GNSS Position Integrity in Urban Environments: A Review of Literature

Ni Zhu, Juliette Marais, David Betaille, Marion Berbineau

Institutions: IFSTTAR

Published on: 05 Jan 2018 - IEEE Transactions on Intelligent Transportation Systems (IEEE)

Related papers:
- Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems
- Baseline advanced RAIM user algorithm and possible improvements
- Multi-Constellation GNSS Performance Evaluation for Urban Canyons Using Large Virtual Reality City Models
- Advanced RAIM user Algorithm Description: Integrity Support Message Processing, Fault Detection, Exclusion, and Protection Level Calculation
- Lane-Level Integrity Provision for Navigation and Map Matching With GNSS, Dead Reckoning, and Enhanced Maps
GNSS Position Integrity in Urban Environments: A Review of Literature

Ni Zhu, Juliette Marais, David Betaille, Marion Berbineau

To cite this version:

Ni Zhu, Juliette Marais, David Betaille, Marion Berbineau. GNSS Position Integrity in Urban Environments: A Review of Literature. IEEE Transactions on Intelligent Transportation Systems, IEEE, 2018, 17p. 10.1109/TITS.2017.2766768. hal-01709519

HAL Id: hal-01709519
https://hal.archives-ouvertes.fr/hal-01709519
Submitted on 15 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—Integrity is one of the criteria to evaluate GNSS performance, which was firstly introduced in the aviation field. It is a measure of trust which can be placed in the correctness of the information supplied by the total system. In recent years, many GNSS-based applications emerge in the urban environment including liability critical ones, so the concept of integrity attracts more and more attention from urban GNSS users. However, the algorithms developed for the aerospace domain cannot be introduced directly to the GNSS land applications. This is because a high data redundancy exists in the aviation domain and the hypothesis that only one failure occurs at a time is made, which is not the case for the urban users. The main objective of this paper is to provide an overview of the past and current literature discussing the GNSS integrity for urban transport applications so as to point out possible challenges faced by GNSS receivers in such scenario. Key differences between integrity monitoring scheme in aviation domain and urban transport field are addressed. And this paper also points out several open research issues in this field.

Index Terms—GNSS, Integrity, urban environment, protection level (PL)

I. INTRODUCTION

The GNSS integrity concept has been firstly developed and formalized in the aviation field for Safety-of-Life (SoL) applications [1]. It is defined as a measure of trust which can be placed in the correctness of the information supplied by the total system [2]. As one of the most essential performance parameters, GNSS integrity has recently attracted interest from other transportation fields especially in the urban environment. This is because the GNSS-based urban applications proved to be a huge and appealing market which is currently in a constant growth [3].

For GNSS land applications such as the rail and the vehicular domains, knowing the certainty of one’s localization is of great importance. The framework of GNSS integrity in urban environment is firstly introduced especially in the vehicle domain, for instance, the famous Liability Critical Applications, here the computed Position, Velocity and/or Time (PVT) are used as the basis for legal decisions or economic transactions [4] [5], such as Electronic Toll Collection (ETC) and Pay as you Drive insurance. In such kinds of scenario, large errors can lead to serious consequences such as wrong legal decisions or wrong charge computation as the example shown in Fig. 1. In addition, an increasing number of Unmanned Aerial Vehicles (UAV) in urban environment require also high integrity performances [6] since multipath effects associate with their low-level flights. Consequently, it is necessary and important to bound the errors and to ensure that the probability of errors not properly bounded is below a certain limit in order to reduce the probability of the harmful effects and to guarantee the correctness and fairness of the decision. These requirements attach extreme importance to the concept of positioning integrity in urban environment.

However, the urban environment presents great challenges to common commercial GNSS receivers [8] [9]. This is mainly because the GNSS positioning performance can be severely degraded by the limited satellite visibility, multipath effect, interference and other undesired impairments such as foliage attenuation [10] [11] [12]. Much research has been developed in terms of techniques to mitigate the effect of multipath interference and Non-line-of-sight (NLOS) signals at different levels, for example, the antenna design techniques [13] [14], the receiver-based techniques [15], as well as the post-receiver techniques [16], which help to improve accuracy and reliability of the GNSS positioning in urban environment. But these techniques are still an issue to be ceaselessly developed especially for its compatibility and robustness to different stringent environments.

Despite the existing difficulties, introducing the integrity concept to urban GNSS receivers is more and more attractive as a result of emerging GNSS-based applications in stringent environments. But the integrity monitoring algorithms developed in the aviation domain cannot be transported directly into the urban vehicle applications. This is because, on the one hand, the integrity monitoring algorithms developed in the aviation context are established on the fact that a high data redundancy exists, which is not the case in the urban environment.
context. On the other hand, the single-fault assumption made in the aerospace applications is not true for urban GNSS receivers due to the potentially large and frequent errors provoked by multipath interference and NLOS. [17]

This paper is organized as follows: Section II introduces definitions and theoretical foundations about GNSS navigation performance criteria as well as some parameters of integrity. Section III presents the traditional integrity monitoring approaches in the aviation context. Then the next section analyzes the limitation of the classic integrity monitoring approaches in the urban context by summarizing the complexity of the GNSS signal reception in the urban environment. Finally, section V gives a structured overview of the existing integrity monitoring approaches for the urban GNSS receivers and the last section draws the conclusion and proposes some perspectives for the future work. The paper also has an appendix section which presents GNSS positioning principles.

II. DEFINITIONS AND THEORETICAL FOUNDATIONS OF GNSS INTEGRITY

A. GNSS Navigation Performance Criteria

Let us define here the concept of integrity in the context of GNSS performance. Generally, when talking about the performance of GNSS, we will necessarily mention the four criteria: accuracy, integrity, continuity and availability which are defined as follows:

Accuracy of an estimated or measured position and velocity of a vehicle at a given time is the degree of conformance of these position and velocity with the true ones of the vehicle [18]. Accuracy is related to the statistical features of merit of position or velocity error. So accuracy metrics are often built from the statistical distribution of the errors. Thus, the accuracy specifications are often given at a certain percentile of the Cumulative Distribution Function (CDF) (e.g., 95th percentile). Generally, for ITS applications, as specified by the European Committee for Standardization (CEN) and European Committee for Electrotechnical Standardization (CENELEC), accuracy is represented with a set of three statistical value given by the 50th, 75th and 95th percentiles of the CDF of the position error [19].

Integrity is conventionally defined as the measure of trust that can be placed in the correctness of the information supplied by a navigation system. This concept is originally introduced in the aviation context in the last decades in order to measure the influence of the navigation performance on the safety. Since the concept of integrity was intended for SoL applications, it also includes the ability of the system to provide timely warnings to users when some system anomaly results in unacceptable navigation accuracy [18] [20]. In summary, it is an indicator of veracity and trustworthiness that can be placed in the information supplied by the navigation system.

Recently, integrity monitoring has been more and more introduced into road transport especially for the liability critical applications. Under this context, the definition of integrity is re-adapted, for instance, by the SaPPART (Satellite Positioning Performance Assessment for Road Transport) project [7] as following:

Integrity is a general performance feature referring to the trust a user can have in the delivered value of a given position or velocity quantity (e.g., horizontal position). This feature applies to 2 additional quantities associated to the value delivered at each epoch of pseudo-range measurement: the Protection Level (PL) and the associated Integrity Risk (IR).

The definitions of these parameters will be detailed hereafter in the following section.

Continuity is the probability that the specified system performance (accuracy and integrity) will be maintained for the duration of a phase of operation, presuming that the system was available at the beginning of that phase of operation. Hence it expresses reliable operation (no failure) of the system during the specific time interval given that the system was operating at the start of the operation.

Under the context of mass-market applications, unlike integrity, which is important for liability critical applications, the concept of continuity is essential especially for the Location-Based Service (LBS) [21]. These kinds of services refer to the software applications for mobile devices that require knowledge about where the mobile device is located. For instance, based on the knowledge of users’ positions, LBS can provide the nearest points of interest (bank, restaurant etc.) For these applications, the continuity of the user positions is more important than other criteria since ideally the service should be available anywhere at anytime. Besides, continuity is an important criteria for railway signaling and train control in order to guarantee the safety of the operations [22] [23]. On the contrary, continuity is not a relevant feature for ITS domain and is therefore replaced by another called timing performance composed of time-stamp accuracy and output latency, update rate, jitter and Time to First Fix (TTFF) [24].

Availability is officially defined by ICAO as the percentage of time that the services of the system are usable by the navigator, which is an indication of the ability of the system to provide reliable information within the specified coverage area. But for the road GNSS applications, this feature can be defined in many different ways according to application needs. For example, for certain applications, availability can be the percentage of the measurement epochs where the considered output is delivered with the required performance or simply where the considered output is delivered by the terminal, whatever its quality.

In fact, the criteria mentioned above come from the Required Navigation Performance (RNP) concept. These criteria are related to each other as shown in Fig.2. We can see that accuracy is the base and the starting point of the performance pyramid which is specified at a certain
Fig. 2. Navigation Performance Pyramid: Accuracy, Integrity, Continuity and Availability

confidence level (e.g., %95). Then, there is a direct link between the definition of integrity and accuracy because the condition when a system should not be used for navigation is a lack of confidence in accuracy. And the continuity is the probability that accuracy and integrity will be maintained over a certain period. So continuity builds upon both accuracy and integrity. Finally, the definition of availability contains the notion of reliable information. To be reliable, the information must meet the accuracy, integrity and continuity specifications. Thus, availability is based on the assumption of certain levels of accuracy, integrity and continuity.

Besides these classic performance criteria from the aeronautical RNP, in the context of urban GNSS applications, other important performance features of GNSS can also include: robustness to spoofing and jamming, indoor penetration etc [25]. This article will only focus on the integrity aspect, which will be detailed in the following text.

B. Basic Definitions of Integrity

Integrity is a measure of trust that can be placed in the correctness of the information supplied by a navigation system and it includes the ability of the system to provide timely warnings to users when the system should not be used for navigation [18] [20]. This definition can be clarified thanks to four main parameters: Alert Limit (AL), Integrity Risk, Time to Alert (TTA) and Protection Level (PL).

Alert Limit represents the largest position error allowable for safe operation, more precisely:

- Horizontal Alert Limit (HAL) is the radius of a circle in the horizontal plane (the local plane tangent to the WGS-84 ellipsoid), with its center being at the true position, which describes the region that is required to contain the indicated horizontal position with the required probability for a particular navigation mode.

- Vertical Alert Limit (VAL) is half the length of a segment on the vertical axis (perpendicular to the horizontal plane of WGS-84 ellipsoid), with its center being at the true position, that describes the region that is required to contain the indicated vertical position with the required probability for a particular navigation mode.

In the urban context, generally we are only interested in the horizontal dimension.

Time to Alert (TTA) is the maximum allowable elapsed time from the onset of a positioning failure until the equipment announces the alert. So with this parameter, the integrity risk can be specified in a time interval.

Integrity Risk is the probability of providing a signal that is out of tolerance without warning the user in a given period of time [18]. It defines the maximum probability with which a receiver is allowed to provide position failures not detected by the integrity monitoring system [26].

Protection Level is a parameter of the integrity concept which will be well highlighted in urban vehicular contexts. It is formally defined as:

- The PL is a statistical error bound computed so as to guarantee that the probability of the absolute position error exceeding the said number is smaller than or equal to the target integrity risk [18].

Similar to the definition of AL, PL is also typically defined separately for the horizontal plane (Horizontal Protection Level, HPL) and the vertical direction (Vertical Protection Level, VPL). And here we only focus on the horizontal dimension which is defined as:

- The HPL is the radius of a circle in the horizontal plane (the local plane tangent to the WGS-84 ellipsoid), with its center being at the true position, that describes the region assured to contain the indicated horizontal position. It is a horizontal region where the missed detection and false alert requirements are met for the chosen set of satellites when autonomous fault detection is used [1].

Generally, the AL is specified by applications and the PL is calculated by users. Since the position error is not observable, the decision of alert is done by comparing the AL specified and the PL calculated, more precisely:

- If PL > AL, the alert triggers;
- If PL < AL, the alert does not trigger.

C. Integrity Events

Integrity Failure is an integrity event that lasts for longer than the TTA and with no alarm raised within the TTA.

Misleading Information (MI) is an integrity event occurring when, being the system declared available, the
position error exceeds the protection level but not the alert limit.

**Hazardously Misleading Information (HMI)** is an integrity event occurring when, being the system declared available, the position error exceeds the alert limit. Typically, in operating an aircraft, the risk for HMI due to navigation system is budgeted at the level of $10^{-7}$ to $10^{-9}$, which is extremely tight in order to guarantee the safety of operations. But the specification of HMI probability for urban applications has not been set yet.

Fig. 3 gives us a clearer illustration of the relationship between integrity parameters and each integrity event. Besides, the Stanford diagram (or Stanford plot) is generally used as a handy tool to explain and illustrate most of these integrity events and their relations (as well as to assess positioning systems performance), which is shown in Fig. 4. But the disadvantage of this tool is that the true position error should be known, which is difficult in practice.

### III. CLASSIC INTEGRITY CONCEPTS IN THE AVIATION DOMAIN

#### A. Traditional Approaches for Integrity Control

Since the early 90s, as the aviation domain depends more and more on GNSS, the integrity concept was introduced as a crucial measure of confidence of the information supplied by the navigation system.

Generally, the GNSS integrity information can be obtained from different ways. The most basic is the GNSS navigation messages, which indicate the anomalies related to the system and satellite operations such as satellite clock errors. But this kind of integrity information cannot be used for the real-time applications since the ground control segment can take a few hours to identify and broadcast the satellite service failure [28]. Thus, additional sources have to be used to deal with the integrity control.

In the aviation field, the information of integrity is provided by the three normalized augmentations known under the terms ABAS (Airborne Based Augmentation System), GBAS (Ground Based Augmentation System) and SBAS (Satellite Based Augmentation System) [29]. Among the three architectures, the GBAS and SBAS have to rely on some external aiding devices, such as sensor stations.

GBAS relies on a network of ground station references. It can provide estimates of common-mode errors and detect GNSS faults and anomalies. And integrity information can be obtained by comparing the true position of the ground reference and the estimated position obtained from the GNSS. This kind of augmentation system is mainly used at a local level, typically in airports.

SBAS transmits differential corrections and integrity messages for navigation satellites that are within sight of a network of stations, typically deployed for an entire continent. All the SBAS satellites signals covering a given zone are monitored in order to update the error model at the raw range measurement level [28] [29].

ABAS provides integrity monitoring for the position solution using redundant information within the GNSS constellation. ABAS is usually referred to as Receiver Autonomous Integrity Monitoring (RAIM) when GNSS information (range measurements) is exclusively used and as Aircraft Autonomous Integrity Monitoring (AAIM) when information from additional on-board sensors (e.g. barometric altimeter, clock and Inertial Navigation System, INS) are also used [29].

Receiver Autonomous Integrity Monitoring (RAIM) is a
technique based on the consistency check of redundancy of range measurements which is initially investigated in the aviation field since the late 1980s [30] [31] [32] [33] [34] [35] [36] [37]. Many different RAIM schemes have been proposed over the past few years, most of which are snapshot algorithms, such as the range comparison RAIM, the parity method RAIM, Least-Squares-Residuals (LS) RAIM and the Separation Solution (SS) RAIM [38] [39] [40] [41] [42]. Except for these snapshot algorithms, several Kalman filter based RAIM/FDE schemes are proposed [43] [44], which will be discussed later in the V-B section.

Fig. 5 gives us an overview of the flowchart of classic RAIM algorithms. Generally speaking, these classic RAIM has following important features:

- The classic RAIM technique mainly aims at large errors caused by satellite service failure. Since the probability of occurrence of two or more satellite service failures is negligible, classic RAIM detect only one fault each time.

- RAIM may include the function of fault detection and fault exclusion (FDE). It requires at least five (six) pseudo-range measurements to realize the fault detection (fault exclusion).

- The RAIM availability check module does not need to employ current measurements, that is to say, a HPL can be predicted with the satellite/user geometry, the nominal error characteristic (error variance) as well as the integrity probability requirements. Only if $HPL < HAL$, can RAIM continue to enter into the FD module. In addition, after the FDE, actual uncertainty level can be calculated with the help of the geometry, the measurements (i.e. the residuals) as well as the error variance. In this case, this level is called the Horizontal Uncertainty Level (HUL).

- Classic RAIM techniques used in the aviation field model the nominal pseudorange error as Gaussian distribution with zero-mean and a known variance.

Till now, no RAIM implementation exists in aviation domain for any flight operations requiring integrity in vertical planes, which has more stringent requirement such as precision approaches. This gave the motivation of developing the second generation RAIM. Under this context, Advanced RAIM (ARAIM) and Relative RAIM (R-RAIM) are proposed as two parallel candidates for future generation integrity monitoring architectures to support precision approach operations with both lateral and vertical guidance [45] [46]. In fact, as reported in GEAS [46] with updated results, ARAIM with MHSS method was adopted as the major architecture and the position domain RRAIM was only be used when ARAIM was not available. Compared to the classic LS RAIM, ARAIM can provide following improvements:

- ARAIM is designed to account for the multi-faults and is possible to exploit the multi-constellation GNSS with dual-frequency observation to remove the first order ionospheric delay [40] [41].

- ARAIM allows explicit computation of the integrity risk allocation while the classic RAIM is mainly based on probability of false alarm and missed detection [47].

Besides, other RAIM architectures exist such as Carrier-Phase-based RAIM (CRAIM) and Extended RAIM (ERAIM). CRAIM is mainly based on the GNSS carrier phase measurements [48] [49]. Although the carrier phase measurement is much more precise compared to the code measurement since a lower level of noise is involved, the ambiguities exists which is difficult to be successfully fixed especially in harsh environments. This is also the reason for which the carrier phase measurements generally cannot be used as an absolute measurement to estimate PVT solutions while they are preferred to be used to estimate the users’ dynamic in GNSS-based relative navigation and positioning. ERAIM uses the hybridization of GNSS and INS measurements to realize the integrity monitoring [50] [51] [52] [53], which is generally based on the EKF filter. Table I makes a summary about the classification of the RAIM techniques.

B. Classic PL Computation

For users, it is of great interest to have an estimation of the confidence in the position information provided by the GNSS receiver. The $PL$ is a statistical tool to bound the position error.

First, we introduce the concept of expected position confidence, which is a statistical measure related to the errors between estimated positions and the true (unknown) position of the receiver. It can be proved from the navigation equation that this factor depends on two important parameters [28]:

- the quality of range measurements performed by the GNSS receiver, which is usually expressed by User Equivalent Range Error (UERE). This parameter can reflect the error budget for a given satellite and generally is based on the computation of the following contributions: orbit determination and synchronization
equivalent error, troposphere residual error, ionosphere residual error, multipath residual error and receiver noise residual error [29]. In open sky, we consider the UERE as a random variable with a zero-mean Gaussian distribution, whose variance is the sum of the variances of all the error components (detailed in Appendix A):

- the user-satellite geometry, expressed by the Dilation Of Precision (DOP) such as the most general one: the Geometric Dilution of Precision (GDOP). GDOP is a geometry factor depending on the reciprocal positions of the user and the satellites in view which figures out the amplification of the standard deviation of the pseudorange measurement errors onto the solution.

The accuracy of the position/time solution provided by the GNSS can be expressed as the product of a pseudorange error factor and a geometry factor. That is to say, the errors in the GNSS solution can be loosely expressed by the following formula [28]:

\[
\text{error in the GNSS solution} = (\text{geometry factor}) \times (\text{pseudorange error factor})
\]  

According to [28], the GDOP is defined as:

\[
GDOP = \frac{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2}}{\sigma_{UERE}}
\]  

where:
\[
\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_t^2
\] are the variances of the position and time solution error respectively.
If we rearrange Equation (2), we can obtain:

\[
\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2} = \sigma_{UERE} \times GDOP
\]  

which has exactly the form given in Equation (1). Thus, the term on the left side of Equation (3) can be regarded as a position confidence, which is redefined as follows:

\[
\sigma_{pos} = \sigma_{UERE} \times GDOP
\]  

Generally, this position confidence is expressed separately in the vertical and horizontal directions since the required accuracy in these two directions is usually different for most applications, which is described as [64]:

- Vertical position confidence (\(\sigma_V\));
- Horizontal position confidence (\(\sigma_H\)).

The position confidence is the basis to calculate the PL, because the PL is a function of the satellite-user geometry and the expected pseudorange error while combining the required integrity risk probability.

For example, the SBAS PL equations are directly specified by [1] as:

\[
XPL = k_X \cdot \sigma_X
\]  

where:
\(
X
\) represents the H or the V dimension;
\(k_X\) is an inflation factor determined from the missed detection probability.

To concretely realize the computation of the PL, we need to know the distribution of the residual position or range error. In open sky areas, the probability distribution function (PDF) of the position error is often supposed to be Gaussian with zero-mean and a known variance in nominal cases. Thus, under this hypothesis, the PL can be computed directly from the position error distribution as well as the integrity risk. A simple example is illustrated in Fig. 6, where \(f_\varepsilon(e)\) represents the PDF of the position error, the surface of the shadow part in red is the integrity risk, i.e., \(P_{IR} = P(\varepsilon > e_{max})\) and so in this case the maximum tolerable position error \(e_{max}\) can

---

**TABLE I**  
**CLASSIFICATION OF RAIM TECHNIQUES**

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Measurement</th>
<th>Algorithms</th>
<th>FDE Capability</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic RAIM</td>
<td>GNSS Code</td>
<td>LS / WLS Residual-based method or parity-based method</td>
<td>FDE for single fault</td>
<td>[30] [31] [32]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[33] [54] [58] [39]</td>
</tr>
<tr>
<td>ARAIM</td>
<td>GNSS Code</td>
<td>Solution Separation (SS) method (Single alternative hypothesis or Multiple hypothesis or Classic Residual-based method)</td>
<td>FDE possible for multiple faults</td>
<td>[47] [55] [40]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[41] [42] [56] [45] [46] [57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[58] [59]</td>
</tr>
<tr>
<td>RRAIM (Range RRAIM or Position RRAIM)</td>
<td>GNSS Code and Time-Differenced Carrier Phase (TDCP)</td>
<td>MHSS or Classic Single alternative hypothesis method</td>
<td>FDE possible for multiple faults</td>
<td>[60] [61] [57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[45] [46] [62]</td>
</tr>
<tr>
<td>CRAIM</td>
<td>GNSS Carrier Phase</td>
<td>EKF innovation-based method &amp; ambiguity resolution methods (e.g., LAMBDA [63])</td>
<td>Only FD is possible</td>
<td>[48] [49]</td>
</tr>
<tr>
<td>ERAIM</td>
<td>GNSS Code and INS</td>
<td>EKF innovation-based MHSS or EKF innovation-based parity method</td>
<td>FDE possible for multiple faults</td>
<td>[50] [51] [52] [53]</td>
</tr>
</tbody>
</table>
be identified as PL.

For SBAS and GBAS, the $k$ factors are fixed for different phases of flight by aviation requirements in [1] [65]. For instance, for $HPL_{SBAS}$ computation, $k_H$ is equal to 6.18 for en route through Lateral Navigation (LNAV); for $VPL_{SBAS}$ computation, $k_V$ is equal to 5.33, corresponding to a $10^{-7}$ integrity risk requirement. Therefore, the airborne equipment can compute the PL with the fixed $k$ factors as well as the error model transmitted by the SBAS or GBAS reference stations.

For different RAIM algorithms, several methods are adapted for the computation of the PL, as detailed in [66] [37] [36] [67].

Hereafter, we will only make details about the classic slope-based PL computation proposed by [37], which is also detailed in [29].

In the RAIM algorithm, two possible pseudorange error scenarios are supposed: fault-free and faulty cases. In the fault-free case, the pseudorange measurements are affected only by nominal errors which are modeled as zero-mean independent Gaussian distributions with a known diagonal covariance matrix $\Sigma$. But in the faulty case, a bias is added in one of the range measurements. Since the position errors are not directly observable, RAIM uses a test statistic to realize the detection of the position error. For example, the most classic one, Least-Squares-Residuals (LSR) RAIM, uses the Normalized Sum of Squared Error (NSSE) $t$ as the test statistic, which is defined as:

$$t = \frac{||r||^2}{\sigma^2}$$

where,
- $r$ represents the pseudorange measurement residual vector which is a discrepancy vector between the current pseudorange measurements and the predicted pseudorange measurements;
- $\sigma$ represents the standard deviation of the pseudorange measurement errors.

As a result, $t$ follows a $\chi^2$ distribution in the fault-free case and a noncentral $\chi^2$ distribution in the faulty case, that is to say:

$$t \sim \begin{cases} \chi^2_k & \text{if } E \sim N(0, \Sigma) \\ \chi^2_{k,\lambda} & \text{if } E \sim N(b, \Sigma) \end{cases}$$

where,
- $E$ is the pseudorange error vector;
- $b$ is the bias vector in the faulty case, and generally, one single bias is supposed;
- $k$ is the number of degrees of freedom for the $\chi^2$ distribution in the two cases, which is the number of redundant pseudorange measurements;
- $\lambda$ is the non-centrality parameter of the $\chi^2$ distribution.

That is to say:

- In the nominal case,

$$\exists \xi_i, t = \sum_{i=1}^{k} \xi_i^2 \ iid, \ \xi_i \sim N(0, 1) \quad (8)$$

Then the probability of false alarm is used to determine the normalized detection threshold $T$ such as:

$$P_{FA} = \int_{T}^{\infty} f_{\chi^2_k}(x) \, dx$$

- In the faulty case,

$$\exists \xi_i, t = \sum_{i=1}^{k} \xi_i^2 \ iid, \ \xi_i \sim N(\mu_i, 1) \quad (10)$$

So $\lambda$ can be expressed by definition as:

$$\lambda = \sum_{i=1}^{k} \mu_i^2$$

With the specified $P_{MD}$ and threshold $T$ obtained previously, we can calculate the the minimum detectable noncentrality parameter $\lambda_{det}$ such that:

$$P_{MD} = \int_{0}^{T} f_{\chi^2_{k,\lambda_{det}}}(x) \, dx \quad (12)$$

The obtained $\lambda_{det}$ is independent of any pseudorange.

Then, a parameter called slope is introduced as a measure of the coupling between the effect of a pseudorange bias (the induced position error) in the observable parameter (test statistic) [34] [68]. Slope can be expressed as:

$$\text{slope}_i = \frac{(H_{N,i}^+)^2 + (H_{E,i}^+)^2}{S_{ii}}$$

where,
- $H^+ = (H^T H)^{-1} H^T$ represents the pseudo inverse of the matrix $H$ in the local navigation frame (east, north, up). $H$ is the observation matrix in the navigation equation;
- $S = I - HH^+$. 

![Fig. 6. A simple illustration of the relationship between the PL and the integrity risk in the nominal case; the error distribution is a centered Gaussian and the zone in red represents the target integrity risk specified according applications. in this case, the maximal tolerable PE, $\hat{e}_{max}$, is our PL.](image-url)
Each satellite has its own slope. The satellite with the highest slope is the most difficult to detect. It is also the one that produces the largest position error (which we want to protect) for a given test statistic [34]. And a pseudorange bias leading to a given non-centrality parameter $\lambda$ will have the highest impact on the positioning error when it appears in the satellite with the highest slope [17].

Finally, the HPL can be obtained [37]:

$$HPL = \sigma \cdot \text{slope}_{\text{max}} \cdot \sqrt{\lambda_{\text{det}}}$$  \hspace{1cm} (14)

Several methods of deriving a HPL exist according to different assumptions. But one important metric to evaluate a HPL is that it can properly bound the errors with a reasonable size which depends strongly on the targeted application.

IV. COMPLEXITY AND LIMITATIONS IN URBAN ENVIRONMENT

A. Complexity of GNSS Signal Reception in Urban Environment

The urban environment presents several challenges to the GNSS signal reception, which could lead to severe degradation of positioning accuracy if no special measures are taken. And these complexities can be sorted into two major issues.

First of all, since obstacles in the urban environment can block GNSS Line-Of-Sight (LOS) signals, the number of satellites in view will be effectively reduced. Yet this situation can be improved by using a multi-constellation receiver in order to obtain sufficient direct-LOS signals for the computation of a position solution [69] [70]. This effect influences also the geometrical distribution of the satellites around the users, i.e., Dilution of Precision (DOP).

Secondly, due to flat surface reflectors presenting in the urban environment, the problems of multipath interference and NLOS reception arise [9], which are illustrated in Fig. 7. In fact, the multipath interference and NLOS reception should be considered as two different phenomena as they can produce different ranging errors. The detailed explanations about these two phenomena are in [71].

The consequences of the two major problems mentioned above are:

1) Distort the correlation function of receiver

In GNSS received signal processing, correlation is an essential step which helps receivers to estimate TOA $\Delta t$ of the GNSS signals, which directly links to pseudorange measurements. In fact, by correlating the received satellite Pseudo Random Noise (PRN) code with the replica generated by the local receiver, the TOA $\Delta t$ can be determined from the maximum of the correlation function as shown in Fig. 8 [72]. The reception of a multipath-contaminated signal will effectively distort the correlation function so that the code and carrier phase tracking accuracy will be degraded. This effect can lead to large range errors as well as inaccurate position solutions. Fig. 9 gives us an illustration of the effect of multipath interference on the correlation function [73]. The resulting code tracking error depends on the receiver design as well as the direct and reflected signal strengths, path delay and phase difference, and this error can be up to half a code chip [74] [75].

2) Increase or decrease the carrier-power-to-noise-density ratio ($C/N_0$) of the received signals

The carrier-power-to-noise-density ratio, i.e., $C/N_0$, represents the ratio of signal power and noise power per unit of bandwidth. In urban environments, constructive multipath interference leads to an increase in $C/N_0$, while destructive multipath interference leads to a decrease. The level of $C/N_0$ will mainly influence the signal tracking performance of GNSS receivers. For instance, the noise of the receivers’ tracking loop is directly linked with $C/N_0$ and the linear domain
of the discriminator output is also strongly affected by the level of $C/N_0$, which will further influence the tracking error [28]. Since the phase of a reflected signal with respect to its directly received counterpart depends on the wavelength, multipath interference may be constructive on one frequency and destructive on another frequency. As a result, these characteristics contribute to new multipath detection technique by comparing the difference in measured $C/N_0$ between two frequencies with what is expected for that signal at the elevation angle [9].

3) **Change the polarization of the signal**

GNSS signals directly received from satellites have Right-Handed Circular Polarization (RHCP). But after one reflection, the polarization becomes Left-Handed Circular Polarization (LHCP). Thus, most reflected signals have LHCP or mixed polarization. Consequently, multipath mitigation techniques can be developed at antenna design level by differentiating the sensitivity of antenna for RHCP and LHCP [9] [13] [14] [76].

4) **Inconsistent GNSS pseudorange measurements**

Because of the stringent environment for the GNSS signal reception, it is possible that the pseudorange measurements provided by one or more GNSS satellites are not consistent with other ones. Hence, it is necessary to implement algorithms to ensure that the pseudorange measurements are all consistent. [77] has implemented the random sample consensus (RANSAC) algorithm, developed for computer vision tasks, in the GNSS context. This algorithm is based on consistency checking and it is capable of detecting multiple fault unlike the RAIM technique, which is compatible to a degraded scenario such as urban environment.

![Correlation function](image)

**Fig. 9. Effect of constructive and destructive multipath interference on the correlation function [73]**

**B. Limitations of the Classic Integrity Concept in Urban Context**

As briefly mentioned previously, the classic integrity concept can not be transported straightforwardly into the urban vehicular context since the limitations are due to the stringent environment. By combining the complexity of GNSS signal reception analyzed in the previous parts, we can see the following constraints.

First of all, integrity control techniques in the aviation field such as RAIM, suppose that the distribution of the range and position domain error is Gaussian with zero-mean and a known variance in the nominal case, while just a single bias is added in the faulty case [34] [35]. The effects of multipath, limited satellite visibility, NLOS due to obstacles are not taken into consideration, which is also the case for the EGNOS [101]. These augmentation systems such as EGNOS can help the low cost commercial receivers to get a better accuracy in open sky conditions but, in a severe environment, their performances degrade, which is proved by experimental data in [101] [102]. Thus, the error models have to be characterized in order to make them more adapted to the urban GNSS applications. We will further address the error models in Section V-A.

Secondly, the satellite visibility is badly degraded in urban canyons [103]. Thus, the availability of traditional augmentation systems such as SBAS will be affected due to bad satellite visibility [102]. In addition, SBAS also adopts the Gaussian model with zero-mean and a known variance. As a result, either the SBAS satellites are not visible or the obtained PLs are too large, which is not usable in the corresponding urban applications.

Thirdly, as already mentioned in section III-A, the RAIM algorithm supposes a scenario of high redundancy and that no more than one failure is detected at a time, which is not true in the urban environment, because the major error that threaten the urban integrity is NLOS. And the errors provoked by NLOS can be very large and frequent. We cannot guarantee the high redundancy either in the urban context due to the poor satellite visibility.

Finally, the typical requirements of integrity risk in aviation are often too conservative for the vehicular applications [104].

These limitations have been proved by several research works. For instance, with real GNSS data, [17] shows that, in the dual-constellation case and a HAL of 50 m, the percentage of epochs in which a RAIM configured with $P_{MD} = 5 \times 10^{-5}$ and $P_{FA} = 5 \times 10^{-3}$ is available decreases from almost 100% in the rural environment to approximately 55% in the urban one. In the GPS case, it decreases from 50% to around 7%. Lower PMD or PFA would still decrease the availability rate. Similar conclusions are obtained with simulations by [105]. In addition, [5] also proves that the HPL calculated by the classic measurement rejection approach is too big for the urban applications (e.g., for a data set of urban Madrid, only 10% of the measurement epochs have a protection level smaller than 100 meters).
<table>
<thead>
<tr>
<th>Approach Classifications</th>
<th>Techniques</th>
<th>Features of Techniques</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenna design</td>
<td>1. Dual polarization antenna</td>
<td>This method cannot detect all the NLOS, especially those reflected by even times (e.g., twice or 4 times).</td>
<td>[9] [13] [14]</td>
</tr>
<tr>
<td></td>
<td>2. Choke Rings</td>
<td>The volume of choke-ring antenna system is too large for most dynamic positioning applications</td>
<td>[76] [78] [79]</td>
</tr>
<tr>
<td></td>
<td>3. Controlled Reception Pattern Antenna (CRPA)</td>
<td>Better performance for high elevation signals, large volume and expensive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Angle of Arrival (AOA) measurement</td>
<td>suitable for NLOS and strong multipath; interference; expensive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Multiple Antennas</td>
<td>suitable for large vehicles (e.g., ships, trains or large aircraft)</td>
<td></td>
</tr>
<tr>
<td>Receiver Design</td>
<td>1. Code Discriminator Design</td>
<td>expensive for manufacturing; huge power consumption</td>
<td>[15] [73]</td>
</tr>
<tr>
<td></td>
<td>2. Early-Late Correlator Comparisons</td>
<td>more effective for dynamic applications</td>
<td>[80] [81]</td>
</tr>
<tr>
<td></td>
<td>3. Vector Tracking</td>
<td>Similar mechanism with carrier smoothing</td>
<td></td>
</tr>
<tr>
<td>Weighting model</td>
<td>1. C/N₀-based Weighting model</td>
<td>Can improve the positioning accuracy; easy to implement</td>
<td>[82] [83] [84] [85]</td>
</tr>
<tr>
<td></td>
<td>2. Satellite elevation-based weighting model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Danish Reweighting Method</td>
<td>it is an empirical procedure with no valid statistical explanation but perform well in practice</td>
<td></td>
</tr>
<tr>
<td>Signal Processing</td>
<td>1. Carrier Smoothing</td>
<td>more adapted to dynamic applications; only multipath interference mitigation, not NLOS mitigation</td>
<td>[86] [87] [88]</td>
</tr>
<tr>
<td></td>
<td>2. Doppler Domain Multipath Mitigation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Multi-frequency C/N₀</td>
<td>reliable for static applications; easy to implement</td>
<td></td>
</tr>
<tr>
<td>Image Processing</td>
<td>1. Fish-eye Camera</td>
<td>Discrimination between LOS/NLOS</td>
<td>[89] [90] [91]</td>
</tr>
<tr>
<td>Consistency Checking</td>
<td>1. RAIM</td>
<td>Measurement redundancy required</td>
<td>[85] [84] [74]</td>
</tr>
<tr>
<td></td>
<td>2. Subset Testing</td>
<td>Performance degraded if a large proportion of signals are NLOS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Forward-Back Testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapping-Aided</td>
<td>1. 2D map-matching (e.g., Urban Trench Model (UTM))</td>
<td>Mapping error exist without knowing true position; computationally intensive; huge work to establish and load 3D city model.</td>
<td>[92] [93] [94]</td>
</tr>
<tr>
<td></td>
<td>2. 3D environment model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Approaches</td>
<td>1. Bayesian approaches</td>
<td>latter state is easy to be contaminated by the former one because of the sequential procedures; high processing load is possible</td>
<td>[95] [96] [97]</td>
</tr>
<tr>
<td></td>
<td>2. Particle filtering</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. KF-based innovation filtering</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Maximum likelihood Estimation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V. EXISTING RESULTS FOR INTEGRITY MONITORING IN URBAN ENVIRONMENT

A. Error Models for Integrity Monitoring in Urban Context

Properly characterizing the GNSS position errors is essential to realize integrity monitoring in urban environment since certain error models established in the aviation field are not valid anymore. Ideally, the true error distribution can be obtained by collecting all the possible error sources. Yet, this is too difficult to realize especially in the urban environment. This is because, firstly, it is not realistic to collect all error sources, which is complicated and varying all the time in urban contexts such as multipath. Secondly, despite the fact that, in the aviation field, the error sources are simply supposed to be independent, the dependency of error sources cannot be negligible for the urban applications [106]. This makes it more difficult to model the error distribution in the urban environment. Finally, the error sources are rarely stationary [106], which once again adds difficulties to the error modeling in urban environment.

Despite the huge difficulties, some research already exists which has addressed the error models in the urban environments. We will class them into two categories: pseudorange domain error models and position domain error models.

1) Pseudorange Domain Error Models: Since the position errors are generally not observable by the receivers, most research starts and focuses on the pseudorange error models in urban environments.

A class of methods of overbounding exists which can deal with the bias due to multipath or NLOS reception. [107] gives a structured overview of the overbounding solutions in the use of SBAS integrity, such as the Probability Density Function (PDF) overbounding and the Cumulative Distribution Function (CDF) overbounding. [108] has analyzed the PDF overbounding and CDF overbounding performance with numerical simulations. It is proven that the CDF overbounding cannot exceed everywhere the PDF of the actual error distribution. So, compared to PDF overbounding, CDF overbounding is less restrictive. That is to say, CDF overbounding can result in more conservative standard deviations for the overbounding distribution.

Overbounding methods have been well addressed in the aviation field [107] [108] [109], yet there has not been much research into overbounding employed specifically for the urban integrity control applications. This may be a promising method which can be continued for urban error modeling.

Except for the overbounding error models, some other pseudorange error models exist in the current literature.
[110] has proposed the model of noise variance jump or mean value jump, which depends on the case of LOS or NLOS reception as well as the dynamic of the vehicles.

[111] proposed the Gaussian Mixed Model (GMM) as the pseudorange error model. The essential of this model is to approximate the pseudorange error distribution in the degraded environment as the weighted sum of several different Gaussian distribution. It is proven that the GMM model can adapt better, especially in the case of NLOS reception, than some other distributions such as Normal distribution, Rayleigh distribution or Laplace distribution [112]. But the GMM has inevitably some potential limitations which are also demonstrated in [112]. For instance, in the case where a sudden change of the reception state occurs, the GMM is not accurate enough since there is a dependence on the past observations for this model. And the number of the sum is always an important parameter to determine, which depends on the reception scenario.

Considering the drawbacks of the GMM model, [112] [97] have proposed the Dirichlet Process Mixture (DPM) model. This model can adapt well to the change in the reception state. This is because the DPM model works in a sequential way with less dependency on the previous range measurements. The performance of the DPM model is shown with simulation and experimental data. But the main drawbacks of the DPM model are the complexity of its implementation and the high computational cost.

These last two pseudorange error models mentioned previously have not been analyzed and implemented in the urban integrity control, which can be an axis for further research.

2) Position Domain Error Models: The position errors are obviously not observable by the receivers. But modeling the position errors seems to be more straightforward than modeling the pseudorange errors since the integrity information is provided in the position domain.

In terms of overbounding in the position domain, [107] has described a model called the tail area overbounding. This kind of overbounding required that the tail area of the overbounding distribution contains more energy than the tail area of the actual distribution [107]. But, the tail area overbounding in the position domain cannot guarantee the overbounding condition in the range domain. What is more, [113] describes a position domain method which can improve the availability of Local Area Augmentation System (LAAS) by reducing the inflation factor for standard deviations of pseudorange correction errors.

Besides, [114] proposed some error models in the position domain for a specific receiver. Experimental data is used in order to get the empirical CDF of the Horizontal Position Error (HPE) and further to identify the distribution characterizations. The position error is obtained from the difference between the measured positions and a trajectory of reference. It is shown that the empirically based HPE distribution has a good fit with Rayleigh distribution in the open sky, while in urban environment it is fitted to the Pareto distribution.

The most obvious advantage of the error modeling in the position domain is the capability to get rid of the unobservable multiple fault conditions. That is to say, the error residual can fade due to the combination of several unhealthy range measurements [5]. However, the error models in the position domain are only valid for specific receivers because different algorithms or techniques are possibly implemented in different receivers for the position estimation. In turn, these position error modeling methodologies allow the classification or selection of the proper receiver for a specific application. But error models in the position domain cannot provide the possibility of detection of an isolated fault. As a result, it is better to combine the error models both in the position domain and the pseudorange domain, which could be a perspective for further research.

B. Integrity Monitoring Approaches in Urban Environment

In recent years, some different possibilities of solutions have been studied in terms of navigation integrity monitoring in urban environments. Often, the hybridization techniques between GNSS and other sensors are used. For example, [115] and [116] proposed integrity monitoring with map-matching techniques for land vehicle applications. [117] proposed a conception of integrity monitoring architecture using a fisheye camera which is not completely implemented yet.

Besides, a new concept which implements Vehicular Ad-Hoc Network (VANET) infrastructures is currently proposed by [118]. That is to say, different vehicles participating to a VANET can share and combine their observations of GNSS signals so that a collaborative spatial/temporal characterization and prediction of the local degradation of the GNSS signals can be implemented.

For the UAVs, multipath effects associate with their low-level flights but the integrity monitoring techniques for urban environments are still in its infancy. GNSS Aircraft-Based Integrity Augmentation (ABIA) technique [6] [119] [120] [121] is introduced as the main role to guarantee the integrity performance for the UAVs. The ABIA system delivers integrity caution (predictive) and warning (reactive) flags, as well as steering information to the electronic commands of the UAV flight control system. These features allow real-time avoidance of safety-critical flight conditions and fast recovery of the required navigation performance in case of GNSS data losses. In fact, this is similar to the concept of the ABAS, in which the integrity processing of GNSS data is performed onboard the UAV itself, and can be aided by additional sensors. And cooperation between different UAVs and exchange with UAV Traffic Management station are also
possible to make in order to realize integrity control.

However, in terms of autonomous integrity monitoring, that is to say, using standalone GNSS receivers, no methods or techniques exist which are well developed and ready to be implemented. Yet, this approach is more promising and attractive for users since it can reduce the complexity of the on-board equipment as well as the costs. Thus, we will concentrate on the integrity monitoring approaches without any other external equipment.

In current literature, two groups of theoretical approaches for integrity monitoring in urban environment exist: the measurement rejection approach (MRA) and the error characterization approach (ECA) [5].

1) MRA approach: The principle of the MRA is to reject faulty range measurements such as the classic concept mentioned previously. This approach not only works well in the open-sky environment but also can work in other environments if the assumption that only a single fault can occur at a time can be got rid of. Yet, removal of such an assumption is really a big challenge.

If multiple simultaneous faulty range measurements are considered, the threat exists that the error sizes can combine with the satellite geometry in such a way as to produce a large position error but very small residuals, thus passing unnoticed to a conventional large position error but very small residuals, thus passing with the satellite geometry in such a way as to produce a considered, the threat exists that the error sizes can combine assumption is really a big challenge.

The main idea of the ECA is to characterize the range measurement errors and be able to compute a PL that actually protects, without the need for identifying and removing degraded range measurements, even if they are contaminated with very large errors [5]. As a result, this approach can possibly lead to large protection levels which cannot suit the requirement of quite particular applications.

Isotropy-Based Protection Level (IBPL) is a patented algorithm as well as an ECA concept implementation which
can provide a PL autonomously.

The basis of the IBPL algorithm consists in using the vector of least square estimation residuals as a characterization of the position error. And the only assumption made is that the range measurement error vector has an isotropic distribution in the measurement space [122]. This means the error vector can point in any direction of the measurement space with the same probability. Then the following relationship is used:

\[ HPL = k \cdot \|r\| \cdot HDOP \]  
\[ \text{(15)} \]

where
- \( r \) is the least square residual vector;
- \( k \) is called Isotropic Confidence Ration (ICR) which depends on the target confidence level \( 1 - \alpha \) (\( \alpha \) is the integrity risk), the number of range measurements \( N \) and the number of unknown to estimate \( m \).

The detailed derivation of \( k \)'s expression as a function of \( \alpha, N \) and \( m \) is in [122]. Some tables of pre-calculated values of \( k \) with different \( \alpha, N \) and \( m \) are available. Other values not in the table can be obtained by the interpolation method.

The IBPL method can perfectly deal with the problem of single fault assumption in the classic RAIM algorithm. It has been proven to be relatively reliable and robust in certain degraded environment. But the disadvantage is that the calculated PL depends too closely on the number of range measurements. That is to say, if the visibility of the satellite is not good enough, the performance of the IBPL method will be badly degraded. Thus, for mono-constellation receivers in urban environment, IBPL algorithm is not very interesting due to bad satellite visibility.

Despite the robustness of the ECA approaches such as IBPL, their common problem is that, since neither range measurement is removed, the size of PL has the risk of being too large. So a trade-off should be made between the size of PL and the level of integrity risk.

In order to resolve the problems and the shortages of the existing IBPL method, GMV has lately expanded the IBPL method to support the Kalman filter, which is called the KIPL method. The KIPL can apply to GNSS-standalone or hybrid GNSS/INS navigation system. And the KIPL is able to provide tight integrity bounds in all kinds of environments for virtually any desired confidence level [123].

Except for the IBPL and KIPL methods, [124] has proposed a composite approach for HPL computation in urban environments. The principle of this method is to treat the biases and noises in a separate way. The PL can be formulated as a sum of noise component, \( PL_n \), plus a bias component, \( PL_b \):

\[ PL = PL_n + PL_b \]  
\[ \text{(16)} \]

And in [124], the bias and noise composites are separated by an autoregressive (AR) model. The noise component of HPL is calculated using the weighting model in [125]. For the additional term which represents the bias, the residuals obtained from the least-square PVT algorithm are used.

The analyses of the performances of HPL computation using this method in urban environments have not been made in detail. But in the open sky, it is proven that its main advantage is a clear final decrease in PL [126]. This is good news for urban integrity controlling with the ECA approach. Thus, further research about this method in urban contexts is needed.

VI. CONCLUSION

The concept of integrity has become a hot topic for urban GNSS users. Many research efforts have been devoted to addressing the problematic of position integrity in urban contexts. Yet there have been no methodologies or requirements of position integrity in urban environments which are mature enough to be implemented by urban GNSS receivers. Thus, the integrity for terrestrial applications is of great necessity.

This paper gives a global and structured review of the fundamentals and the state of the art of the integrity in urban contexts. Since the urban environment has its own particularity compared to the open sky environment, the integrity concept in urban context is more challenging. In order to guarantee the expected integrity level, the following two requirements should be met:

- The navigation solution should exist with an associated PL. That is to say, there ought to be enough measurements (code or carrier phase).
- The size of PL should be small enough in order to meet the specifications of the applications.

Consequently, the integrity monitoring method in urban contexts always confronts the compromise between the size of PL and the desired level of integrity. Since the integrity requirements are application dependent, specifications and algorithms for different urban applications are needed.

Considering all the complexities and the existing research for the urban integrity monitoring mentioned in this paper, we have the following perspectives:

- Error modeling in urban environments is still an important aspect to be conquered, especially the characterization of the local effects, which can contribute to the integrity monitoring;
- The proper way to remove the constraint assumption in the classic RAIM approach with a low computational cost should be fully addressed in the future. If this can be achieved, the implementation can be facilitated;
- Improvement of the existing urban integrity algorithms is necessary in terms of the trade-off between the size of PL and the criterion of the integrity. And other new algorithms can be developed based on the combination of current methods.
Fig. 10. GNSS-based Trilateration. A pseudorange is estimated by the user for each visible satellite. The intersection of the spheres (centered on satellites, with the corresponding measured pseudorange as radius) will be computed as the user position. At least three visible satellites are necessary to compute a 3D position and another one is needed to compute the clock offset.

- The specifications of the integrity for different urban applications should be developed and tested. For this, the methodology used in the aviation domain can be partly taken.

A number of issues remain in terms of integrity in urban environment. This promising topic is waiting for innovations.

ACKNOWLEDGMENT

The authors would like to thank the IFSTTAR (the French Institute of Science and Technology for Transport, Development and Network), CNES (Centre National d’Études Spatiales) and the region Hauts-de-France for their financial support.

APPENDIX A

GNSS POSITIONING PRINCIPLES

Global Navigation Satellite Systems (GNSS) refer to satellite navigation systems which provide continuous positioning over the globe [72]. Generally speaking, a GNSS is composed of three main segments: the space segment, the control segment and the user segment. GNSS receiver utilizes the concept of one-way Time of Arrival (TOA) ranging and trilateration mechanism to determine its position on the surface of the earth [28] as shown in Fig.10.

GNSS receivers can provide two main types of pseudorange measurements from satellites: code and carrier phase measurements. The code measurement typically includes high level of noise and the carrier phase measurement is more precise than the former one but the ambiguity problem exists, which results from cycle slips of carrier tracking. As a result, generally, the carrier phase measurement cannot be used as an absolute measurement to estimate PVT especially in harsh environment for practical applications since it is too difficult to successfully fix all the ambiguity. But they are preferred to be used to estimate users motions.

Thus, the carrier phase measurements are not as robust as the code pseudorange measurements. We will focus on the code measurement hereafter.

We can express the code pseudorange measurements for the satellite \(i\) as a function of the receiver true position and of the satellite positions as follows:

\[
P^i(k) = \sqrt{(x(k) - x^i(k))^2 + (y(k) - y^i(k))^2 + (z(k) - z^i(k))^2} + b_u(k) + e^i(k)
\]

where:

- \(x, y, z\) are the Cartesian coordinates of the receiver antenna at the time of signal reception expressed in an Earth-centered Earth-fixed (ECEF) reference frame;
- \(x^i, y^i, z^i\) are the Cartesian coordinates of the satellite antenna at the time of signal emission expressed in an ECEF reference frame;
- \(b_u(k)\) represents the receiver clock bias expressed in meter with \(b_u(k) = c\delta t_u(k)\);
- \(e^i(k)\) represents the sum of the code measurement errors due to ionospheric and tropospheric propagation delay residual, multipath, noise, satellite clock residuals with \(e^i(k) = P^i(k) + T^i(k) + D_{\text{mult}}(k) + n^i(k) - c\delta t_u(k)\).

Different error sources in \(e^i(k)\) have been studied and standardized error models exist. Some of them apply for both aviation domain and the urban context, such as ionospheric and tropospheric error. Some errors, however, such as multipath, cannot be applied directly in urban framework since the local effects are completely different from that of the aviation applications. Table III [28] shows us the GNSS error budget for the standard L1 C/A error model. What should be emphasized is that, the multipath error can even achieve several kilometers in challenging environments, which is more serious than open-sky cases.

With the raw code measurements, different estimators can be used to compute the positions of users such as Least Square (LS) Estimator and the Kalman Filter (KF) [28].

LS estimation algorithm and its variants such as Weighted Least Square (WLS) are basic methods to obtain a navigation solution. Its objective is to estimate user position in a iterative way by using the linearization of the range measurement model around successive estimate of the receiver position. The WLS estimator is the best linear unbiased estimator which reaches the Cramer-Rao lower bound [127].

The KF [128] (and its variants such as the extended Kalman filter [73]) is one of the most celebrated and popular data fusion algorithms. It is much used for the integration of GNSS and inertial sensors. KF is a statistical technique that combines knowledge of the statistical nature of system errors with knowledge of system dynamics. The state estimate utilizes a weighting function, called the Kalman gain, which is optimized to produce a minimum error variance. For this reason, the KF is called an optimal filter.
TABLE III
GNSS ERROR BUDGET (STANDARD ERROR MODEL FOR L1 C/A)

<table>
<thead>
<tr>
<th>error source</th>
<th>Ephemeris data</th>
<th>Satellite clock</th>
<th>Ionosphere</th>
<th>Troposphere</th>
<th>Multipath</th>
<th>Receiver noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 - \sigma$</td>
<td>1.1-2.1</td>
<td>1.1-2.1</td>
<td>4.0-7.0</td>
<td>0.2-0.7</td>
<td>0.2-1.4</td>
<td>0.1-0.5</td>
</tr>
</tbody>
</table>

GLOSSARY

AAIM  Aircraft Autonomous Integrity Monitoring

ABAS  Airborne Based Augmentation System

AL/HAL/VAL  Alert Limit/ Horizontal AL/ Vertical AL

AOA  Angle of Arrival

CDF  Cumulative Distribution Function

CEN  European Committee for Standardization

CENELEC  European Committee for Electrotechnical Standardization

CRPA  Controlled Reception Pattern Antenna

DOP/GDOP  Dilution of Precision/ Geometric DOP

DPM  Dirichlet Process Mixture (DPM)

ECR  Error Characterization Approach

EGNOS  European Geostationary Navigation Overlay Service

ETC  Electronic Toll Collection

FDE  Fault Detection and Exclusion

GBAS  Ground Based Augmentation System

GMM  Gaussian Mixed Model

GNSS  Global Navigation Satellite System

IBPL  Isotropy-Based Protection Level

ICAO  International Civil Aviation Organization

INS  Inertial Navigation System

IR  Integrity Risk

ITS  Intelligent Transport Systems

KIPL  Kalman Integrated Protection Level

LAAS  Local Area Augmentation System

LAMBDA  Least-Square AMBiguity

LBS  Location-Based Service

LNAV  Lateral Navigation

LOS/NLOS  Line-of-Sight/ Non-line-of-sight

LS/WLS  Least Square/ Weighted LS

MHSS  Multiple Hypothesis Solution Separation

MI/HMI  Misleading Information/ Hazardous MI

MRA  Measurement Rejection Approach

NSSE  Normalized Sum of Squared Error

PDF  Probability Distribution Function

PE/HPE  Position Error/ Horizontal PE

PL/HPL/VPL  Protection Level/ Horizontal PL/ Vertical PL

PRN  Pseudo Random Noise

PVT  Position, Velocity and/or Time

RAIM  Receiver Autonomous Integrity Monitoring

RANSAC  Random Sample Consensus

RHCP/LHCP  Right/Left-Handed Circular Polarization

SaPPART  Satellite Positioning Performance Assessment for Road Transport

SBAS  Satellite Based Augmentation System

SoL  Safety-of-Life

TDPC  Time-Differenced Carrier Phase

TTA  Time to Alert

TTFF  Time to First Fix

UERE  User Equivalent Range Error

VANET  Vehicular Adhoc Network

WAAS  Wide Area Augmentation System

REFERENCES


R. Sabatini, T. Moore, C. Hill, and S. Ramasamy, “Investigation of gns integrity augmentation synergies with unmanned aircraft systems sense-and-avoid.”


