EGNOS vertical protection level assessment

David Comby, Rick Farnworth, Christophe Macabiau

EGNOS Vertical Protection Level Assessment

David COMBY, *French Air Force*

Rick FARNWORTH, *Eurocontrol Experimental Centre*

Christophe MACABIAU, *ENAC*

BIOGRAPHY

David Comby is an officer of the French Air Force. He graduated as an aeronautical engineer in 1995 from the Air Force academy (Ecole de l'air) in Salon de Provence. In 2002, he obtained his Master degree in satellite-based Communication Navigation Surveillance from ENAC (Ecole Nationale de l'Aviation Civile) in Toulouse, France. He graduated as a Civil Aviation Engineer the same year after the work he performed in the satellite navigation domain within the Eurocontrol Experimental Centre.

Rick Farnworth graduated with a BSc in electronic Engineering from the University of Wales in 1988 and was awarded a PhD in 1992 for his work on LORAN-C coverage prediction modeling. He then joined the UK's CAA's National Air Traffic services to work on R&D projects relating to the application of satellite navigation systems in civil aviation. Since February 1996 he has been working for Eurocontrol at their experimental centre in the GNSS Project Office where is responsible for various R&D projects related to satellite navigation.

Christophe Macabiau graduated as an electronics engineer in 1992 from ENAC (Ecole Nationale de l'Aviation Civile) in Toulouse, France. Since 1994, he has been working on the application of satellite navigation techniques to civil aviation. He received his Ph.D. in 1997 and has been in charge of the signal processing lab of the ENAC since 2000.

ABSTRACT

The EGNOS Signal In Space (SIS) performance is defined in terms of accuracy, integrity, continuity of service and availability. For Civil Aviation, those four components of the performance shall fulfill ICAO requirements (particularly stringent for the integrity).

EGNOS is expected to be operational by the first quarter of 2004. One of the major issues for Civil aviation in the perspective of operating EGNOS, is to prove that the system is safe to use. In this context, the demonstration of the EGNOS compliance with the integrity requirements is of the utmost importance. This raises the following question:

How to assess the integrity performance of the EGNOS SIS?

In the frame of Eurocontrol work supporting the operational validation of EGNOS, a number of techniques are under investigation to evaluate measurement data and provide an assessment of the integrity achieved.

It is not expected, even using a combination of different techniques that compliance with integrity requirement will be exhaustively demonstrated as this would require an amount of data that is impracticable to collect. The main contributor to the integrity validation is the analysis performed during the system design. However, it is necessary to use techniques such as the one presented in this paper to examine the behavior of the protection level in relation to the position error in order to gain a better understanding of how the system is performing. This technique will also help to identify when the system is not performing as it should be, even during times when it appears to be functioning correctly.

The aim of the work presented in this paper is to develop a methodology for the assessment of the EGNOS Vertical Protection Level (VPL). The presented analysis is based on the processing of data from the EGNOS System Test Bed that is received by a network of data collection stations distributed throughout Europe.

The objective is to make a statistical analysis of the Vertical Position Error (VPE) that can be measured at the output of the data collection receivers. An estimated VPL can be computed from the position error data enabling an evaluation to be made as to whether the VPL provided by the EGNOS system is a conservative bound on the VPE or not. For a given processed data set, a confidence level in the VPL is defined as an estimated probability that the VPL is a bound of the VPE.

The analysis aims to provide the confidence levels as indicators of the quality of the VPL over selected data sets.

INTRODUCTION

Several statistical analyses dealing with the integrity of GNSS have been carried out in the past. They often addressed the issue of the VPL definition from the observations performed in the range domain.

This paper proposes a VPL analysis in the position domain and aims to address the following question:

Is the EGNOS provided VPL overbounding the position error with a sufficient level of confidence?

The basic goal of this method is to appraise the VPL quality thanks to a statistical analysis of the Vertical Position Error (VPE). It is a kind of "plug and play" tool to be applied at the output of receivers[1] from the ESTB DCN[2]. The analysis requires a few assumptions among them the independence of the processed samples. It is based on the following methodology:

- The Eurocontrol software PEGASUS processes data from the EGNOS System Test Bed over a minimum period of 6 days on a given site. The output of this is a data file containing VPE and ESTB VPL,
- The data is split into small intervals around targeted VPL (few centimeters),
- For each targeted VPL, VPE samples that are at least 360 seconds apart are selected (to ensure independence of the samples). Data subsets are then totally determined and ready to be processed,
- From each selected subset, a Gaussian modeling that bounds the VPE distribution (in the cdf overbounding sense) is defined,
- From this previous modeling of the VPE, a statistical analysis is then performed over each selected subset in order to compute an estimated probability that the VPL is a 10^{-7} bound of the VPE,
- This probability is denoted "confidence level in the VPL".

The output of the presented analysis is the confidence level for each targeted VPL on the given site.

The paper begins with a short introduction of the VPL concept for EGNOS.

---

[1] In order to assess the EGNOS SIS, the assumption of fault-free receivers has to be sought.

[2] EGNOS System Test Bed data collection network. It is composed of four permanent sites in Barcelona, Lisbon, Toulouse and Delft.

I. VPL AND INTEGRITY CONCEPT FOR EGNOS

I.1 Definition of the VPL

According to the SARPs [1], the SBAS Vertical Protection Level (VPL) is defined as a bound on the Vertical Position Error (VPE) with a probability derived from the integrity requirement. It is thus one of the major component of the SBAS integrity mechanism.

The VPL is computed by the user receiver using a bound of the standard deviation of the range measurements corrected with the data broadcast by the SBAS SIS.

For each satellite contributing to the position solution, the SBAS corrected range measurement errors are assumed to have a Gaussian, independent and centered distributions. The SBAS corrected range measurement errors components are:

- The error on the differential correction on each satellite excluding atmospheric effects and receiver errors ($\sigma_{i,flt}$ should bound the standard deviation of this error for the i^{th} satellite),
- The error on the ionospheric correction on each satellite ($\sigma_{i,UIRE}$ should bound the standard deviation of this error for the i^{th} satellite),
- The aircraft pseudo range errors due to the combination of receiver and aircraft multipath ($\sigma_{i,air}$ should bound the standard deviation of this error for the i^{th} satellite),
- The residual pseudo range of a tropospheric correction model which is defined by a standard mentioned in the SARPs ($\sigma_{i,tropo}$ should bound the standard deviation of this error for the i^{th} satellite).

The overall error distribution in the measurement domain (for the i^{th} satellite) is also assumed to be centered with a standard deviation bounded by:

$$\sigma_i=\sqrt{\sigma_{i,flt}^2+\sigma_{i,UIRE}^2+\sigma_{i,air}^2+\sigma_{i,tropo}^2} \qquad (1)$$

The knowledge of the bounds of the N standard deviations in the measurement domain (related to the N satellites which contributes to the position solution) allows the computation of the bound of the standard deviation for each component of the position domain. For the vertical component it gives:

$$\sigma_{V,integrity}=\sqrt{\sum_{i=1}^{N} S_{V,i}^2\cdot\sigma_i^2} \qquad (2)$$

$S_{V,i}$ are geometrical parameters: they are the partial derivative of the position error in the vertical direction with respect to pseudo range error on the i^{th} satellite. They characterize the relative geometry between the N satellites and the user receiver.

Finally, the VPL is computed by multiplying by a K factor:

$$VPL = K_V \cdot \sigma_{V,integrity} \qquad (3)$$

with $K_V = \text{Normal\ cdf}^{-1}\left(1 - \frac{10^{-7}}{2}\right) = 5.33$

I.2 Link between the VPL and the integrity requirement

The goal of this sub-section is to derive the EGNOS VPL requirement that will be used in the analysis from the high level integrity risk defined by ICAO [1] and Eurocontrol [2] .

First, it could be written that:

Integrity Risk = $P_{md_{VPL}} \cdot P(\text{no range bias})$, $P_{md_{VPL}}$ being equal to the conditional probability P(VPE>VPL/no range bias).

After application of SBAS corrections, the range measurements are assumed to be non affected by any range bias[1], meaning that P(no range bias) is close to 1. Therefore, it could be written:

$$\text{Integrity Risk} = P_{md_{VPL}} = P(VPE>VPL) \qquad (4)$$

According to [1], ICAO defines an acceptable integrity risk for an approach using GNSS:

$$\text{Acceptable integrity risk} = 10^{-7} \text{ per approach} \qquad (5)$$

From (4) and (5) it could be deduced the following EGNOS requirement:

$$P(VPE>VPL) = 10^{-7} \text{ per approach} \qquad (6)$$

According to Eurocontrol [2], one approach is assumed to last 150 seconds. Furthermore, it is commonly accepted that the maximum correlation time for the position error using EGNOS is 360 seconds (mainly due to ionosphere)[2]. Therefore, it could be assumed that there is one independent sample per approach. This leads to the EGNOS requirement that will be the basis of the analysis:

$$P(VPE>VPL) = 10^{-7} \text{ per independent sample} \qquad (7)$$

I.3 VPL issue

The integrity performance of EGNOS is a significant contributor to safety at user level. Therefore, the assessment of the VPL quality is a real issue and should answer the question:

Is the VPL a reliable bound of the VPE related to a probability of 10^{-7} ?

The next sections try to address this issue.

II. METHODOLOGY OF THE VPL ASSESSMENT

II.1 Objective and rational

The objective is to assess whether the values of the VPL for a given data set are reliable bounds of the VPE related to a probability of 10^{-7} .

In this framework, the paper proposes a statistical analysis of the VPE when the VPL is around a targeted value (+/- a tolerance). For each tested value of the VPL, the information being sought is the probability that the targeted VPL is a reliable bound of the VPE related to a probability of 10^{-7}. This probability is denoted "confidence level" in the targeted VPL.

The paper presents a statistical analysis that aims to compute the confidence level in targeted VPL values from a 6 day (at least) data set.

From this data set, the final results are collected in a graph showing the confidence level in the VPL with respect to the tested values of the VPL.

II.2 Selection of the processed data

First level of the data selection:

From a 6 day (at least) data file containing VPE and ESTB VPL, it is first necessary to select the GPS epochs where the ESTB VPL is around a specific targeted VPL, +/- a specified tolerance (nominal value of 0.125meters). The VPE observed at these GPS epochs are then stored.

This selection process begins by selecting the targeted VPL as the lowest value of the VPL in the file and increasing to a maximum value of 50 meters[3].

Second level of the data selection:

Then, in line with the SARPs, the analysis is based on the assumption that 360 seconds are enough to ensure independence of the samples. Therefore, the next step of the data selection is to identify samples that are at least 360 seconds apart among the samples that were selected for each targeted VPL. This process is performed for each targeted VPL and should systematically contain the maximum value of the VPE (in absolute value) and its corresponding VPL.

---

[1] After application of SBAS corrections, the range measurements are assumed to be only affected by noise.

[2] This correlation time is assumed in [3].

[3] According to [2], 50 meters corresponds to the maximum Vertical Alert Limit applicable for Civil Aviation. If the VPL exceeds this value, the system becomes unavailable for the user.

Third level of the data selection:

Finally, in order for the analysis to make statistical sense, the required minimum number of selected independent samples is fixed at 200. If this requirement is not satisfied, the size of the VPL interval around the targeted VPL is grown until at least 200 independent samples are included. A trade-off should be found to maximize the number of processed samples while minimizing the size of the selected VPL intervals.

Output of the data selection:

From an initial data set file containing VPE and ESTB VPL, data are stored in several data subsets over which:

- The VPL is almost constant (if possible within a 0.25 meter interval),
- Samples of each selected subsets are at least 360 seconds apart to ensure independence of the VPE samples; however they systematically include the maximum absolute value of the VPE related to the targeted VPL,
- Each selected subset contains at least 200 samples.

II.3 Description of the statistical analysis

The proposed statistical analysis implemented on selected data subsets is based on several mathematical tools that will be presented step by step in this section.

First step: ideal case

This first step makes two very simple assumptions for each of the selected data subsets:

- The VPE follows a perfect Gaussian distribution,
- The mean value μ and the standard deviation σ of the VPE are perfectly known.

In reality these assumptions are not met. Indeed, on each selected data subsets the VPE might not follow a Gaussian distribution. Furthermore, the mean value and the standard deviation of the VPE, which can be computed are only estimates based on a limited number of observations.

However, this first step is necessary in order to provide the reader with a simple introduction of the process.

Under both simple assumptions mentioned above, from the statistical properties of the VPE, the error related to a probability of 10^{-7} can be simply derived. This error can be denoted VPL_{ideal} insofar as this value should be the ideal VPL for the processed data set. VPL_{ideal} should be the solution of the following equation:

$$1-\frac{1}{\sigma\cdot\sqrt{2\pi}}\int_{-VPL_{ideal}}^{VPL_{ideal}} e^{-\frac{(t-\mu)^2}{2\cdot\sigma^2}}\cdot dt = 10^{-7} \qquad (8)$$

which is exactly the same as:

$$1-\frac{1}{2}\cdot\left(\operatorname{erf}\left(\frac{VPL_{ideal}+\mu}{\sigma\cdot\sqrt{2}}\right)+\operatorname{erf}\left(\frac{VPL_{ideal}-\mu}{\sigma\cdot\sqrt{2}}\right)\right)=10^{-7} \qquad (9)$$

with $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\cdot\int_0^x e^{-t^2}\cdot dt$

The following figure illustrates this statement:

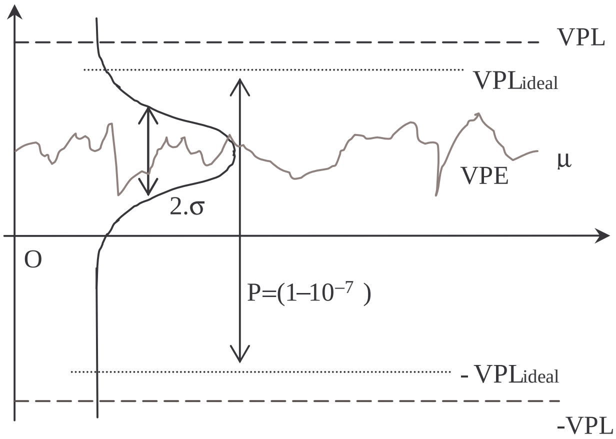

Figure 1: VPE perfectly Gaussian, μ and σ known

The next steps address the fact that those simple assumptions are not met.

Second step: non Gaussian distribution of the VPE

In general, the VPE does not follow a Gaussian distribution. This is especially the case when the VPE has a secondary peak in the tails of its distribution:

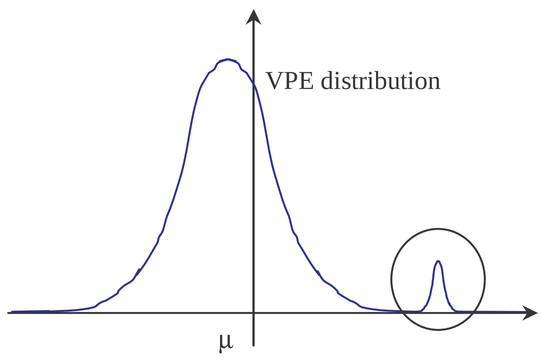

Figure 2: example of a non Gaussian distribution

In order to cope with such events, it should be noted that a Gaussian modeling can be applied to a non Gaussian VPE distribution provided that the modeling properly bounds the tails of the VPE distribution. However, the following question should be addressed:

With which Gaussian modeling, could the VPE distribution be bounded?

The idea is to inflate the VPE standard deviation σ up to a value Σ so that the Gaussian modeling $G(\mu,\Sigma)$ bounds (in the cdf overbounding sense) the tails of the VPE distribution, μ being the VPE mean value. This means that the energy contained in the tails of the Gaussian modeling should be above the energy contained in the tails of the VPE distribution. This process can be mathematically expressed:

If pdf(e) is the probability density function of a distribution, the energy of the distribution between $-\infty$ and L is:

$$cdf(L)=\int_{-\infty}^{L} pdf(e)\cdot de \qquad (10)$$

Therefore, if L1 is the limit of the left tail and L2 is the limit of the right tail of the distribution, the overbounding conditions are:

$$\begin{cases} cdf_{overbounding}(L1) \geq cdf_{VPE}(L1) \\ 1-cdf_{overbounding}(L2) \geq 1-cdf_{VPE}(L2) \end{cases} \qquad (11)$$

Σ is thus computed on the basis of the following identities:

$$\begin{cases} cdf_{overbounding}(L1) = cdf_{VPE}(L1) \\ 1-cdf_{overbounding}(L2) = 1-cdf_{VPE}(L2) \end{cases} \qquad (12)$$

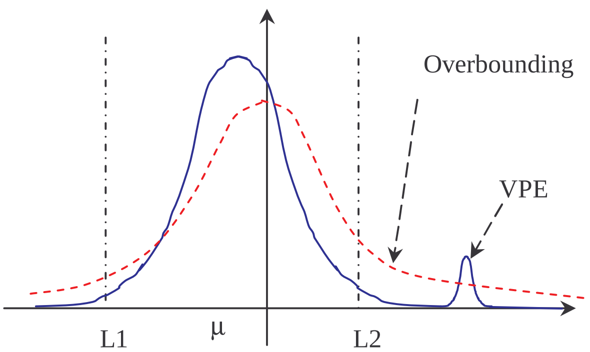

Figure 3: overbounding the tails

Unfortunately, μ and σ are only estimated through the respective values $\hat{m}$ and $\hat{s}$. Therefore, instead of defining a Gaussian modeling $G(\mu,\Sigma)$ as presented above, a Gaussian modeling $G(\hat{m},\hat{\Sigma})$ based on estimates is defined. Practically, the implementation of the cdf overbounding process consists, on the basis of another Gaussian modeling $G(\hat{m},\hat{s})$, in inflating $\hat{s}$ so that the new Gaussian modeling $G(\hat{m},\hat{\Sigma})$ bounds the tails of the VPE distribution.

It can be shown that the estimate of the energy contained in the left tail of the cdf overbounding Gaussian distribution is:

$$E_{left}(L1)=cdf_{overbounding}(L1)=\frac{1}{2}+\frac{1}{2}\cdot erf\left(\frac{L1-\hat{m}}{\hat{\Sigma}\cdot\sqrt{2}}\right) \qquad (13)$$

It can be shown that the estimate of the energy contained in the right tail of the cdf overbounding Gaussian distribution is:

$$E_{right}(L2)=1-cdf_{overbounding}(L2)=\frac{1}{2}-\frac{1}{2}\cdot erf\left(\frac{L2-\hat{m}}{\hat{\Sigma}\cdot\sqrt{2}}\right) \qquad (14)$$

The cdf overbounding should be sought whatever $L1 \leq \hat{m}$ and $L2 \geq \hat{m}$. This means that the condition (11) should be satisfied whenever $L1 \leq \hat{m}$ and $L2 \geq \hat{m}$.

However, in order not to introduce unnecessary sigma inflation due to errors close to $\hat{m}$, the maximum value of L1 and the minimum value of L2 are defined as followed:

$$MAX\,|L1| = \left| E_{left}^{-1}(40\%) \right| \qquad (15)$$

$$MIN\,|L2| = \left| E_{right}^{-1}(40\%) \right| \qquad (16)$$

(15) and (16) assume that the sigma inflation copes with 80% of the left tail and 80% of the right tail, disregarding 20 % of the overall error within the core of the distribution and centered on μ (being estimated through $\hat{m}$).

It should be noted that this cdf overbounding process does not aim to extrapolate a 10^{-7} error by multiplying the inflated estimate of the standard deviation by a K factor as does the SBAS integrity mechanism. The only purpose of this process is to define two parameters $\hat{m}$ and $\hat{\Sigma}$ to be used in the modeling of the uncertainty of μ and σ (addressed in the next step). In this context, the assumption of a centered VPE distribution is not necessary as it is in the context of the EGNOS VPL computation.

The next two figures illustrate the implementation of the cdf overbounding algorithm on selected data from the EGNOS System Test Bed.

Figure 4 shows that when the tail of the VPE distribution has no secondary peak, the estimated standard deviation of the Gaussian modeling is weakly inflated by a cdf overbounding algorithm. On this figure, it could be noted:

$$\begin{cases} \hat{s} = 1.13 \text{ meters} \\ \hat{\Sigma} = 1.22 \text{ meters} \end{cases}$$

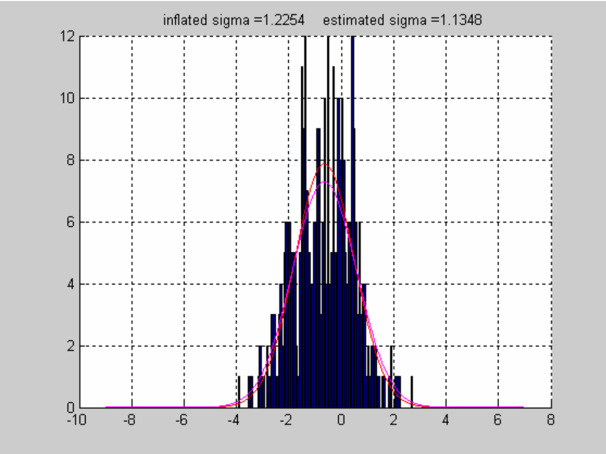

Figure 4: weak effect of cdf overbounding

On the other hand, figure 5 shows that when there is a strong secondary peak in the tail of the VPE distribution, the estimated standard deviation of the Gaussian modeling can be strongly inflated by a cdf overbounding algorithm. On figure 5 it should be noted:

$$\begin{cases} \hat{s} = 1.37 \text{ meters} \\ \hat{\Sigma} = 2.56 \text{ meters} \end{cases}$$

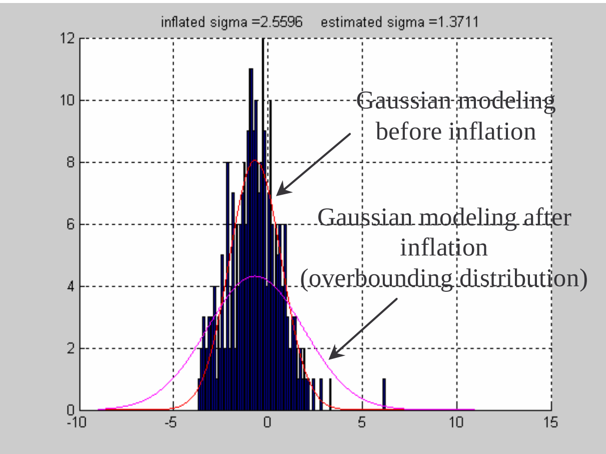

Figure 5 : strong effect of cdf overbounding

It should be felt that the smaller the number of processed samples, the higher the sigma inflation due to a single anomaly in the tails of the VPE distribution. Appendix 1 proves this statement and justifies the choice of selecting at least 200 samples for a data subset.

At this level of the analysis, an estimate of the mean value ($\hat{m}$) and an inflated estimate of the standard deviation ($\hat{\Sigma}$) of the VPE are determined. These two values are fundamental parameters to be used for modeling the uncertainty of μ and σ.

The third step of the analysis addresses the determination of this uncertainty: this will be the basis of the VPL confidence level determination.

Third step: allowing for the uncertainty on μ and σ

Because the statistical analysis is performed over a limited number of samples, the mean value and the standard deviation (μ and σ) of the underlying random process of the VPE cannot be perfectly known. Only estimates of them are computed ($\hat{m}$ and $\hat{\Sigma}$) through the previous step. Therefore, under this assumption it is not possible to compute an ideal VPL for a selected data subset (as was done in the first step thanks to the unrealistic assumption that μ and σ are perfectly known).

However, a modeling of μ and σ based on the knowledge of $\hat{m}$ and $\hat{\Sigma}$ will allow the estimation of the probability that the targeted VPL fulfills the EGNOS requirements.

In the frame of the analysis, this estimated probability is the confidence level in the VPL.

In other words, the confidence level in the VPL is an estimated probability that the VPL computed using EGNOS broadcast data is a 10^{-7} bound of the VPE.

The mathematical modeling of μ and σ and the theoretical process to compute the VPL confidence level are explained in appendix 2.

III. ESTB DATA PROCESSING FROM ENAC

III.1 Presentation of the data processing

The methodology presented in the previous sections was applied to data from ESTB DCN sites such as the ENAC site in Toulouse, France, between 28th May and 3rd June 2003.

On this 6 day data set, the following algorithm was implemented:

- Selection of subsets of VPE samples related to targeted VPL (from the lowest VPL value up to a 50 meters maximum)
- Implementation of the cdf overbounding algorithm in order to compute $\hat{\Sigma}$,
- Computation of the confidence level for each targeted VPL thanks to a modeling of μ and σ based on n, $\hat{m}$ and $\hat{\Sigma}$,

n, $\hat{m}$ and $\hat{\Sigma}$ being respectively the number of samples, the estimate of the VPE mean value and the inflated estimate of the VPE standard deviation of the subset related to a targeted VPL.

III.2 Overall results

The results are summarized in a graph presenting the confidence level in the VPL with respect to targeted VPL (cf. figure 6).

As depicted in figure 6, the processing of a 6 day data set from ENAC (from 28th May 00:00 UTC to 3rd June 2003 12:14 UTC), makes disparate results conspicuous.

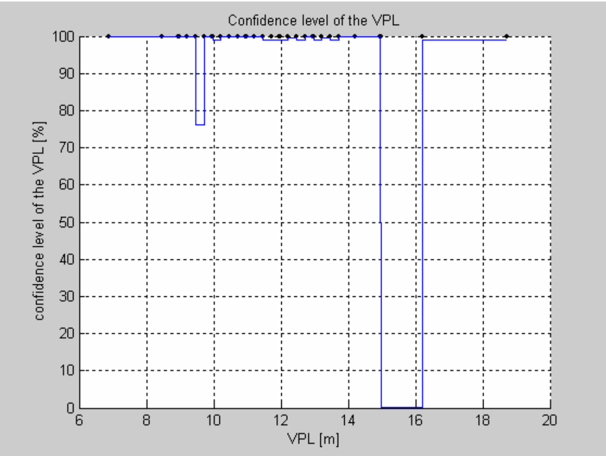

Figure 6: results from Toulouse

Good confidence levels can be noted except for a range of the VPL between 15 and 16 meters.

III.3 Analysis of the poor confidence level in ENAC

As showed in Figure 7, the poor confidence level is due to an anomaly in the tail of the selected VPE distribution when the VPL is comprised between 15 and 16 meters.

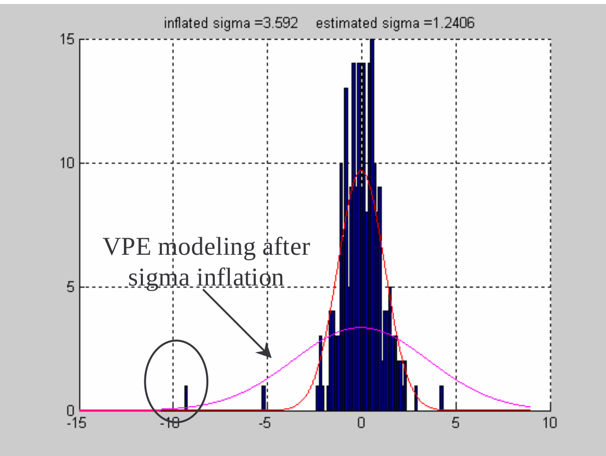

Figure 7: Poor confidence level in ENAC

The anomaly in the left tail of the VPE distribution induces a strong sigma inflation leading to a very poor confidence level below 1 %. Furthermore, for the selected data subset, figure 8 presents:

- The selected VPE samples and their corresponding targeted VPL,
- The estimated ideal VPL related to the selected VPE distribution,
- A value denoted VPL99 corresponding to a VPL that would have a 99 % confidence level according to the selected VPE,
- A value denoted VPL1 corresponding to a VPL that would have a 1 % confidence level according to the selected VPE,

being understood that the estimated probability that the ideal VPL is below VPL99 is equal to 99%. In the same way, the estimated probability that the ideal VPL is below VPL1 is equal to 1 %. The ideal VPL is therefore likely to be between VPL1 and VPL99[1].

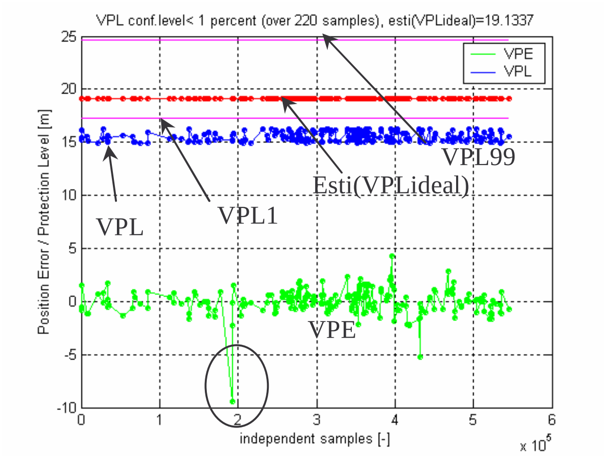

Figure 8: poor VPL confidence level in ENAC

Figure 8 shows that the EGNOS provided VPL is below the computed value VPL1. This means that the VPL confidence level is below 1 % for VPL values between 15 and 16 meters.

However, on figure 8, it should be noted that, even with this low confidence in the VPL the position error is always exceeded by the VPL so there is no integrity failure.

[1] The reliability of VPL1 and VPL99 has been verified through a simulation described in [5].

III.4 Analysis of a high confidence level in ENAC

In the majority of the cases, the confidence level in the VPL is very good such as in the following example.

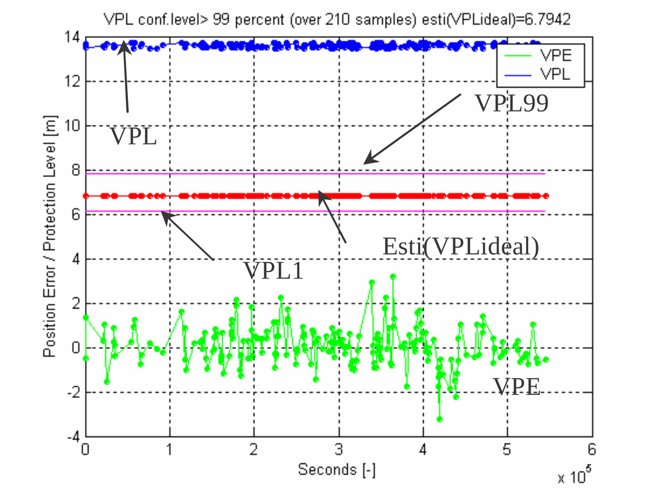

Figure 9: good confidence level in ENAC

This example shows that the confidence level in the analyzed VPL is very high (by far above 99%). In a pure statistical sense, such a VPL seems to be very conservative with respect to the selected VPE.

The VPE distribution related to the former example is presented in figure 10:

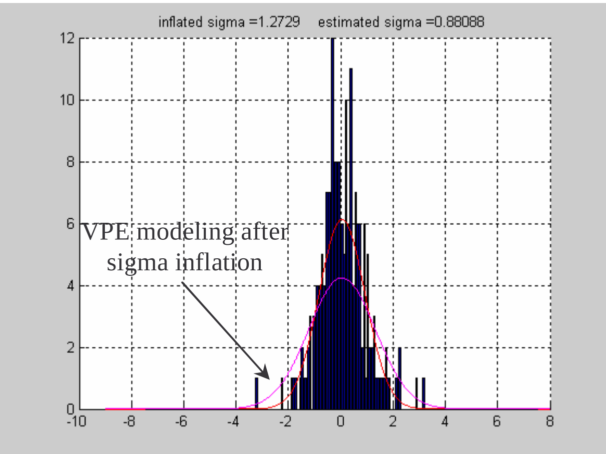

Figure 10: good confidence level in ENAC

From figure 10, it can be seen that only a small sigma inflation has been necessary to allow for the non Gaussian behaviour of the VPE distribution in its tails: indeed, both estimated and inflated sigma are close values. They respectively equal 0.88 meters and 1.27 meters.

More generally, all the performed analyses from several sites have shown that poor confidence levels in the VPL are always due to a strong sigma inflation (which is not the case in the previous example).

III. CONCLUSION AND FUTURE WORK

More than providing results, this paper proposes a methodology aiming at the assessment of the ESTB VPL.

It should be noted that the proposed statistical analysis of the position error over few hundreds of independent samples cannot formally be a complete validation tool. On the other hand, this methodology may provide indicators of the quality of the VPL.

Therefore, it will contribute to a better knowledge and understanding of the integrity mechanism and it will serve to highlight periods when the integrity performance should be analyzed in detail using other methods.

The implementation of the technique on real data sets has showed that no anomaly in the tails of the VPE distribution leads to high VPL confidence levels and poor VPL confidence levels are due to anomalies in these tails. Therefore, further analyses should be carried out in order to identify why such anomalies occur while not being compensated by a VPL increase.

Furthermore, the presented analysis is based upon ESTB data. The end of the EGNOS development phase being planned for the first quarter of 2004, the proposed methodology should be applied to the EGNOS Signal In Space when it will be available.

Finally, as this analysis has only addressed the vertical protection levels a similar methodology could be developed to assess the HPL.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of the people from the DCN for their contributions to the provision of data to be analyzed. Special thanks are due to P.B. Ober for the very beneficial advise he addressed during the writing of the paper.

REFERENCES

[1] ICAO, Standards and Recommended Practices, “Annex 10”, Volume 1, Radio Navigation Aids.

[2] “Civil Aviation requirements for EGNOS”, issue 3.3, Doc. Ref. OCR/DP/157, EUROCONTROL/GNSS programme, 2nd August 2001.

[3] ”The SBAS integrity concept standardized by ICAO. Application to EGNOS”, Benoit Roturier, Eric Chatre, Javier Ventura-Traveset, Seville, May 2001.

[4] ”Probability, random variables and stochastic processes”, third edition, McGRAW-HILL international editions, Athanasios Papoulis, 1991.

[5] ”EGNOS, a first step towards Galileo”, volume 2, David Comby, EUROCONTROL/GNSS programme, September 2002.

APPENDIX 1: Impact of the number of the processed samples on sigma inflation

The sigma inflation applied on the data (on the basis of the cdf overbounding algorithm) is necessary to take the non Gaussian behavior of the VPE distribution into account generally due to an anomaly in the tails of the distribution.

Indeed, through the analysis of the ESTB data it should be noted that the non Gaussian behavior of the VPE is generally due to an anomaly in the tails of the distribution.

In order to appraise the impact of the processed samples number on the sigma inflation, a simulation has been carried out. The simulation assumes a centered VPE, having a 1 meter standard deviation and following a Gaussian distribution except the tails that contain a single VPE sample of 10 meters as depicted in the following figure:

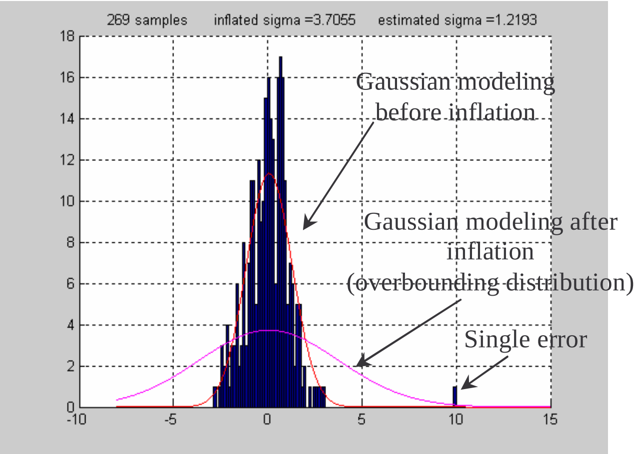

Figure 11: simulated VPE distribution (269 samples)

The question addressed in this appendix is:

How does the number of the processed samples help to mitigate the inflation of the VPE standard deviation due to a single error in the tails of the VPE distribution?

The scenario of simulation is based on the error previously introduced (considering a 10 meter single VPE in the tail of the distribution) with the number of samples varying from 30 to 1000.

Intuitively it can be felt that the higher the number of samples, the smaller the sigma inflation due to a single error in the tails of the distribution. This conclusion is confirmed by simulation results as illustrated in the next figure:

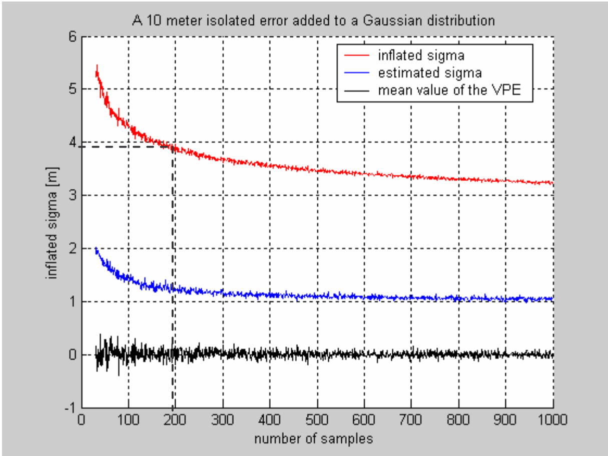

Figure 12: sigma inflation vs number of samples

Figure 12 shows that the sigma inflation applied in order to cover a tail anomaly becomes strong below 200 samples. From 30 to 200 samples, the inflated sigma is divided by a 1.3 factor whilst being divided by a 1.2 factor from 200 to 1000 samples.

The same conclusion could be drawn by considering a smaller anomaly in the tails (for instance a 4 meter single error) as depicted in figure 13:

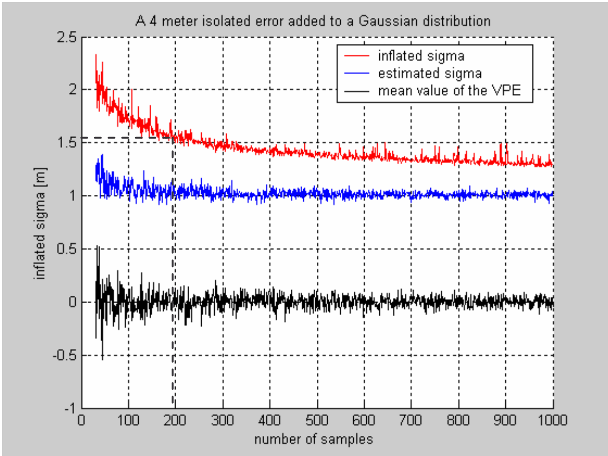

Figure 13: sigma inflation vs number of samples

Therefore, the selection of at least 200 samples seems to be a good trade-off between a reasonable sigma inflation and a reasonable size of a processed data file.

APPENDIX 2: UNCERTAINTY ON μ AND σ

Modeling of μ and σ

For each selected data subset related to a targeted VPL, an estimate of μ and σ (that are respectively the mean value and the standard deviation of the VPE) can be computed. Those estimates are theoretically random; μ and σ are unknown and deterministic.

In the frame of the analysis, roles are switched insofar as μ and σ are considered as two random variables and the estimates will be considered as being deterministic. This allows the computation of the VPL confidence level[1].

In this theoretical context, the purpose of this section is to determine the joint probability function of μ and σ. It will be noted that this function depends on three parameters:

- n, the number of independent samples over the selected data subset,
- $\hat{m}$, the estimate of μ,
- $\hat{\Sigma}$, the inflated estimate[2] of σ.

In order to express the joint probability density function f of μ and σ, it is useful to express both probability density functions of μ and σ, respectively $f_\mu(m)$ and $f_\sigma(s)$.

Knowing that $f_\mu(m)$ is the derivative function of the distribution function $F_\mu(m)$ of μ, it is useful to express $F_\mu(m)$. By definition, $F_\mu(m) = P(\mu \le m)$

Assuming that $t^{-1}_{(n-1)}$ is the inverse function of the student-t law with (n-1) degrees of freedom, and by writing:

$$m = \hat{m} + t^{-1}_{(n-1)}(\delta) \cdot \frac{\hat{\Sigma}}{\sqrt{n}} \tag{17}$$

a theorem of [4], states that:

$$F_\mu(m) = P\left(\mu \le m = \hat{m} + t^{-1}_{(n-1)}(\delta) \cdot \frac{\hat{\Sigma}}{\sqrt{n}}\right) = \delta \tag{18}$$

Using another property mentioned in [4], under the assumption that n>30 (a fortiori when n>200 as requested in appendix 1), the following approximation can be made:

$$t^{-1}_{(n-1)}(\delta) = z^{-1}(\delta) \cdot \sqrt{\frac{n}{n-2}} \tag{19}$$

(where $z^{-1}(\delta)$ is the inverse of the normal law N(0,1))

Therefore, $$m = \hat{m} + z^{-1}(\delta) \cdot \frac{\hat{\Sigma}}{\sqrt{n-2}} \tag{20}$$

And then, $$z^{-1}(\delta) = \frac{m-\hat{m}}{\hat{\Sigma}} \cdot \sqrt{n-2} \tag{21}$$

By definition, $$\delta = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z^{-1}(\delta)} e^{-\frac{u^2}{2}} \cdot du \tag{22}$$

Therefore, by using (18), (21) and (22) the following expression of $F_\mu(m)$ can be written:

$$F_\mu(m) = P(\mu \le m) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{m-\hat{m}}{\hat{\Sigma}} \cdot \sqrt{n-2}} e^{-\frac{u^2}{2}} \cdot du \tag{23}$$

Finally, knowing that $f_\mu(m) = \frac{dF_\mu(m)}{dm}$, the probability density function $f_\mu(m)$ of μ can be expressed:

$$f_\mu(m) = \sqrt{\frac{n-2}{2\pi}} \cdot \frac{1}{\hat{\Sigma}} \cdot e^{-\frac{n-2}{2 \cdot \hat{\Sigma}^2} \cdot (m-\hat{m})^2} \tag{24}$$

In the same way, knowing that $f_\sigma(s)$ is the derivative function of the distribution function $F_\sigma(s)$ of σ, it is useful to express $F_\sigma(s)$. By definition, $F_\sigma(s) = P(\sigma \le s)$.

In order to find an expression of $F_\sigma(s)$, an intermediate random variable X should be defined:

$$X = \frac{(n-1) \cdot \hat{\Sigma}^2}{\sigma^2} \tag{25}$$

Thanks to the assumption of independence of the n selected samples, it can be stated that X follows a chi-square law with (n-1) degrees of freedom:

$$P(X \ge x) = \int_x^{+\infty} \chi^2_{(n-1)}(t) \cdot dt \tag{26}$$

By writing $x = \frac{(n-1) \cdot \hat{\Sigma}^2}{s^2}$, it could be deduced:

$$\{X \ge x\} \Leftrightarrow \left\{\frac{(n-1) \cdot \hat{\Sigma}^2}{\sigma^2} \ge \frac{(n-1) \cdot \hat{\Sigma}^2}{s^2}\right\} \Leftrightarrow \{\sigma \le s\} \tag{27}$$

This allow deducing $P(\sigma \le s)$:

[1] The confidence level is an estimated probability that the VPL is a 10^{-7} bound of the VPE.

[2] This estimate is inflated in order to cope with the anomalies in tails of the VPE distribution.

$$F_\sigma(s) = P(\sigma \le s) = P\left(X \ge \frac{(n-1)\cdot\hat{\Sigma}^2}{s^2}\right), \text{ or:}$$

$$F_\sigma(s) = \int_{\frac{(n-1)\cdot\hat{\Sigma}^2}{s^2}}^{+\infty} \chi^2_{(n-1)}(t)\cdot dt, \text{ or:}$$

$$F_\sigma(s) = 1 - \int_0^{\frac{(n-1)\cdot\hat{\Sigma}^2}{s^2}} \chi^2_{(n-1)}(t)\cdot dt \qquad (28)$$

Finally, knowing that $f_\sigma(s) = \frac{dF_\sigma(s)}{ds}$, the probability density function $f_\sigma(s)$ can be expressed:

$$f_\sigma(s) = \frac{2\cdot(n-1)\cdot\hat{\Sigma}^2}{s^3}\cdot\chi^2_{(n-1)}\left(\frac{(n-1)\cdot\hat{\Sigma}^2}{s^2}\right) \qquad (29)$$

It is now possible to conclude on the expression of the joint probability density function of both random variables μ and σ. Indeed, knowing that μ and σ are independent, it can be stated:

$$f(m,s) = f_\mu(m)\cdot f_\sigma(s) \qquad (30)$$

Therefore, through the knowledge of $f_\mu(m)$ and $f_\sigma(s)$ given by (24) and (29), the function $f(\mu,\sigma)$ is now totally determined:

$$f(\mu,\sigma) =$$

$$(n-1)\cdot\sqrt{\frac{2\cdot(n-2)}{\pi}}\cdot\left(\frac{\hat{\Sigma}}{\sigma^3}\right)\cdot\chi^2_{(n-1)}\left(\frac{(n-1)\cdot\hat{\Sigma}^2}{\sigma^2}\right)\cdot e^{-\left(\frac{(n-2)}{2\cdot\hat{\Sigma}^2}\cdot(\mu-\hat{m})^2\right)}$$

(31)

Computation of the VPL confidence level

Introduction

For each selected data subset related to a targeted VPL, the modeling of μ and σ introduced in the previous section allows the computation of an estimated probability that the VPL is a 10^{-7} bound of the VPE: this probability is referred to as the confidence level in the targeted VPL.

The computation process is presented through several steps.

First step of the process presentation

A theoretical assumption should be made : over each selected data subset related to a targeted VPL, each sample of the VPE is assumed to follow the same Gaussian distribution totally determined by its mean value μ and its standard deviation σ[1]. From this assumption, the following statement can be concluded: The probability that at least one VPE sample exceeds the VPL over the selected data subset is a function of both μ and σ. This function is denoted $g(\mu,\sigma)$ and will be expressed later on.

Second step of the process presentation

For each selected data subset, μ and σ are modeled as two random variables on the basis of:

- n, being the number of samples for the selected data subset,
- $\hat{m}$, being the estimate of the VPE mean value for the selected data subset,
- $\hat{\Sigma}$, being the inflated estimate of the VPE standard deviation for the selected data subset after a cdf overbounding process.

Therefore, the probability that the VPE exceeds the targeted VPL over the selected data subset could be modeled as a random variable that could be denoted Z:

$$Z = g(\mu,\sigma) \qquad (32)$$

Third step of the process presentation

From the previous step, a theoretical expression of the VPL confidence level (denoted CL) over the selected data subset can be written:

$$CL = P\{Z \le R_n\} \qquad (33)$$

where R_n is the acceptable probability that at least one of the n independent VPE samples over the subset exceeds the VPL. R_n will be expressed later on thanks to the EGNOS requirement for the VPL introduced in §I.2.

In another way, an equivalent equation can be expressed:

$$CL = \iint_D f(\mu,\sigma)\cdot d\mu\cdot d\sigma \qquad (34)$$

where $f(\mu,\sigma)$ is the joint probability density function of both random variables μ and σ expressed in (31), and the domain D of integration is defined as follows:

[1] In fact, the VPE does not follow a Gaussian distribution. However, one can make this assumption because the Gaussian modeling should bound the VPE distribution through a cdf overbounding algorithm.

$$\{ (\mu,\sigma) \in D \} \Leftrightarrow \{ Z \le R_n \} \quad (35)$$

$$\{ (\mu,\sigma) \in D \} \Leftrightarrow \{ g(\mu,\sigma) \le R_n \} \quad (36)$$

The computation of (34) requires:

- the expression of R_n,
- the expression of $g(\mu,\sigma)$,
- a frame of the domain D of integration.

Expression of R_n

For the i^{th} sample of the selected data subset, the event denoted A_i is defined:

$$A_i = \{ VPE_i > VPL_i \} \quad (37)$$

The probability (denoted P_n) that at least one VPE of the subset exceeds its corresponding VPL can be expressed by using A_i:

$$P_n = P\left(\bigcup_{i=1}^{n} A_i\right) \quad (38)$$

By assuming that the n events A_i are independent, the following identity can be written:

$$P_n = 1 - \prod_{i=1}^{n} P\left(\overline{A_i}\right) = 1 - \prod_{i=1}^{n} \left[1 - P(A_i) \right] \quad (39)$$

Knowing that the EGNOS requirement leads to $P(A_i) \le 10^{-7}$ per independent sample (see § I.2), the requirement R_n for a n sample data subset can be expressed:

$$P_n \le R_n = 1 - \prod_{i=1}^{n} \left[1 - 10^{-7} \right] \quad (40)$$

Finally,

$$R_n = 1 - \left(1 - 10^{-7} \right)^n \cong n \cdot 10^{-7} \quad (41)$$

Expression of $g(\mu,\sigma)$

Assuming that over each selected subset every VPE sample follows the same Gaussian distribution[1] (totally determined by μ its mean value and σ its standard deviation), $g(\mu,\sigma)$ should express the probability P_n that at least one VPE sample of the subset exceeds the VPL.

[1] This is assumed to be met because the selected VPE samples are related to close values of the VPL.

The probability $P(A_i) = P(VPE_i \ge VPL_i)$ can be written in the following way:

$$\forall\, i \in \{ 1,\ldots,n \},\ P(A_i) = 1 - \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-VPL_i}^{+VPL_i} e^{-\frac{(t-\mu)^2}{2\cdot\sigma^2}} \cdot dt \quad (42)$$

From (42) an equivalent expression can be derived:

$\forall\, i \in \{ 1,\ldots,n \}$,

$$P(A_i) = 1 - \frac{1}{2} \cdot \left[\mathrm{erf}\left(\frac{VPL_i + \mu}{\sigma \cdot \sqrt{2}} \right) + \mathrm{erf}\left(\frac{VPL_i - \mu}{\sigma \cdot \sqrt{2}} \right) \right] \quad (43)$$

Finally, knowing that $g(\mu,\sigma) = P_n$, an expression of $g(\mu,\sigma)$ can be derived from (39) and (43):

$$g(\mu,\sigma) = 1 - \frac{1}{2^n} \cdot \prod_{i=1}^{n} \left[\mathrm{erf}\left(\frac{VPL_i + \mu}{\sigma \cdot \sqrt{2}} \right) + \mathrm{erf}\left(\frac{VPL_i - \mu}{\sigma \cdot \sqrt{2}} \right) \right] \quad (44)$$

Frame of the domain D of integration

In order to compute the confidence level in the VPL for a selected data subset, one should have an idea of the frame of the integration domain D for the integral (34). The knowledge of a frame of D would allow the implementation of an algorithm for computing (34).

From (44) it can be proven that:

$$\forall\, \mu \ne 0, \forall\, \sigma > 0,\ g(0,\sigma) \le g(\mu,\sigma) \quad (45)$$

Furthermore, $\forall \mu$ the partial function $p(\sigma) = g(\mu,\sigma)$ is increasing. It can be deduced that the biggest value of σ for which $g(\mu,\sigma)$ remains below R_n is obtained when $\mu=0$. Therefore, the maximum value of σ within the integration domain D belongs to the $O\sigma$ axis (when $\mu = 0$). It can also be shown that the maximum absolute value of μ within D is necessarily below the maximum value of the VPL over the selected interval:

$$\forall\, (\mu,\sigma) \in D,\ |\mu| \le \underset{i \in \{1,\ldots,n\}}{MAX} (VPL_i) \quad (46)$$

A frame of the domain D of integration can thus be illustrated by the following figure:

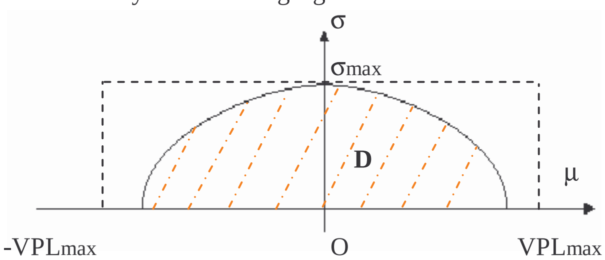

Figure 14: domain D of integration