. The technique was first introduced by DeCleene (2000) and gave rise to numerous alternatives, including the Paired Overbounding and Excess Mass Functions Overbounding (Rife et al. 2004) among others. However, while these models provide solutions to specific situations, none offers a credible universal alternative to the Gaussian model (Panagiotakopoulos 2009).

4. THE GEV DISTRIBUTION

In the CLT, the normal distribution is the limiting distribution of sample averages. The Extreme Value Theory (EVT) represents a parallel idea in which a class of extreme value distributions characterizes the possible distributions of sample maxima. The Fisher-Tippet theorem (Fisher and Tippet 1928) is analogous to the CLT and uses a tail index to unify the possible characterizations of the density function of an extreme value distribution. These functions represent three possibilities for the decay of the density function in the tail (Embrechts et al. 1997). The tail can decay (i) exponentially with all finite moments as in the standard cases of the normal, log normal, and gamma, (ii) by a power as in the stable Paretian, Cauchy and Student's t distributions, and (iii) with a finite tail index. These three possibilities can be mapped onto different distributions, called Gumbel, Frechet and Weibull, which can be combined in a single functional form called the Generalized Extreme Value (GEV) distribution. The cumulative distribution function (CDF) of the GEV is given by:

$$GEV(x|\xi, \mu, \sigma) = \begin{cases} e^{-\left\{1 + \xi \frac{x - \mu}{\sigma}\right\}^{-\frac{1}{\xi}}} & \xi \neq 0 \\ e^{-e^{-(x - \mu)/\sigma}} & \xi = 0 \end{cases} \quad (11)$$

The distribution function $GEV(x|\xi, \mu, \sigma)$ is referred to as the GEV distribution with shape parameter $\xi = P$, location parameter $\mu \in P$ and scale parameter $\sigma > 0$. The case $\xi = 0$ is interpreted as $\xi \rightarrow 0$, resulting in the Gumbel-type distribution which is thin-tailed as its tail rapidly thins out. The case $\xi < 0$ corresponds to the Weibull with no tail as after a certain point there are no extremes, and $\xi > 0$ to the Frechet distribution, fat tailed as its tail slowly thins out. The standard case is obtained for $\mu = 0$ and $\sigma = 1$. Intuitively, residual navigation errors should follow either a Gumbel or Frechet-type distributions. This is because the former does not have any upper or lower endpoints and is defined for the entire real line, and the latter has a heavy-tailed distribution. Although the Gumbel distribution is considered as thin-tailed, it has a heavier tail than the Gaussian distribution. In the literature, the Gumbel distribution is commonly referred to as the 'Extreme-Value (EV) distribution', not to be confused with the GEV model.

EVT provides a framework to study the behaviour of the tails of a distribution by using extreme observations to measure the density of the tail. This measure can then be extrapolated to parts of the distribution still to be observed in the empirical data. It can also be mapped onto distributions with specific tail behaviour. In this way a theoretical process can be simulated that captures the extreme features of the empirical data and improves the accuracy of estimated probabilities of extraordinary events. The nature of extreme events is such that none or very little directly relevant data exist and therefore, the probability of an extreme event is evaluated by assembling disparate sources of relevant evidence, none of which are reliable in isolation.

Though a statistical tool, EVT is currently the most robust approach for predicting the size of a rare event. Whilst there is always some doubt when extrapolating into unknown regions, EVT makes the best use of available data on extreme phenomena. These features have ensured the success of EVT for many applications including in hydrology for flood frequency analysis, finance, insurance, meteorology for wind strength and rainfall assessment, environmental analysis, earthquake risk assessment and many engineering areas including corrosion and fatigue prediction (Embrechts et al. 1997).

The main advantage of the GEV model is that by using the three parameters shape, location and scale, many distributions can be described, including the Gaussian model. The additional shape parameter ξ also known as tail index, allows for the characterization of various tail decay characteristics. This is important in safety-of-life applications such as aviation, where the focus is on the mitigation of the effects of rare events that can have catastrophic consequences. Fig. 2 illustrates the versatility of the GEV. It can be seen that by varying the shape parameter ξ different tails can be characterized. The green line represents a normal distribution with zero mean and standard deviation equal to 1, the common navigation assumption.

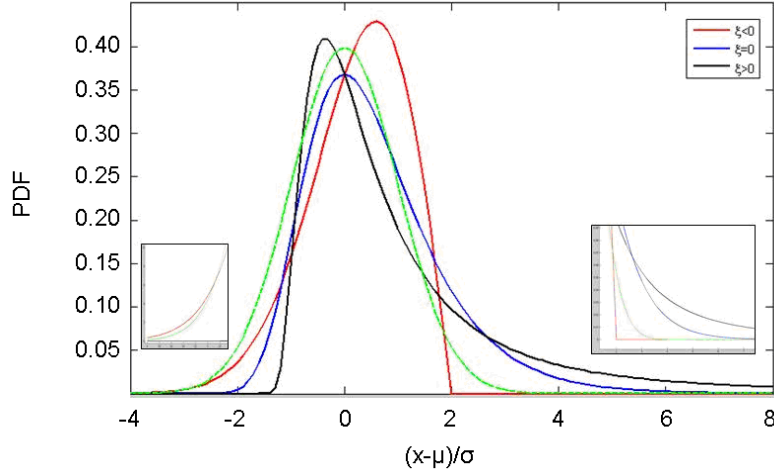


Fig. 2: GEV Versatility with respect to ξ

The potential of EVT in aviation and GNSS applications lies in:

- *Accidents and incidents modeling*: due to the lack of data from underreporting and the rare occurrence of these events, traditional statistical methods cannot adequately characterize them. Panagiotakopoulos et al. (2009) proposed the use of EVT to model accidents and incidents in aviation.
- *Navigation atmospheric errors*: EVT has been used extensively in meteorology and modeling of extreme environmental conditions. Hence, the extreme behaviour of the troposphere and ionosphere could also be studied in the context of EVT. In this regard, Collins and Langley (1998) provided some insight into the modeling of the troposphere using EVT.
- *Residual navigation estimation error*: the potential of applying EVT to characterize residual navigation errors has been recognized in the context of SBAS by both Ober (2003) and Azais et al. (2009).

To date there has been no research in the application of EVT to characterize residual navigation errors for RAIM. Therefore, this paper is unique in developing a RAIM algorithm based on EVT.

4.1 GEV-RAIM

The application of the GEV into a RAIM method employs the same test statistic used in traditional RAIM, i.e. the SSE, but with a different threshold computed on the basis of the P_{MD} and degrees of freedom. Since the GEV includes an additional shape parameter the threshold, T_{GEV} , is determined as follows.

$$T_{GEV} = \left(inv \left[GEV_{1/(n-4), \sigma, (xAL \cdot f \cdot / (SLOPE_{xMAX}))} \left(P_{MD} \right) \right] \right) / (n-4) \quad (12)$$

It is noticeable that the $Pbias_{max}$ parameter, i.e. $Pbias_{max} = xAL \cdot f \cdot / (SLOPE_{xMAX})$, is neither initially standardized by the standard deviation, σ , nor squared as when using the noncentral chi-square distribution. Instead σ is directly used as standard deviation for the GEV and the $Pbias_{max}$ as the mean. Applying this threshold forms the basis of the optimized RAIM algorithm, herein called GEV-RAIM.

5. STATISTICAL ANALYSIS

Analysis of the applicability of the GEV distribution to air-navigation was carried out using the methodology captured in Fig. 3. The Gaussian and Laplace distributions were also considered for modeling the data sets chosen as they are currently the models recommended to characterize the residual navigation errors. Table 1 presents the nine stations considered in this analysis, three in each latitude region to capture variations in ionospheric activity. To capture tropospheric effects, the stations were chosen at different heights and longitudes. The dates chosen for the analysis are the Day of Year (DOY), 89, 90, 96, 97, 196, 197, 303 and 305 for each of the years 2001, 2003, 2005 and 2006. These dates were chosen to study the impact of different atmospheric conditions, as the years 2001 and 2003 experienced high solar activity and ionospheric storm events, respectively. The residual data was generated at a sampling rate of three minutes to ensure an appreciable level of temporal de-correlation of measurements (ICAO, 2005). This resulted in a total of 15360 epochs processed for each station.

Table 1: Stations Chosen for Analysis

Region	IGS ID	City	Country	Longitude (°E)	Latitude (°N)	Height (m)
Equatorial	KOUR	Kourou	French Guyana	307.1940	5.2522	-25.5700
	MAS1	Maspalomas	Spain	344.3667	27.7637	197.3000
	MKEA	Mauna Kea	USA	204.5437	19.8014	3755.000
Mid-latitudes	PTBB	Braunschweig	Germany	10.4597	52.2962	130.2000
	NNOR	New Norcia	Australia	116.1927	-31.0487	234.9840
	OUS2	Dunedin	New Zealand	170.5109	-45.8695	26.1000
High-latitudes	FAIR	Fairbanks	USA	212.5008	64.9780	319.0002
	KIRU	Kiruna	Sweden	20.9684	67.8573	319.1000
	HOFN	Hoefn	Iceland	344.8132	64.2673	82.5000

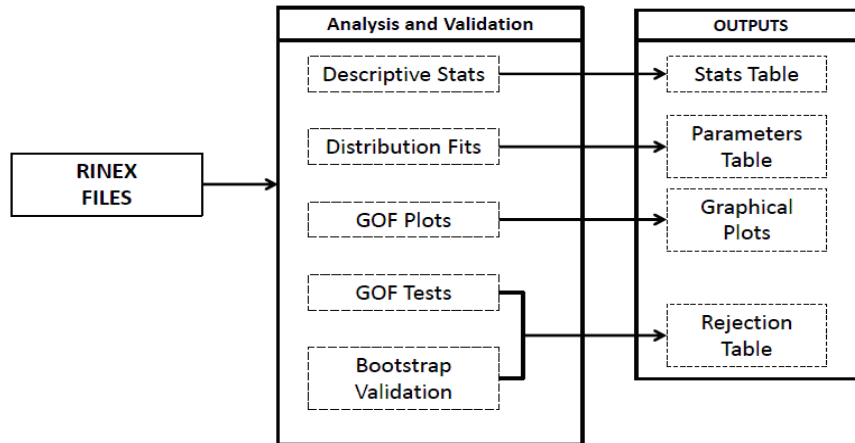


Fig. 3. Methodology for statistical analysis

The main steps in the analysis process were:

- Residual Generation:** the first step of the analysis was to generate the residuals from the data. This was done by first computing the position solution using the Least-Squares method given the pseudo-ranges and then comparing this to the known position for each of the stations.
- Descriptive Statistics:** Eight statistics were computed to summarize the main features of the residuals generated; maximum, minimum, mean and median to capture the central tendency of the data; Standard Deviation (SD), variance (VAR) to capture the dispersion with respect to the mean; skewness and kurtosis to capture the symmetry and the heaviness of the tail respectively. Kurtosis is an important statistic as it is a measure of the outlier-prone nature of a distribution and indicates the potential of applying the GEV distribution to characterize residuals with heavier than Gaussian tails. The higher the kurtosis, the more of the variance that is due to infrequent extreme deviations, as opposed to frequent medium-size deviations. Excess kurtosis which equals simple kurtosis minus 3, provides an easier comparison between distributions according to their negative, nil or positive excess kurtosis. A low kurtosis distribution has a sharper peak and thinner tails, while a high kurtosis distribution has a more rounded peak and heavier tails.
- Distribution fitting:** each of the distributions considered (Normal, Laplace, GEV) were fitted to the residual data by estimating their respective parameters. Maximum-likelihood estimation was used in all three cases.
- Graphical plots:** each of the fits from the previous step were first assessed graphically using a histogram plot of the residual data overlaid by the three fits, quantile-quantile (qq) plots which plot the sample data order statistics

against the order statistics predicted by the fits and probability plots using the Empirical Distribution Function (EDF) of the data plotted against the CDF of each of the fits.

- e) **Goodness-of-fit (GoF) Tests:** these measures the compatibility of the residual sample with the distribution fits in a quantitative way. The GoF tests used were the Kolmogorov-Smirnov, Kuiper's, Anderson and Darling, the Cramer-von-Mises, and Watson (Stephens 1992).

The tests described in e) above are generally used to verify the assumption that a certain distribution fits the data being considered with a certain significance level, α , which in turn is derived from the distribution itself. This means that each distribution will have different critical values for a certain level. A comprehensive account of these tests and their critical values is given in D'Agostino and Stephens (1986) for a number of parametric distributions, including the normal and extreme value. However, the critical values computed for each test highly depend on the parameters estimated for each distribution. In the case that $F(x)$ is specified completely with known parameters, asymptotic distributions of goodness-of-fit statistics are known and critical points can be tabulated. However, in the general case when parameters are estimated from data, distributions of statistics have to be approximated. A relatively straightforward method to determine the critical points is to approximate sampling distributions of goodness-of-fit statistics by the bootstrap method. This approach selects subsets of the data considered and calculates the critical values of the test statistic corresponding to each distribution fit. The asymptotic validity of the bootstrap method for the above-mentioned goodness-of-fit tests was established in Beran and Ducharme (1991).

6. STATICAL ANALYSIS RESULTS

6.1. Goodness-of-fit for Gaussian and Laplace distributions

The statistical summaries are given in Tables 2 and 3 for all periods and for each year respectively. The statistics consists of the percentage of the residual data which skewness exceeds 0.5 (i.e. $>SK0.5$) indicating asymmetry and which have five different cases of Excess Kurtosis (EK), namely $EK0 \leq 0$, $0 < EK1 \leq 1$, $2 < EK3 \leq 2$, $2 < EK3 \leq 3$ and $EK1 \geq 3$ to capture tail behaviour.

Table 2. Residual Data Set Statistical Summary Across all periods (%)

Stations	$>SK0.5$	EK0	EK1	EK2	EK3	EK4
MAS1	98.96	0.00	0.00	0.52	2.08	97.40
KOUR	86.63	6.42	9.03	6.60	2.43	75.52
MKEA	81.77	9.37	11.83	10.04	3.72	65.03
PTBB	86.46	2.78	14.24	7.99	3.13	71.88
NNOR	95.83	0.52	0.52	1.56	1.56	95.83
OUS2	89.78	5.06	8.04	4.91	2.08	79.91
FAIR	85.94	0.63	3.75	0.63	3.13	91.88
KIRU	88.62	2.75	6.47	10.12	6.32	74.33
HOFN	88.89	8.51	16.67	24.13	22.92	27.78
<i>Average</i>	89.21	4.00	7.84	7.39	5.26	75.51

Table 3: Residual Data Set Statistical Summary per year for the equatorial stations (%)

Residual	2001		2003		2005		2006	
	$>SK0.5$	EK4	$>SK0.5$	EK4	$>SK0.5$	EK4	$>SK0.5$	EK4
\underline{v}_{vLS}	100	100	100	100	100	100	87.5	75
\underline{v}_{vLS}	100	100	100	100	100	100	100	100
\underline{v}_{zLS}	100	100	100	87.5	100	100	100	100
\underline{v}_{vWLS}	100	100	100	87.5	100	100	100	100
\underline{v}_{vWLS}	100	100	100	100	100	100	87.5	100
\underline{v}_{zWLS}	100	100	100	100	100	100	100	100

On average, across all periods and stations, approximately 90% of the residual data are asymmetric, while 96% have heavy tails, with 88% with very heavy tails (i.e. with $EK > 2$). Table 3 presents these non-Gaussian features, which are particularly acute for the equatorial stations and for the periods of high ionospheric activity, i.e. year 2001. For the periods of low ionospheric activity the residuals are better behaved with lighter tails except for the period of around October 2003.

The following three cases of EK are used to analyses the goodness-of-fit of the Gaussian and Laplace distributions using the methodology outlined in section 5: EK0: station KOUR for Day 89 of year 2005; EK3: station PTBB for Day 197 of year 2001; EK4: station PTBB for Day 305 of year 2003.

1) Descriptive statistics and distribution parameters estimation tables

The descriptive statistics for the horizontal residual error of each of the EK cases are given in Table 4. It should be noted that although the standard deviation increases with increasing EK, the relationship is not linear between the two

statistical parameters. For this reason, although the weighted least-squares position estimation gives a more accurate result than ordinary least-squares, it does not necessarily decrease the heaviness of the tails.

Table 4: Residuals Descriptive Statistics (units m)

Residuals Descriptive Statistics								
EK Cases	Maximum	Minimum	Mean	Median	SD	VAR	Skew	EK
EK0	2.921	0.123	0.947	0.794	0.732	0.563	0.387	-0.809
EK3	193.3	569.3	61.23	-26	144.9	20996	-1.436	2.143
EK4	512.2	544.7	3.059	9.198	130	16905	-1.343	5.858

2) Graphical Assessment

Fig. 4 shows the graphical assessment of the distribution fits using the parameters estimated in the previous step. The Gaussian distribution being symmetric and with ‘normal’ tails has a skewness and an EK close to 0. Thus, it should be expected that data samples with non-zero skewness and high EK are not effectively described by a Gaussian. In general, the higher the EK of a distribution, the longer its tails are and hence the less likely it is to be described by a Gaussian. Fig. 4 show that, as EK increases the performances of the fits degrade in characterizing the residual distributions.

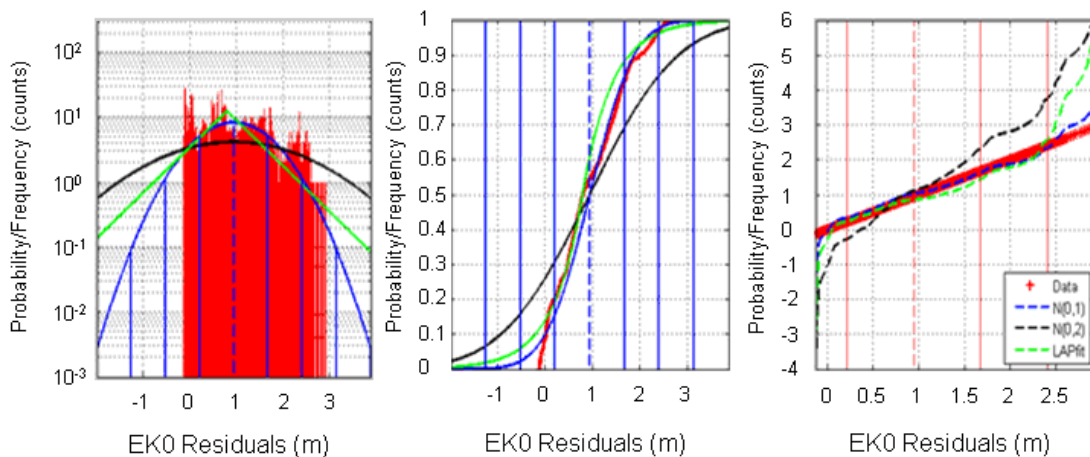
With respect to the overbounding technique, the PDF and CDF plots indicate that the Gaussian model with unity Standard Deviation (SD) and Laplace can perform this function for the Case EK0 but not for the higher EK cases. The Gaussian with twice the SD can provide an overbounding distribution for all cases, although from the quantile-quantile plots it can be concluded that as a model, it is not such a good fit as it does not efficiently follow the shape of the residual data.

3) Rejection Tables from Bootstrapping

The final step in assessing the goodness-of-fit is by applying the bootstrap algorithm and test statistics. The number of bootstrap samples was set to 10000. The percentages of the number of times each of the fits were rejected across all test statistics is given in Table 5. This table clearly complements the conclusions made from the graphical assessment that the performance of the distributions in describing the considered residual data sets is poorer for increasing EK case. It is clear that the higher the EK of the data samples, the more the distributions fail to efficiently describe them. Although the distributions with heavier tails, i.e. Gaussian with 2σ and Laplace are better fits for the EK3 and EK4 data sets than the Gaussian with zero mean and unity SD, they are respectively, still rejected 70% and 90% of the time.

Table 5: Rejection Tables for Residuals: Summary across all statistics (%)

	Gaussian	Gaussian (2σ)	Laplace
EK0	25.30	44.19	38.85
EK3	84.08	70.02	70.10
EK4	98.25	96.38	93.99



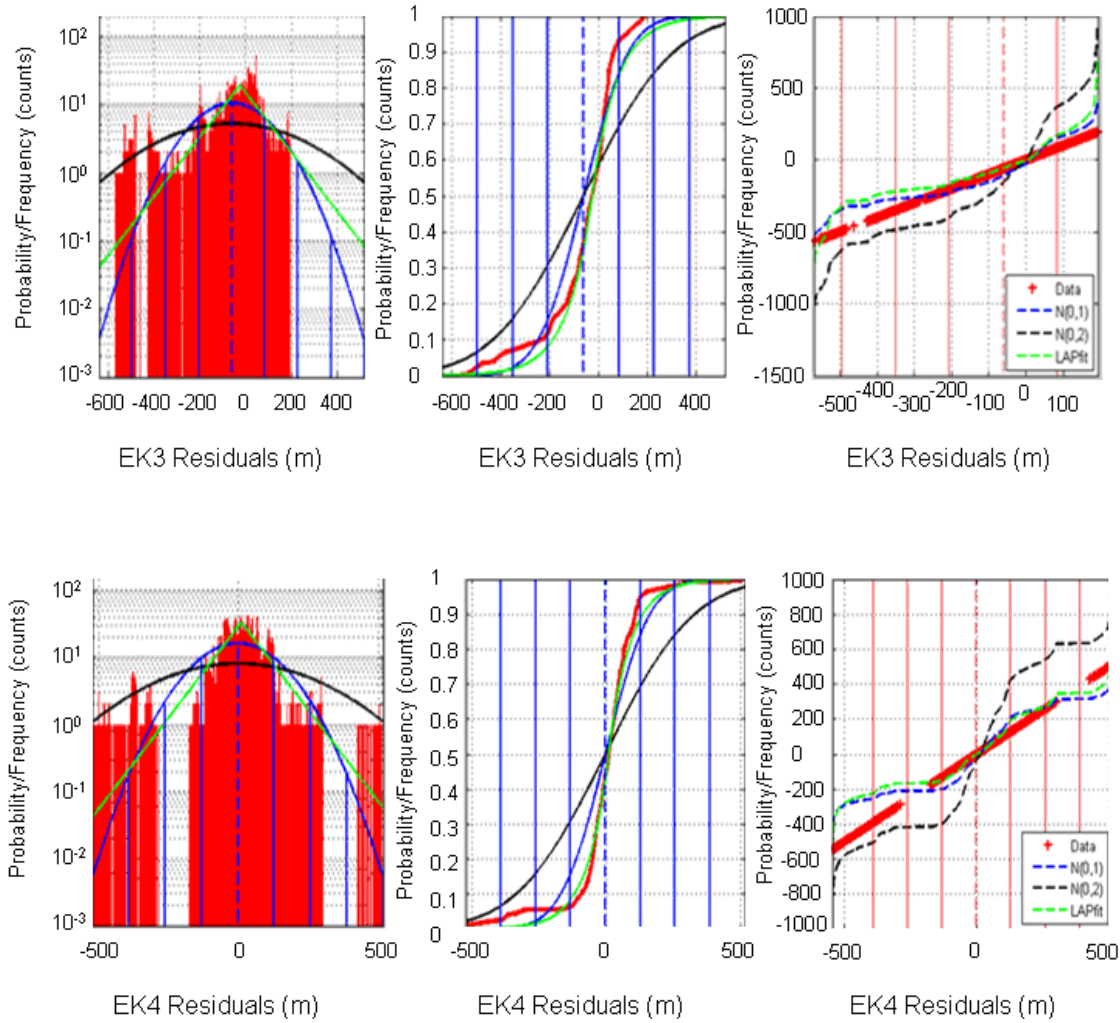


Fig. 4: Goodness-of-fit for the Gaussian and Laplace distributions for the residuals for cases EK0, EK3 and EK4

6.2. Impact on RAIM

The RAIM algorithms described in section 2 were applied to the entire data sets of 9 stations and also to the three EK cases individually to study their performance. The parameters for Non Precision Approach were used, i.e. $P_{FA} = 3.33 \times 10^{-7}$, $P_{MD} = 0.001$, with a measurement standard deviation of 12.5m, a mask angle of 7.7° , and $f=0.95$. Table 6 shows the results for the equatorial stations for each period for the horizontal component. Table 7 shows the results for the individual EK cases. The tables show the proportions of time when RAIM was not available (NA) and when missed detections (MD) and false alerts (FA) were experienced.

Table 6: RAIM Performance for equatorial stations: Horizontal Component (%)

	MAS1						KOUR						MKEA					
	LS-RAIM			WLS-RAIM			LS-RAIM			WLS-RAIM			LS-RAIM			WLS-RAIM		
	NA	MD	FA	NA	MD	FA	NA	MD	FA	NA	MD	FA	NA	MD	FA	NA	MD	FA
2001	31.72	9.52	6.39	21.50	8.12	7.23	18.71	11.23	0.00	12.54	9.65	4.56	31.46	13.26	19.25	12.55	10.63	3.78
2003	41.40	12.46	10.82	24.60	8.32	8.55	34.76	9.83	4.23	13.64	8.51	7.52	23.73	11.47	13.14	13.44	9.54	5.36
2005	27.44	11.63	10.61	23.54	9.16	10.60	47.42	8.24	8.33	15.42	7.23	8.24	31.99	12.53	5.48	13.66	10.49	4.57
2006	35.86	10.78	13.35	22.10	9.49	11.24	27.97	9.46	0.00	25.61	7.47	6.54	32.36	10.64	4.23	13.57	9.48	4.55

Table 7: RAIM Performance for EK Scenarios: Horizontal Component (%)

	LS-RAIM			WLS-RAIM		
	NA	MD	FA	NA	MD	FA
	EK0	0.69	1.56	2.50	0.30	0.98

EK3	15.24	11.84	13.20	5.60	8.46	7.44
EK4	24.50	19.36	22.30	11.20	12.48	8.40

Tables 6 and 7 show the impact of the statistical characteristics of the residuals on the performance of RAIM, in particular for the equatorial stations. LS-RAIM is unavailable on average 32% of the time with 8% of the epochs processed resulting in a false alert and 11% in a missed detection, while for the WLS-RAIM the values are 17.7%, 6.9% and 9% respectively. Furthermore from Table 7, it can be seen that the unavailability, false alerts and missed detections of both the LS-RAIM and WLS-RAIM processes increase with increasing EK.

6.3 Goodness-of-fit of GEV distribution

1) Distribution parameters estimation

The parameters for the GEV distribution are estimated using the Maximum Likelihood Estimation (MLE) technique. The Log-Likelihood (LL) function for the GEV distribution is given by:

$$LL_{GEV}(\xi, \mu, \sigma) = \begin{cases} -n \ln \sigma - (1+1/\xi) \sum_{i=1}^n \ln[1+\xi(\frac{x_i-\mu}{\sigma})] - \sum_{i=1}^n [1+\xi(\frac{x_i-\mu}{\sigma})]^{-1/\xi} & \xi \neq 0 \\ -n \ln \sigma - \sum_{i=1}^n (\frac{x_i-\mu}{\sigma}) - \sum_{i=1}^n \exp[-(\frac{x_i-\mu}{\sigma})] & \xi = 0 \end{cases} \quad (13)$$

Differentiating these functions with respect to μ , σ and ξ , and equating them to 0, results in the simultaneous equations to be solved to estimate the parameters. Clearly, no explicit solution exists to these non-linear equations and thus numerical procedures or search algorithms are used such as the Newton-Raphson scheme (Prescott and Walden 1980). The resulting MLEs for each parameter can be found in Stephens (1977) which give the equations to be solved.

The parameters estimated for the GEV distribution for the three EK cases are shown in Table 8. These are the parameters associated with the horizontal residual errors.

Table 8: Residuals GEV Distribution Parameter Estimates

EK Cases	GEV Distribution Estimates			Confidence Intervals (95%)		
	ξ	Σ	M	ξ	Σ	M
EK0	-0.1003	0.6347	0.63	-0.1141; -0.0887	0.6304; 0.6500	0.6167; 0.6546
EK3	-0.5348	155.3354	-94.098	-0.5424; -0.5232	152.0967; 157.0129	-96.6247; -92.1897
EK4	-0.2619	152.999	-52.4472	-0.2719; -0.2501	150.2080; 154.7720	-54.9595; -48.4397

The precision of the estimated parameters in Table 8 was verified by changing slightly the input variables and ensuring that the change does not impact the stability of the estimation process, i.e. that the parameters estimated remain largely the same. This uncertainty sensitivity analysis confirmed that each of the small changes in the input data resulted in estimated parameters that remain within their respective 95% confidence intervals as shown in Table 8. Additionally, the 95% confidence intervals for each of the estimated parameters are stable and without significant variation, which adds to the evidence that for the datasets considered the MLE is an accurate and stable estimator.

2) Graphical Assessment

The Gaussian distribution has a skewness and an EK of near 0 and hence, it fails to characterize efficiently data samples with high EK. It is thus expected that distributions with high EK (i.e. with heavy tails) would be better characterized by the GEV, because it is a distribution specifically designed for this purpose. Fig. 5 shows the performance of the GEV fit in characterizing the three EK cases. In all cases, the GEV closely follows the shape of the data considered and models it in a more accurate way than both the Gaussian and Laplace. Furthermore, the near-linear pattern of the q-q plot supports this conclusion. For the EK4 case, although the q-q plot is not as linear as in the other cases, it is closer to the actual data than the Gaussian and Laplace.

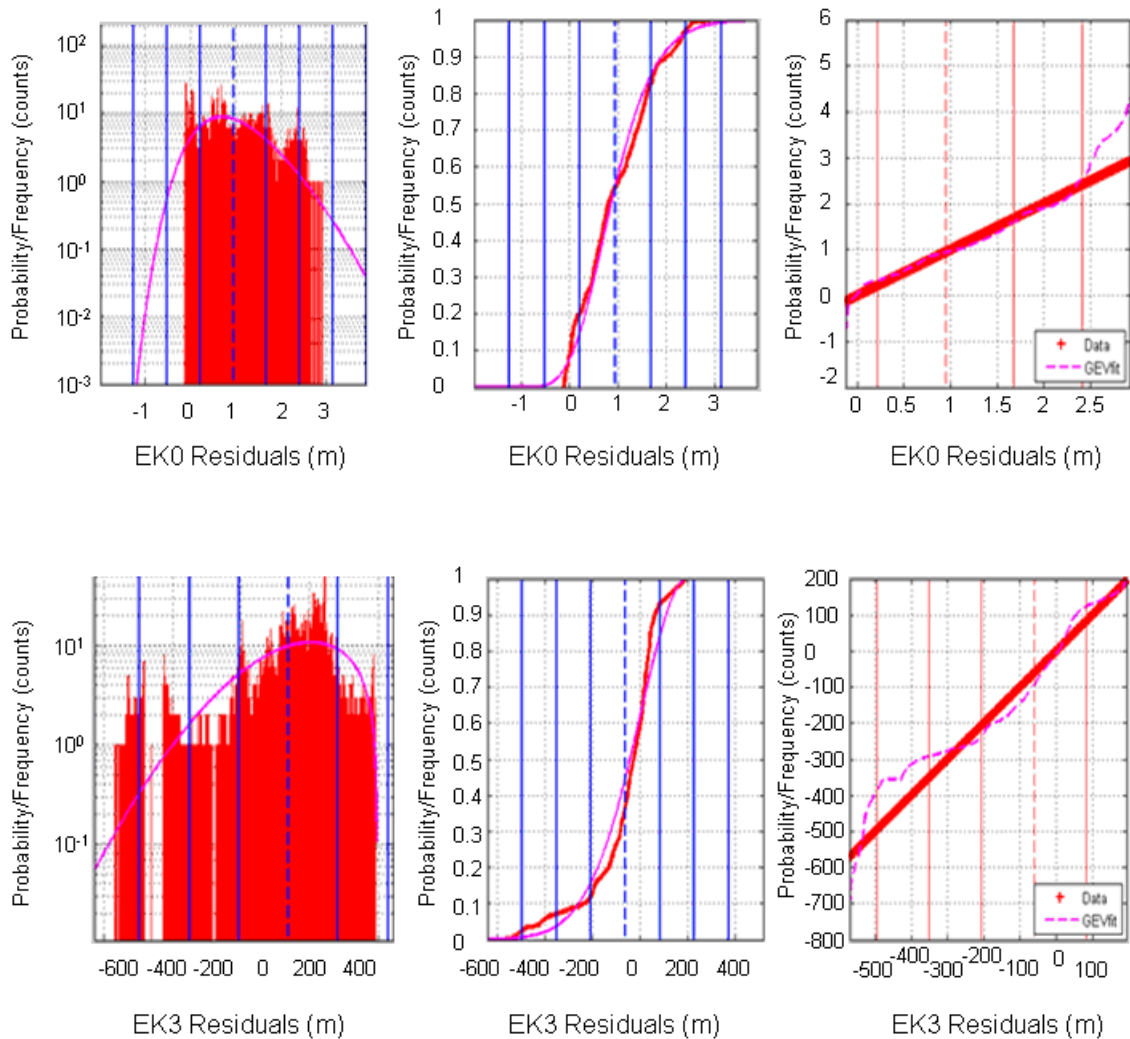
3) Rejection Tables from Bootstrapping

The results from the bootstrap algorithm are shown in Table 9, which presents the average number of times the GEV fit was rejected across all statistics in percentage. Comparing with the rejection tables of the traditional residual

distributions in Table 5, it can be seen that the GEV is a better fit by an order of magnitude, to the data sets considered. The GEV is rejected 3.2%, 3.8% and 4.7% of the time for the EK0, EK3 and EK4 cases respectively, against on average 40%, 70% and 90% for the traditional models. The findings complement the previous graphical assessment in that the GEV is a better fit than the Gaussian and Laplace distributions in characterizing the data sets considered. From Table 9 and Fig. 5, it is evident that GEV is a better fit than both the Gaussian and Laplace for these residual data sets.

Table 9: GEV Rejection Tables for residual data sets: Percentage summary across all statistics

EK Cases	GEV
EK0	3.19
EK3	3.83
EK4	4.72



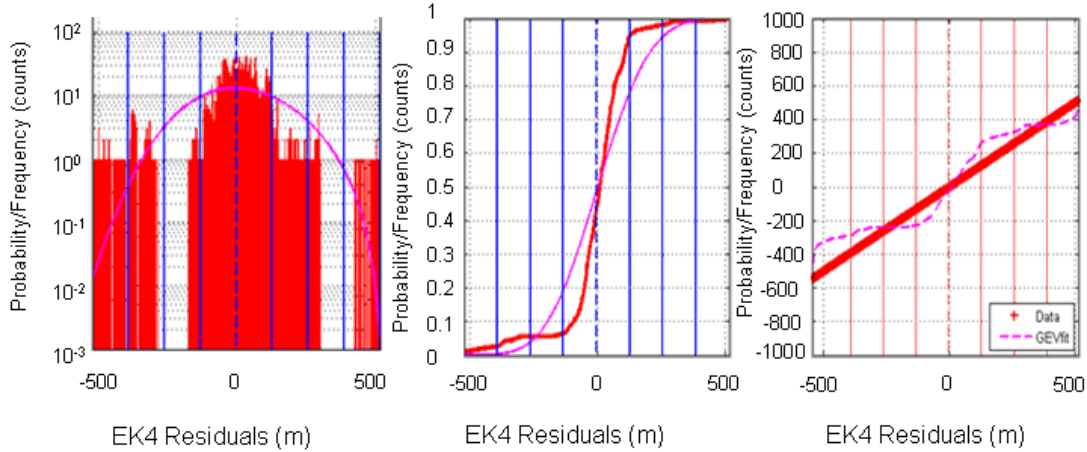


Fig. 5: Graphical Assessment of the goodness-of-fit of the GEV for the residual data and Cases EK0, EK3 and EK4

This section clearly demonstrates the potential of using the GEV distribution to model the residual navigation estimation errors that result from the typical least-squares positioning algorithm widely used in GNSS receivers. The findings imply that the GEV is a better fit to describe residuals than the Gaussian model. In the cases where the residuals are heavy-tailed and the Gaussian assumption fails to provide a good model, the GEV functional model provides a very good alternative and thus should be considered in the design of positioning and integrity algorithms.

6.4 Performance of GEV-RAIM

The data sets for the particular cases of EK were also used to study the performance of the GEV-RAIM algorithm. The results are shown in Table 10 for horizontal positioning. It can be seen that GEV-RAIM performs very well with no false alerts and, an order of magnitude less in RAIM unavailability, i.e. 0.08% on average across the EK cases and missed detections, 2% on average, compared to traditional RAIM algorithms. The corresponding figures for LS-RAIM are 13%, 11% and 13.5% respectively; and, 6%, 6% and 7% respectively for WLS-RAIM.

Table 10: GEV-RAIM Performance for Data and EK Cases (%)

EK Cases	Horizontal Component		
	GEV-RAIM		
	NA	MD	FA
EK0	0.00	0.34	0.00
EK3	0.13	2.51	0.00
EK4	0.12	2.86	0.00

Furthermore, the GEV-RAIM is stable as the tail heaviness of the residual data set used for analysis increases compared to LS-RAIM and WLS-RAIM, whose performances degrade substantially from EK3 to EK4. This is expected, as GEV-RAIM incorporates an additional parameter compared to LS-RAIM and WLS-RAIM to characterize the tail heaviness of the residual data distribution, i.e. the tail index ξ . The tail index enables GEV-RAIM model to be more versatile and a better RAIM algorithm, particularly when applied to heavy tailed residual errors. Therefore, the use of the GEV distribution in the derivation of the optimized threshold in GEV-RAIM enables the threshold to be less pessimistic than the one derived in the traditional RAIM algorithms, which in turn results in a tighter PL and hence allows the GEV-RAIM to have a better performance in terms of unavailability, false alerts and missed detections.

8. CONCLUSIONS

The paper has characterized the tails of residual error distributions in navigation and challenged the assumption of normally distributed errors which is the ‘de facto’ model currently implemented in most navigation algorithms, such as in differential GPS, RAIM and SBAS. It has been shown that a good alternative to the Gaussian model, in these extreme though rare cases is the Generalized Extreme Value (GEV) distribution, which incorporates a shape parameter to represent different rates of tail decay. Although further tests are required, both the graphical assessment and goodness-of-fit tests show that the GEV is more powerful in characterizing the tails than the Gaussian model. In terms of integrity monitoring, it has been shown that GEV-RAIM provides a safety threshold that is less conservative than that derived from the official missed detection probability, while still within the bounds imposed by the false alert probability thereby not affecting availability. This offers a good trade-off between continuity and integrity risk for air-

navigation. Furthermore, the better statistical model should result, overall, in better position estimation quality. Therefore, it is proposed that the GEV model is used for the development of future integrity monitoring algorithms.

ACKNOWLEDGEMENTS

The authors thank the UK's Engineering and Physical Sciences Research Council and Logica for funding this research.

REFERENCES

- Azais J-M, Gadat S, Lévy J-C, Rols B, Mercadier C, Jordan C, Suard N (2009) GNSS Integrity Achievement by using Extreme Value Theory. ION-GNSS-2009, Savannah, GA, 1281-1287.
- Anderson EW, Ellis DM (1971) Error distributions in navigation. *Journal of Navigation*, 24, 429-442.
- Beran R, Ducharme G (1991) *Asymptotic Theory for Bootstrap Methods in Statistics*. Centre de recherches mathématiques, Université de Montréal.
- Bhatti U (2008) Improved integrity algorithms for the integrated GPS/INS systems in the presence of slowly growing errors. PhD Thesis, Imperial College London
- Brooker P (2004) Radar Inaccuracies and Mid-Air Collision Risk: Parts 1 and 2 - A Dynamic Methodology. *Journal of Navigation*, (57)1, 25-37; 39-51.
- Brown RG (1992) A baseline GPS RAIM scheme and a note on the equivalence of three RAIM methods. *Navigation*, (39)3, 301-316.
- Brown RG (1996) Receiver Autonomous Integrity Monitoring. *Global Positioning System: Theory and Applications*. Volume II. Published by American Institute of Aeronautics and Astronautics, 143-165.
- Collins JP, Langley R (1998) The residual tropospheric propagation deal: How bad can it get?, ION-GPS, 15-18.
- D'Agostino RB, Stephens MA (1986) *Goodness-of-Fit Techniques*, Marcel Dekker Ltd.
- DeCleene B (2000) Defining Pseudo Range Integrity – Over-bounding. ION-GPS.
- Embrechts P, Kluppelberg C, Mikosch T (1997) *Modelling Extremal Events for Insurance and Finance*, Springer.
- Feng S, Ochieng WY, Walsh D, Ioannides R (2006) A measurement domain receiver autonomous integrity monitoring algorithm. *GPS Solutions Journal*, 10(2), 85-96.
- Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest number of the sample. *Proc. Cambridge Philos. Soc.* (24).
- Hsu DA (1979) Long-Tailed Distributions for Position Errors in Navigation. *Applied Statistics*. 28(1), 62-72.
- ICAO (2005) *Standards and Recommended Practices Specifications*, ICAO Annex 10, Vol. I; "Radionavigation Aids.
- Moek G, Lutz E, Mosberg W (2001) Risk Assessment of RNP10 RVSM in the South Atlantic Flight Identification Regions. ARINC Corporation.
- Ober PB (2003) *Integrity Prediction and Monitoring of Navigation Systems*", Integricom Publishers.
- Ochieng W Y, Sheridan KF, Sauer K, Han X (2002) An assessment of the RAIM performance of a combined Galileo/GPS navigation System using the marginally detectable errors (MDE) algorithm. *GPS Solutions*, (5)3, 42-51.
- Panagiotakopoulos D (2009) Robust statistical framework for monitoring the integrity of space-based navigation systems, and preparing the marketplace for integrity-based services. PhD Thesis, Imperial College London.
- Panagiotakopoulos D, Majumdar A, Ochieng WY (2009) Characterising the distribution of safety occurrences in aviation: an approach using extreme value theory. *Transportation Research Records*, 2106, 129-140
- Parker JB (1972) Error distributions in navigation. *Journal of Navigation*, 25, 250-252.
- Prescott P, Walden AT (1980) Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution. *Biometrika*, 67, 723-724.
- Rife J, Walter T, Blanch J (2004) Overbounding SBAS and GBAS Error Distributions with Excess-Mass Functions. The 2004 International Symposium on GNSS/GPS, 6-8 December 2004, Sydney, Australia.
- Shively C (2000) An Overbound Concept for Pseudorange Error from the LAAS Ground Facility. ION-GPS 2000.
- Stephens MA (1992) *Breakthrough in Statistics*, Springer Verlag, New York & USA.
- Stephens MA (1977) Goodness-of-Fit for the Extreme Value Distribution. *Biometrika*, (64), 583-588.
- Walter T, Enge P, DeCleene B (2002) Integrity Lessons from the WAAS Integrity Panel (WIPP), ION-GPS 2002.