Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2002-310

A practical interpretation of performance requirements for a Global Navigation Satellite System

What does the user really need?

A.N. van den Berg and P. Dieleman

Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2002-310

A practical interpretation of performance requirements for a Global Navigation Satellite System

What does the user really need?

A.N. van den Berg and P. Dieleman

This report is based on a presentation held at ENC-GNSS2002, Copenhagen, Denmark, 27-30 May 2002.

Customer:	National Aerospace Laboratory NLR
Working Plan number:	R.1.B.3
Owner:	National Aerospace Laboratory NLR
Division:	Space
Distribution:	Unlimited
Classification title:	Unclassified
	June 2002

Contents

A practical interpretation of performance requirements for a Global Navigation Satellite System

What does the user really need?

Axel van den Berg and Peter Dieleman,
avdberg@nlr.nl, dieleman@nlr.nl
National Aerospace Laboratory **NLR**
Amsterdam, The Netherlands

Biography

Axel van den Berg was born in Amsterdam, the Netherlands in 1971. After obtaining the degree of MSc. in Aerospace Engineering at the Delft University of Technology in 1997 he joined the Space Division of the National Aerospace Laboratory NLR. Starting on the GNSS-1 Operational Validation study for Eurocontrol he became a GNSS safety and performance expert. Now, his main activity is Galileo System RAMS analysis in co-operation with Galileo Industries (GaIn).

Peter Dieleman was born in the Netherlands in 1959. After obtaining an MSc. in Electrical Engineering at the Delft University of Technology in 1983 he worked for several universities, research institutions, and industries. His work focussed on control engineering, and automated systems development with a focus on system design and software engineering. In 1999 he joined the Space Division of the National Aerospace Laboratory NLR where he has been involved in the definition of a GNSS performance monitoring facility and since 2000 in the definition of GALILEO with focus on RAMS and System Verification in co-operation with European space industry.

Background

A uniform interpretation of the user requirements for satellite navigation systems appears quite difficult to achieve. System users, system designers and (navigation) specialists are examples of actors that play a role in defining, designing, building and using such systems. All of these actors use similar words and phrases but use a different language. Obviously miscommunications are the result.

Navigation specialists in the field of aviation have gone through an extensive process to define exactly what they require from a (satellite) navigation system for operational purposes. After many reviews, they came up with a short and simple set of definitions and requirements (SARPs [1.]), which gradually became the basis for all user requirements stated in terms of Accuracy, Integrity, Continuity and Availability.

Space systems engineers take the resulting set of definitions and requirements as a starting point for defining their system requirements for a satellite navigation system to be newly built. This sounds okay so far. However, the carefully chosen wording in the SARPS [1.] is easily overlooked, and due to an overlap with the vocabulary used by system engineers and navigation experts, many terms obtain multiple meanings or interpretations. All this results in confusion and wrong interpretations that severely affect the system's design and verification.

On top of this it should be noted that the GNSS user requirements originate from an already existing navigation system (i.e. GPS [4.]). These requirements do not sufficiently take into account the flexibility one has when designing a completely new system from scratch. Furthermore, political and commercial interests force the system to be defined in a more complex form than necessary, in order to make it suitable to *any* user domain. While the core (limited) nature of the system: "provision of accurate and reliable ranging sources" is overshadowed.

The result of all this could lead to a complex, and difficult-to-verify set of system requirements for the European GNSS (i.e. GALILEO [2.], [3.]). Therefore the following simple question needs a clear and unambiguous answer:

"What does the Safety of Life user really need from a satellite navigation system?"

The user community should be assured that "the right system is being designed" by checking the correctness/usefulness and the interpretation of the system requirements. The industry and customer institutions who drive the development of the system should be assured that "the system is designed right" by system verification. All actors would benefit from a clear and verifiable set of requirements with one commonly agreed interpretation.

System description

For clarification a short description of GNSS is provided, as seen from a user receiver perspective. GPS [4.] is the actual operational instantiation of GNSS, that is being augmented with space based regional overlay systems such as EGNOS [3.], WAAS [5.] and MSAS [6.]. Next to this, ground based augmentation systems are being developed such as LAAS [5.]. In addition Europe has started the development of its own GNSS: GALILEO. The GALILEO [2.] definition with its system requirements as presently available is used as an example here, as it is supposed to be developed in accordance with up-to-date user needs, and state-of-the-art satellite navigation concepts and technology.

GALILEO will provide a number of ranging sources from the GALILEO Space Segment: a constellation of 27 satellites, controlled by the GALILEO Ground Segment. For the satellite-only services provided by GALILEO all information for the user is provided through these ranging signals. In addition to the information required to determine the Pseudo Ranges (PR) between satellites-in-view and the user's receiver, each ranging source contains error-bounding information concerning the PR (SISA) that can be used with the momentary satellite geometry to determine an error-bound in the position domain: the Protection Level (PL). As SISA is not updated sufficiently frequently (order of hours) for many applications, in addition a near real-time (order of seconds) "Integrity Flag" (IF) is provided, that can be used to exclude unreliable signals from the position solution, as determined by the receiver, within the specified Time To Alert (TTA). This TTA defines the time between the occurrence of a condition in the GALILEO system that should lead to an alert for the user, and the actual alert being provided to the user by the receiver.

It is important to note that the performance of GNSS for a user ultimately depends on the achievable Protection Level as computed in the user receiver. As stated above, this PL depends on the geometry of the satellites providing the signals used (i.e. after exclusion of unreliable ones) and the error bounds (SISA) provided for each ranging signal, adjusted by the local/receiver ranging error budget estimates.

System boundary

Although it is clear that some minimum functionality of the user receiver (i.e. Minimum Operational Performance Specifications) should be considered while designing GNSS, the user receiver should be placed outside the system boundary. The main reason for this is the wide variety of receiver implementations and configurations possible, even for comparable applications. It complicates system requirements dramatically, in case all possible receiver configurations including additional aiding sensors, have to be addressed.

It is therefore proposed to define the system performance entirely in terms of the Signal in Space (SIS) domain. The detailed specification of the user receiver's maximum error budgets and its response to information provided by the system should be sufficient to determine, in combination with simulation of the geometry and a set of environmental assumptions, the performance in the (user) position domain.

Apart from the technical benefits in system design, system implementation and system verification, placing the user receiver outside the system boundary would enable a clear distinction to be made between:

- Service provider responsibility: restricted to SIS and constellation geometry
- Receiver certification: restricted to receiver performance assuming various nominal and degraded operational modes of GNSS service provision
- User operations approval: separation of service provision quality, receiver quality, and operational procedures

Interpretation of performance requirements

In this paper the emphasis is on the safety critical performance requirements of Integrity and Continuity. The main objective of this discussion is to arrive at an interpretation of the user requirements that is practical for the system designer and acceptable for interpretation by the user.

Service provision

For a GNSS, many different navigation service levels can be described. However, the general functionality will be the same.

The main objective for a GNSS is to provide positioning (PVT) information to the user. The quality of this positioning information is dimensioned in terms of Accuracy, Integrity, Continuity and Availability. Active provision of an accuracy guarantee (Protection Level) is a method to optimise the balance between Accuracy, Integrity and Availability. This functionality is often referred to as Integrity Service, thereby creating even more confusion around the meaning of the word "Integrity".

In this paper the performance requirements, Accuracy, Integrity, Continuity and Availability are treated as quality parameters of the combined service of positioning and the accuracy guarantee/alert mechanism.

Accuracy

Accurate positioning is the primary goal of a GNSS. It is crucial to distinguish the following forms of accuracy for a GNSS:

- Desired Accuracy statistics or the distribution of the (true) error statistics. This is normally stated in terms of a **95% Navigation System Error** bound (NSE95%)
- Instantaneous Error, the real error of the position solution, which is not known to the user. Also called the **Position Error** (PE). Note that this could occasionally be higher than the NSE95% target.
- Instantaneous Accuracy guarantee, the information as provided through the system, which bounds the true instantaneous Position Error. Also called the **Protection Level** (PL). Note that this is the "known" accuracy for the user.
- Required Accuracy, the maximum allowable Position Error for a certain operation. Also called **Alert Limit** (AL).

The user (receiver) will determine his availability primarily by the equation PL < AL.

Integrity

Integrity is a quality of the positioning and accuracy guarantee service provided.

For a user, Integrity is not assured when the true Positioning Error (PE) exceeds the user's maximum required accuracy (or Alert Limit) without an Alert for longer than the Time To Alert (TTA). An Alert to the user is generated based on the Protection Level exceeding the Alert Limit, which implies that, for the system point of view, Integrity is not assured when the true Positioning Error exceeds the Protection Level (PL). This undesired event is also called an Integrity Event, meaning an event when the system provides Hazardous Misleading Information (HMI) at the user output (or Integrity Risk is not assured). Note that this has nothing to do with statistics.

Integrity Risk is the user-defined acceptable level of risk to experience an Integrity Event. This Risk should be assured anywhere and any time a user is relying (within coverage and operational constraints) on the GNSS for a (critical) operation. Consequently, it is not permitted to average out the probability of an Integrity Event over time or location and therefore, the Integrity Risk is stated in terms of probability of occurrence per 150 seconds. The 150 seconds relate to a typical duration of a critical operation for aviation (final approach).

For the analysis of Integrity through the system, the above implies that the worst geometry and system configuration should be considered under which **a service is declared available**.

The following explains the difference between Integrity seen from the user point of view and as seen from the system point of view. In figure 1 all possible configurations for the Position Error (PE) +, with respect to the Alert Limit (AL) and the Protection Level (PL) are shown (in only two dimensions). Note that in the positioning domain there is no relevance for an Integrity Flag (IF), as these are only defined at SIS level

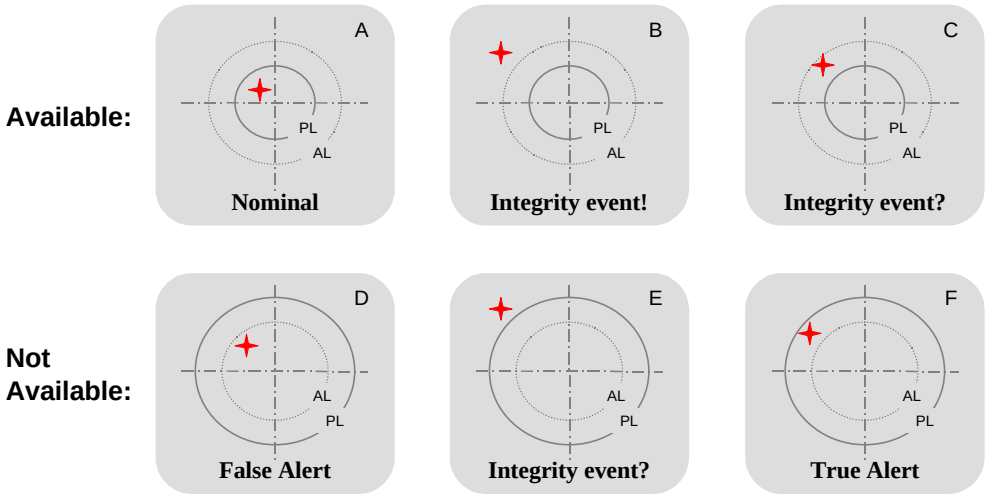


Figure 1 Integrity definition in a two dimensional plane

and they are applied before computing the Protection Level.

A. is the **nominal** situation in which the system is available and correctly bounding the Positioning Error;

B. is the most feared situation where the user with a sufficiently low Alert Limit experiences an undesirable **Integrity Event***;*

C. although for a user with a sufficiently high Alert Limit there is nothing wrong, the system is not correctly bounding the Positioning Error, i.e. a **System Integrity Event**

D. when the system provides a much too conservative bound, the user might be unnecessarily alerted for a non existing user Integrity Event, i.e. a **False Alert**;

E. although a user with a sufficiently low Alert Limit is correctly alerted for an unacceptable Position Error, the system is not correctly bounding the Position Error, i.e. a **System Integrity Event** ;

F. this is no doubt a safe situation, as in this case the Position Error is properly bound by the Protection Level and the user is alerted not to use the system, i.e. a **True Alert**;

When considering that the Alert Limit could differ for every user or operation, this AL should not be the scope of the system's Integrity Risk. It is assumed that (in the worst case) the Protection Level equals the Alert Limit. Therefore the *Protection Level* should be considered as the guaranteed limit on which the Integrity Risk is based when defining the *system* performance. Leaving out the dotted circles for the Alert Limit has no impact on the interpretation of Integrity. It only impacts Availability.

Because the Protection Level still comprises local receiver effects such as signal selection and local environmental effects, it is advisable to define the system Integrity Risk in the Signal in

	IF = Good	IF = Alert	IF = Not Monitored
SISA > SISE	• **Available** • **Integrity good**	**False Alert** • **Unavailable** • **Integrity good**	• **Unavailable** • **Integrity good**
SISA < SISE	**Hazardous Misleading Information**	• **Unavailable** • **Integrity good**	**Misleading Information** • **Unavailable**
No SISA	• **Unavailable**		

Table 1 Integrity and availability in the SIS domain.

Space (SIS) domain. The same logic can be applied: the Position Error (PE) can be replaced by the SIS Error (SISE) and the Protection Level (PL) can be replaced by the SIS Accuracy (SISA). In the SIS domain there is one additional protection in the form of an Integrity Flag (IF), which informs the user in near real-time whenever it is detected that a SISA parameter is exceeded by the SISE. The resulting configurations are described in Table 1, where the only real Integrity problem is the Hazardous Misleading Information provided by one Signal. It can be (conservatively) assumed that only one hazardous misleading ranging source (SIS HMI) could already result in an (user) Integrity Event.

Continuity

Continuous provision of positioning and accuracy guarantee data is a quality of the service.

For a user, continuity is not assured when he has to abort his current operation. To translate this to the system, it is necessary to know on what basis the decision to abort is taken. A discontinuity, or Continuity Event, can be caused by any of the following events:

- Protection Level changes and exceeds the Alert Limit from: PL < AL into PL > AL. Note that this could be caused by both True and False Alert conditions.
- No Position solution update is generated for more than (suggestion) six seconds. Note that according to the Integrity requirement a user is allowed to experience six seconds of HMI guidance, which is more unsafe than continuing six seconds blind on an initially safe course.
- No Integrity Message is received at the receiver for more than (suggestion) two seconds. Two seconds would allow permitting short outages (e.g. data packet loss), while the probability that any feared event occurs at the same instant is negligible (even if it would be the case, only the Time to Alert would be increased with two seconds.
- A Receiver Autonomous Integrity Monitoring (RAIM) alert is generated.
- *Known* geometry or system status changes into a configuration where the Integrity Risk is *known* to be unacceptable (e.g. no RAIM available). It could be justified to permit to continue an operation for a short period.

It is assumed that when any of these events occur, the user is instantly notified by an “abort” alert, and will abort the current operation unless he has other means to proceed and finalise his activity in a sufficiently safe manner. Note that the list of Continuity Event causes does not cover the Integrity Event. If this was known to a user, it would not be an Integrity Event.

Continuity of the service has to be assured anywhere, any time over 15 seconds following a certain point in time at which there was no reason to expect a discontinuity of the service with a probability higher than the Continuity requirement. This implies that the user assumes a decision moment. The only reasonable assumption would be that this decision moment is at the start of a (short-term) critical operation during which an abort is undesirable. A typical critical operation for aviation lasts about 150 seconds (see also the Integrity requirement). The statement anywhere, any time already prohibits to average over location or time and implies to use a worst-case geometry scenario for system performance analysis of continuity. Therefore it is strange, unnecessary and only confusing to state the continuity requirement in a different interval than the Integrity requirement (15 seconds opposed to 150 seconds).

It is recommended to state the continuity requirement in a Probability/150 seconds and to define a minimum duration for positioning and accuracy guarantee information interruptions.

The nature of the requirement implies that for the assessment of Continuity Risk (TLH-2) one should consider the worst case condition under which **a service is declared available (at T_0)**.

Availability

A definition of availability [American National Standard for Telecommunications]:

> “The degree to which a system, subsystem, or equipment is operable and in a committable state at the start of a mission, when the mission is called for at an unknown, i.e., a random, time. Note: The conditions determining operability and committability must be specified”

The theoretical "Availability of Accuracy" that is related to the statistical Accuracy requirement (NSE95%) does not relate to the actual real term user Availability. In practise, this availability can only be determined with hindsight or by experiments. This is not one of the conditions mentioned above.

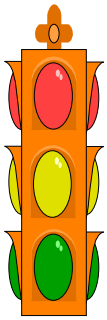
The actual situation, when the user will not be able to make use of the service, is governed by the known information and decisions in the user receiver. The following decision triggers can be assumed (in other words, the service is declared available if):

- PL < AL, assuming PL and PVT are available;
- Minimum redundancy for critical functions is assured (e.g. for Galileo and EGNOS Integrity Messages are available via two or more independent links);
- Receiver Autonomous Integrity Monitoring (RAIM) status is correct;
- RAIM function is available;
- Short-term (e.g. 150 seconds) prediction based on current available local geometry does not reveal eminent discontinuity events or increased risk of discontinuity. This prediction could include satellite orbit propagation, user propagation, masking, and navigation data validity time.

Note that [PE < NSE95%] or [PE < AL] is not in the list of unavailability events because if the real position error was known in the receiver, one would not need GNSS in the first place!

From the discussion on Integrity and Continuity it follows that it is not necessary to specify separately the availability of Integrity and Continuity as this is already implicit in the I&C requirements.

For availability, the following conditions should be identified:



1. Alert to "**Abor**t" safety critical operation because PL > AL or PVT is lost
2. warning: "**Do not initiate**" new critical operation, due to increased probability of discontinuity.
3. **Nominal** condition: where the system is usable and all performance requirements are met.

The user geometry is predictable and known within the user receiver at each point in time-space. This information could be used to decide at user-level if a service is considered available.

Real-time information on the system's internal configuration status (e.g. failed back-ups) is not known at the user receiver, unless specific information is provided to (and used by) the receiver. Degraded system configurations could increase the risk of Continuity or Integrity Events. It is not possible to inform the user of all degraded system configurations. Moreover, this would result in more unwanted False Alerts. It can be justified that as long as for redundant elements the ratio MTTR/MTBF is sufficiently small (<<1), i.e. the first failure is repaired before second failure can be expected) the full redundant configuration can be considered. For non (or long-term) repairable redundant elements the worst case configuration for which a sub-system is still declared operational should be considered (e.g. satellites).

Summary performance interpretation

For safety analysis, the minimum ranging source geometry should be considered under which a user receiver does not generate a "don't use" warning (e.g. for degraded Integrity performance).

- Consequence for Integrity: if RAIM is not available under a certain geometry, this should either result in a "do not use" or "abort" at receiver output or the Integrity Risk requirement should be met even without RAIM.

- Consequence for Continuity: When the user is allowed to initiate a safety critical operation under the condition that he experiences one or more critical satellites[1], the continuity performance must be met, even with the single point failure of one single satellite (including outages, True and False Alerts)

- The start of a new operational phase (T_0) can be considered a decision moment for availability, with the two options: available or "don't use"
- The receiver should be able to generate output in three levels:
  - "nominal" or "use", when there is no reason to expect problems in the short term (150 seconds)
  - "don't use" or "do not initiate", when the known Accuracy performance (i.e. PL) is still available and within limits, but discontinuity risk is known to be not within acceptable limits
  - "abort" when a Continuity Event occurs (either PL exceeds AL or the Integrity risk is known to be not within acceptable limits)

Conclusions

The most important findings are the following:

Availability depends on the system status as it is known to the user, and the receiver Alert logic. It does not depend on the actual Positioning Error, which is, by definition, unknown to the user receiver.

A user receiver has three output states:
- Nominal
- "do not initiate"
- "Abort"

Integrity Risk and Continuity Risk should be specified in similar time periods, preferably the typical duration of a critical operation (150 seconds).

Interruptions that result in a discontinuity must have a minimum duration specified
- position data (x seconds)
- integrity data (y seconds)

Proper interpretation of user requirements and translation into system requirements is not a trivial task. This paper suggests a practical way to make this interpretation, in order to simplify both performance analysis and system verification and validation. Some areas, where the user requirements are somewhat ambiguous, have been identified.

It is considered important to closely involve the user community in further iterations to reach commonly understood and agreed system requirements between users and developers, such that the question: **"Will the right system be developed?"** will eventually be answered positively.

References

[1.] ICAO GNSS Panel, Standards and Recommended Practices, version 8.0, February 1999.
[2.] European Commission Galileo website: http://www.europa.eu.int/comm/energy_transport/en/gal_en.html
[3.] ESA Navigation website: http://www.esa.int/export/esaSA/navigation.html
[4.] USCG GPS website: http://www.navcen.uscg.gov/gps/default.htm
[5.] FAA WAAS and LAAS website: http://gps.faa.gov/Programs/index.htm
[6.] Japanese MSAS website: http://www.mlit.go.jp/koku/ats/e/mtsat/miss/03.html

Acknowledgements

This paper represents the authors' personal view. Special thanks should go to Stephane Journo (Alcatel Space) and Edward Breeuwer (ESA).

[1] Critical satellite refers to a ranging source that is required to keep the PL > AL, if it is lost or excluded the PL will drop below the AL and thus cause a discontinuity.