We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Abstract

Safety-critical applications in transportation require GNSS-based positioning with high levels of continuity, accuracy and integrity. The system needs to detect and exclude faults and to raise an alarm in the event of unsafe positioning. This capability is referred to as integrity monitoring (IM). While IM was considered until recently only in aviation, it is currently a key performance parameter in land applications, such as Intelligent Transport Systems (ITS). In this chapter the IM concepts, models and methods developed so far are compared. In particular, Fault Detection and Exclusion (FDE) and bounding of positioning errors methods borrowed from aviation (i.e. Weighted RAIM and ARAIM) are discussed in detail, in view of their possible adoption for land applications. Their strengths and limitations, and the modifications needed for application in the different context are highlighted. A practical demonstration of IM in ITS is presented.

Keywords: ITS, C-ITS, integrity monitoring, RAIM, FDE, SBAS

1. Introduction

Integrity is a key performance parameter in positioning for ITS safety applications [1, 2]. To provide absolute positioning in safety-critical and mission-critical applications, satellite navigation shall maintain a very high level of service. Correctness — within tight bounds — of the position solution, shall be guaranteed to extremely high levels of probability. In aviation, the risk for so-called Hazardous Misleading Information (HMI) due to the navigation system is typically budgeted at the $10^{-7}$ to $10^{-9}$ level, and a similar level of safety is expected to be required in land applications in the era of fully automated transportation. In more formal terms, integrity is about the trust that a user (or the AI in charge of the vehicle) can have in the indicated position information. The trust is measured by the probability of HMI (or
integrity risk), which is the probability that the position error exceeds a certain tolerance, without being detected and an alert being raised in time. The given position information will then be misleading, as it is not correct within specified bounds, and the user is not aware of the potentially hazardous situation.

While IM was considered until recently only in aviation, it is currently a key performance parameter in safety-critical land applications. Even though integrity requirements in vehicular transport have not been defined yet, the demand for higher levels of automation in an increasing number of applications is pushing the relevant authorities to fill this gap.

1.1. Integrity monitoring in aviation

Today, integrity monitoring in aviation is implemented in two different ways, at system level or at user level. At system level, two types of external augmentation systems can be distinguished, Space-Based Augmentation Systems (SBAS), see [3, 4], and Ground-Based Augmentation Systems (GBAS), see [5]. Both are Differential GPS systems (DGPS). SBAS and GBAS develop corrections that improve the accuracy of the measurements and generate real-time error bounds. These bounds are called Protection Levels (PL) and must exceed the actual error under all conditions with very high probability [6]. SBAS and GBAS are both very powerful means of guaranteeing integrity, but they present the drawback of needing a very complex and costly infrastructure.

At user level the GNSS integrity can be monitored by exploiting the redundancy of the GNSS signals as collected at the receiver. This is done by performing calculations within the user equipment itself to check the measurements consistency. This method is known as Receiver Autonomous Integrity Monitoring (RAIM). RAIM is possible as long as a number of observations larger than the minimum necessary for a position fix are available. RAIM strictly relies on the strength of the satellite geometry. With the deployment of the new GNSS constellations many more satellite signals will soon be available: this will increase the redundancy of measurements and the RAIM power.

1.2. Integrity monitoring on land

Both SBAS/GBAS and RAIM methods can in principle be adopted for IM in land applications, since the fundamental positioning problem is the same. However, some important differences in the applications may make the task not straightforward. GNSS positioning in aviation is generally restricted to Single Point Positioning (SPP), based on code observations on the civil frequencies, L1 (E1 for Galileo) and soon L5 (E5). With SPP, accuracy of few meters is attainable. However, most current and future land applications (such as ITS) require lane-level accuracy, i.e., sub-meter accuracy [7]. As such level of accuracy is considered unattainable with SPS, ITS applications are foreseen to be relying on Satellite Based Augmentation Systems (SBAS), RTK or Precise Point Positioning (PPP) techniques [7].

The different positioning methods and the corresponding higher precisions involved bring with them a new set of specific vulnerabilities. For instance, anomalies that would create positioning errors of too small magnitude in an SPP context, and could therefore been
neglected, would now need to be taken into account. In case carrier phase observations are to be used, cycle-slip monitoring shall be included, as well as IM for ambiguity resolution. Another difference from the aviation case that shall be taken into account is the environment in which positioning is to take place. Land users are often located in urban environment, which is characterized by the presence of high-rise buildings: as a result, GNSS observations are highly more likely to be affected by multipath and Non-Line-of-Sight (NLOS). Furthermore, the urban environment brings extra vulnerabilities linked to the higher risk of interference.

As the SBAS integrity monitoring concept has not been defined yet for ITS applications, this chapter focus is on the RAIM concept. This is in fact the most versatile integrity monitoring approach, generally applicable to any estimation problem. The chapter is organized as follows: in Section 2 the integrity as a navigation performance parameter is introduced and the focus moves to the RAIM approach. The RAIM problem is defined and the most important performance parameters of RAIM algorithms (PL, Probability of HMI, etc.) are introduced. In Section 3, a number of possible approaches to deal with the RAIM problem are introduced, whereas in Section 4 the most popular RAIM methods developed in aviation are described. In Section 5 the challenges related to the adaptation of current aviation RAIM methods to land applications are illustrated, and in Section 6 an example of preliminary results of an IM prototype method in ITS is shown. Finally, in Section 7 conclusions on the state-of-art in IM and directions of present and future work are given.

2. Integrity and RAIM

2.1. Navigation performance parameters

The navigation system’s role is to collect and process measurements or other input data and deliver a position/state estimation, and guide the user to reach their destination. Based on the input data, called observables\(^1\), the parameters of interest are estimated. In the GNSS case, the model for the estimation problem is non-linear, but it is standard practice to transform it into a linear form, such as:

\[
y = Ax + \varepsilon
\]  

(1)

where \(y\) is a vector of \(m\) observables, \(x\) is the state vector (\(n\) components) of the parameters on which the observables depend, among which are the parameters of interest, the \(m \times n\) matrix \(A\) is the design matrix and \(\varepsilon\) is a vector of measurement errors. \(y\) and \(\varepsilon\) are random variables (indicated by an underscore).

In the GNSS case, the observable \(y\) is constituted by the range measurements (code and carrier phase) from each visible satellite, and in some cases by the Doppler observations, to determine

\(^1\)The term observable is used to refer to the random variable, while the term observation refers to its realization.
the velocity of the user. Such observable can be further augmented with external measurements/information, such as estimates of the ionosphere, troposphere, corrections for biases, or by other navigation systems, such as INS. The design matrix $A$ in (1) is determined by the geometrical configuration of the satellites in view, which links the range measurements to the unknowns, and by all the linear relations that link the eventually available additional information (e.g. INS or external corrections) to the unknowns.

In ITS, integrity is listed among the navigation key performance parameters (KPP), which have been identified [1, 2] as:

- **Accuracy.** Accuracy defines how well the estimated or measured position agrees with the true position. It is usually measured by the 95% confidence level for the position error, or by the Root Mean Square Error. Accuracy is computed assuming that the system is working in fault-free conditions, with standard performance.

- **Integrity.** Integrity defines the level of trust that can be given to the system. It is the ability of the positioning system to identify when a pre-defined Alert Limit (a bound to the position error) has been exceeded and to then provide timely warnings to drivers. Integrity is measured by either: a) the Probability of Hazardous Misleading Information, $P_{HMI}$, which is the probability that a position error larger than an Alert Limit (AL) occurs without a warning being timely raised, or b) the Protection Levels, which are the largest position error that may occur without any warning being timely raised, with probability smaller than the maximum allowed $P_{HMI}$.

- **Continuity.** Continuity is the capability of the navigation system to provide a navigation output with the specified level of accuracy and integrity throughout the intended period of operation (POP). Continuity is expressed as the probability that during the POP the system is providing trustworthy navigation information, without any disruption or Alert being raised.

- **Availability.** Availability is the fraction of time the navigation function is usable, as determined by its compliance with accuracy, integrity and continuity requirements. At any epoch of time, the navigation system is deemed either available or unavailable, depending on whether the availability, integrity and continuity requirements are satisfied.

The KPPs are inter-related. In particular, integrity is tightly connected with continuity, since raising an Alert constitutes a disruption to the continuity of the operations.

### 2.2. RAIM problem definition

Assume a single epoch scenario in which a user at an unknown position receives signals from the GNSS satellites, and eventual positioning information from other augmentation systems/external linkage. In this scenario, the RAIM problem is defined as: for any satellite geometry, to which corresponds a certain statistical distribution of the observable $y$, find an ‘acceptance’ region $\Omega \subseteq \mathbb{R}^m$ (sub-domain of $\mathbb{R}^m$) and an estimation/detection function $F(y)$ that to the observable $y \in \Omega$ assigns a position estimator $\hat{x}$:
\[
y \in \Omega \implies \hat{x} = F(y)
\]  

such that:

\[
P(\hat{x} \neq x \in \Omega_{\text{AL}} \cap \Omega) = P_{\text{HMI}} \leq P_{\text{HMI}} \forall x
\]  

and

\[
P(y \not\in \Omega) = P_{\text{FA}} \leq P_{\text{FA}} \forall x
\]

where:

- \( P_{\text{FA}} \) is the requirement of False Alert probability, the maximum allowable probability that an Alert is raised by the algorithm and the continuity of the operation is interrupted, without any actual reason. \( P_{\text{FA}} \) is a sub-allocation of the full continuity requirement \( \tilde{r} \), which has to account also for justified Alert (e.g. in the occurrence of an actual hazardous anomaly).

- \( \Omega_{\text{AL}} \) is the ‘integrity region’ around the true position which boundaries are the Alert Limits (AL). Fundamentally the position error is required to lie within the boundaries defined by the ALs (therefore inside \( \Omega_{\text{AL}} \)) with an extremely high probability, \( 1 - P_{\text{HMI}} \). While in aviation this region is cylindrical, with the radius of the cylinder defined by the Horizontal Alert Limit (HAL) and height defined by the Vertical Alert Limit (VAL), in ITS the shape of this region has not been defined yet, and possibly will be dependent on the specific application. It is expected that in most land applications the vertical error will not need to be monitored, and only limits in the horizontal plane will be considered. On the horizontal plane, distinction shall be made between along-track (AT) and cross-track (CT) directions of motion. A rectangular integrity region could be used, defined by the ALs in the two directions, \( AL_{\text{AT}} \) and \( AL_{\text{CT}} \). Alternatively, an ellipsoidal region could be adopted with semi-axes \( AL_{\text{AT}} \) and \( AL_{\text{CT}} \).

- \( P_{\text{HMI}} \), the (maximum allowed) Probability of Hazardous Misleading Information \( P_{\text{HMI}} \), is the integrity requirement per epoch. This is the probability that the information on the vehicle position is wrong by an amount larger than the ALs, without any alert or warning on possibly present anomaly being provided along. In aviation \( P_{\text{HMI}} \) values range from \( 10^{-7} \) to \( 10^{-9} \) per operation (e.g. approach), whereas for ITS there are yet no candidate values apart from those for aviation.

![Figure 1. Integrity regions in aviation and in ITS.](http://dx.doi.org/10.5772/intechopen.75777)
The acceptance region \( \Omega \) fundamentally defines the set of all the measurements \( y \) from which it is possible to determine a safe position estimate \( \hat{x} \), i.e., for which the requirement on the \( P_{HMI} \) is satisfied.

In any geometry, the rule can be optimized in different ways. The two extreme approaches would be: 1) minimizing the \( P_{HMI} \) given the requirement on the continuity is satisfied, or viceversa 2) minimizing the \( P_{FA} \) (maximizing the continuity \( c \)) given the requirement on the \( P_{HMI} \) is satisfied. The first is usually the preferred approach.

### 2.3. Protection levels (PL)

To define the PLs the total requirement on the \( P_{HMI} \), the \( P_{HMI} \), must be split into the different position components. In aviation, it is to be split into horizontal and vertical allocations, \( P_{HMI}^{hor} \) and \( P_{HMI}^{ver} \). In ITS instead, it is to be split between the horizontal along-track (AT) and cross-track (CT) components, \( P_{HMI}^{AT} \) and \( P_{HMI}^{CT} \), whereas the vertical component is (generally) not of concern. PL\(_{AT} \) and PL\(_{CT} \) are defined as the maximum position error size (in the AT direction and in the CT direction) that can pass undetected with a probability smaller or equal to the probability requirements, \( P_{AT}^{HMI} \) and \( P_{CT}^{HMI} \), i.e.,

\[
\begin{align*}
\text{PL}_{AT} &= \arg \min_{\delta} P(\hat{x}_{AT} - x_{AT} > \delta | \text{No Alert}) \leq P_{AT}^{HMI} \\
\text{PL}_{CT} &= \arg \min_{\delta} P(\hat{x}_{CT} - x_{CT} > \delta | \text{No Alert}) \leq P_{CT}^{HMI}
\end{align*}
\] (5)

with \( P_{HMI}^{AT} + P_{HMI}^{CT} = P_{HMI} \). To satisfy the navigation availability requirement it has to be:

\[
\begin{align*}
\text{PL}_{AT} &\leq \text{AL}_{AT} \quad \text{and} \quad \text{PL}_{CT} \leq \text{AL}_{CT}
\end{align*}
\] (6)

If those equations are satisfied integrity is maintained for the epoch under consideration. Instead of computing the PLs, the integrity monitoring system can simply compute the actual \( P_{HMI} \) or an upperbound for it, and then compare it to the requirement \( P_{HMI} \). If \( P_{HMI} \leq P_{HMI} \), integrity is maintained.

### 2.4. RAIM input, output and performance parameters

In this Section the input and output parameters of a RAIM algorithm are summarized. Figure 2 shows a schematic representation of a RAIM algorithm. A RAIM algorithm is constituted of two blocks: the first one assesses the geometry or model strength the second one processes the real time observations and assesses their coherency.

The model strength assessment takes as input the design matrix \( A \) and the distribution function of the observable \( f_y \), i.e., the observation model, at each epoch. Output of this first assessment are the PLs and/or the \( P_{HMI} \) and consequently the availability prediction for that epoch: if any PL > AL, or equivalently \( P_{HMI} > P_{HMI} \), the navigation service is declared...
unavailable. Such availability assessment can be made at each epoch on the basis of the model strength, before the actual measurements are taken.

The observation coherency assessment takes as input the observations \( y \) at each epoch. The output of the observations processing is the issue of a state of either Alert or No Alert for that epoch; in case of No Alert, a position solution is provided to the user. In this step, a real time check of the observations is performed. Alert is declared in case the sample measurement taken is too inconsistent: the case \( y \notin \Omega \) introduced in the definition of RAIM problem.

Both blocks of the RAIM structure require as input the navigation requirements on integrity and continuity, i.e., \( \Omega_{AL} \), the integrity region, \( P_{HMI} \), the maximum allowed \( P_{HMI} \), and \( P_{FA} \), the False Alert (or continuity) requirement. The performance of a RAIM algorithm can be measured over time by computing (estimating) the actual \( P_{FA} \), \( P_{HMI} \) and PLs.

### 3. RAIM approaches

Since RAIM is linked with the estimation method, two approaches to RAIM can be distinguished:

- **Fault Detection and Exclusion (FDE) procedure**: one adopts a standard estimation rule, for instance the Best Linear Unbiased Estimation (BLUE) [8], characterized by highest accuracy in fault-free conditions; in case the BLUE is not satisfying the integrity requirements (i.e., too large \( P_{HMI} \) because a fault is suspected), one can switch to a different estimator, e.g., a BLUE applied on a subset of the original measurements set. In this way the suspected fault is excluded, and the associated bias in the estimation removed.

- **Robust estimation**: one adopts an estimation rule tailored to integrity. Instead of employing the BLUE, one can sacrifice on some accuracy in fault-free conditions to gain in integrity.

A combination of both methods listed above is also possible. Here only FDE procedures are analyzed.
3.1. FDE procedure

In an FDE procedure one assumes the possible occurrence of different hypotheses, the fault-free case (null hypothesis $H_0$) and the occurrence of fault/anomalies (alternative hypotheses $H_i$). An FDE procedure is applied to detect whether an anomaly is affecting the system, and, in case of detection, exclude the anomalous observations. In a common FDE procedure, typically the BLUE is applied to the model corresponding to the hypothesis $H_i$ that is more likely (or safer to use). Once it has been decided which hypothesis is most likely to hold true (this decision is made through a statistical testing procedure), the estimator to be used is the BLUE for the model corresponding to that hypothesis. The BLUE for the unknown $x$ in the linear model (1), assuming known dispersion of $\varepsilon$, i.e., $D(\varepsilon) = Q_x$, reads:

$$\hat{x} = S y$$

(7)

with $S = (A^T Q_x^{-1} A)^{-1} A^T Q_x^{-1}$ the pseudo-inverse of matrix $A$ in the metric defined by $Q_x$.

Fundamentally $\hat{x}(y)$ in this approach will be constituted by different linear functions of the observable: it will be in the form of (7) when the null hypothesis $H_0$ is considered most likely, or conversely different forms $\hat{x}$ when they one of the alternative hypothesis is designated to be most likely.

3.2. Statistical hypothesis testing

FDE procedures are based on statistical hypothesis testing [9]. In an FDE procedure statistical tests are performed to determine which hypothesis (fault-free/faulty) on the system state is most likely to hold, and determine the observable domain subdivision discussed in Section 3.1. In this chapter only linear models are analyzed, therefore a special attention shall be given to statistical hypothesis testing in linear models. The aim is to decide between competing linear models that could describe the observed phenomenon or process, once an observation has been made. Furthermore the observables are assumed to have normal distributions, and different hypotheses differ only in the specification of the functional model. The models considered are thus Gauss-Markov models [10].

Given the linear model of Eq. (1), we assume the random noise distribution to be known, Gaussian and zero mean:

$$\varepsilon \sim N(0, Q_x)$$

(8)

The linear system in (1) represents the state of standard or nominal operations, that is the case in which the system is working properly without any fault. This state is considered as the null hypothesis $H_0$. The case of a fault affecting the system constitutes instead a different state, described by an alternative hypothesis $H_a$, under which the linear model assumes a different form. Therefore:
\[ H_0 : \quad y = Ax + \varepsilon \]
\[ H_a : \quad y = Ax + C_y \nabla + \varepsilon \]

where \( C_y \) is a \( m \times q \) matrix which represents the ‘signature’ of the errors in the measurements and \( \nabla \) is a \( q \)-sized vector that contains the sizes of the biases in each degree of freedom (\( q \)) of \( C_y \).

To test \( H_a \) against \( H_0 \), the Uniformly Most Powerful Invariant (UMPI) test statistic (through application of the Generalized Likelihood Ratio (GLR) criterion) reads:

\[ T_q = \frac{\mathbf{e}_0^T Q_y^{-1} \left( C_y^T Q_y^{-1} Q_y \mathbf{e}_0 Q_y^T C_y \right)^{-1} C_y^T Q_y^{-1} \mathbf{e}_0}{\mathbf{e}_0^T Q_y^{-1} \mathbf{e}_0} \]

where \( \mathbf{e}_0 = y - A \hat{x}_0 \) is the vector of residuals computed considering the null hypothesis holding true (\( \hat{x}_0 \) being the position estimator under the null hypothesis, obtained by Eq. (7)).

The test statistic \( T_q \) is \( \chi^2 \) distributed:

\[ H_0 : \quad T_q \sim \chi^2(q, 0) \quad \text{and} \quad H_a : \quad T_q \sim \chi^2(q, \lambda) \]

with non-centrality parameter:

\[ \lambda = \nabla^T Q_y^{-1} \nabla \]

Knowing (though only partially in case of \( H_a \)) the distributions of the test statistic under the different hypotheses, one can define a critical region \( K \) (to reject the null hypothesis) on the basis of type I and type II error probabilities. The critical region is one sided, of the type:

\[ K : \quad T_q > k \]

with \( k \) the test threshold (or critical value).

The theory above constitutes the basis of statistical hypothesis testing in linear models, that allows to build the specific test in the simple binary case of null versus one alternative hypothesis. In case one has to choose among multiple alternative hypotheses, one option is to employ a set of binary tests. However, a number of different methods exist in statistics, aiming to answer this more complex problem. Hypothesis testing based methods are known as Multiple Comparisons methods [11], while other methods that do not recur to hypothesis testing are known as Subset Selection methods [12].

4. The aviation legacy

In this section, first the observation models and the typical assumptions adopted in civil aviation applications are described, and next the two most popular RAIM algorithms developed for such
applications, i.e., the Weighted RAIM [13] and the Advanced RAIM (ARAIM) [14] algorithms, are introduced.

4.1. GNSS anomalies and their models

The main categories of High Dynamics Threats (HDTs) to be monitored in aviation applications, which rely on code-based SPP, are here listed. The HDTs are threats that cannot be monitored by the GNSS ground control system, as opposed to the Low Dynamics Threats (LDTs) [14]. They are categorized into:

- Clock and ephemeris estimation errors, see [15];
- Signal deformations, see [16];
- Code-carrier incoherency, see [17];

From the snapshot perspective (considering a single epoch of time), and working with carrier-phase smoothed code measurements, an outlier in a single satellite is believed to be the main threat (in terms of probability of occurrence). Simultaneous outliers on multiple satellites (wide failure errors) can occur, but with a much lower likelihood [14]. Among these are the constellation faults (e.g., upload of incorrect navigation messages that may impact a full constellation).

Errors/anomalies in signal propagation, as ionosphere, troposphere and multipath, shall not be considered hazardous for aviation when the new civil frequency L5/E5, and new certified receivers, are available: the tropospheric delay has typically a small effect (and one can correct sufficiently well for this error source), ionosphere gradients/front effects are supposed to cancel out with the use of ionosphere-free combination, and multipath depends on the local satellite-receiver geometry and can be considered on a per satellite basis (typically outlier-like).

4.2. General distribution of the observable

In the previous section the main threats possibly affecting the positioning system in aviation applications have been described: on this basis a model to describe the distribution of the observable, able to take into account the possible occurrence of anomalies, shall be formulated. The pdf of $y$ is generally supposed to be known in standard fault free conditions, but it cannot be fully defined in the presence of anomalies. However, it is assumed that anomalies in the systems will occur with a low failure rate.

Different hypotheses can be defined to represent the state of the system: a fault free (null) hypothesis $\mathcal{H}_0$ and $N_\alpha$ alternative hypotheses $\mathcal{H}_i$, representing the different possible types of anomalies affecting the system, with $i = 1, \ldots, N_\alpha$. Here only linear models are considered, and hypotheses of the type of Eq. (9), i.e., $\mathcal{H}_i: y = Ax + \nabla y_i + \xi$.

Single satellite faults and constellation faults can be modeled by different $C_i$ matrices: in case of single satellite faults, or combinations of independent single satellite faults, the main $C_i$ to consider shall be the canonical unit vectors of $\mathbb{R}^m$ or $m \times q$ matrices made up of different canonical unit vectors of $\mathbb{R}^m$, respectively; in case of constellation faults, a matrix $C_i$ of $m - n$ columns, fully complementing $A$ in the vector space $\mathbb{R}^m$, shall be used.
The distribution of the observable $y$ depends on the state of the system. Under each hypothesis, $y$ is assumed to be distributed as a multivariate normal distribution (Eqs. (1), (8) and (9)). It is possible to associate prior probabilities to the occurrence of the different hypotheses, in such a way that the variable $H$, representing the state of the system, has a prior Probability Mass Function (PMF), with discrete values $p_i$ for each realization. Thus $H$ and $y$ marginal distributions are:

$$
\begin{align*}
H &\sim \begin{cases} 
P(H = H_0) = p_0 \\
P(H = H_1) = p_1 \\
P(H = H_{N_a}) = p_{N_a}
\end{cases}
\Rightarrow y \sim p_0 \cdot f_{y|H_0} + \sum_{i=1}^{N_a} p_i \cdot f_{y|H_i}
\end{align*}
$$

(14)

At this point the uncertainty about the $y$ distribution is expressed by its dependence on the unknown variable $\nabla_i$ beside $x$. To tackle this uncertainty, most RAIM algorithms assume worst-case bias scenarios or compute bounds for the worst-case risk that could result, see for instance [18].

4.3. Weighted RAIM

In [13] a Weighted RAIM implementation is described. This constitutes one of the first relevant RAIM algorithms conceived and is still in use today, typically implemented in aviation grade GPS receivers, to provide low-precision lateral integrity only. The method consists of the two steps defined in Section 2.4, the model strength assessment and the real time observation coherency assessment. Even though not theorized in the original paper, the method is based on the assumption of the observable distribution described in previous section, with the constraint that only single satellite faults are possibly occurring.

A single test, the OMT, is used to judge the quality of the observations at each epoch [13]. The OMT, also known as $\chi^2$ test, is a UMPI test that employs a test statistic of the form of Eq. (10), and addressing a most generic anomaly, i.e., with $q = m - n$. Such test statistic coincides with the Weighted Sum of Squared Errors (WSSE), defined as:

$$\text{WSSE} = e^T Q^{-1} e$$

(15)

If this statistic exceeds a certain threshold $k$, the estimated position is assumed significantly biased; otherwise, it is assumed acceptable. This threshold is chosen to meet the probability of False Alert requirement, $P_{FA}$, knowing that in the fault-free hypothesis, the WSSE is distributed as a central $\chi^2$ with $m - 4$ degrees of freedom (using GPS only).

4.4. Model strength assessment

If a range error from one measurement occurs, the expected value of the test statistic grows, along with, proportionally, the expected position error. The satellite geometry determines how
the error in the range domain propagates into the position domain. The original Weighted RAIM algorithm focuses on monitoring only the vertical component of the position solution, but the same reasoning can be made for the other components. In a simple two-dimensional graph, plotting $\sqrt{\text{WSSE}}$ on the horizontal axis and the vertical position error on the vertical axis, their relation can be represented by a straight line (see Figure 3), with a steepness (slope), for satellite $i$ given by:

$$V_{\text{slope}} = \frac{|S_{[3,6]}|\sigma_i}{\sqrt{1 - P_{A_{0,6}}}}$$

with $S = (A^TQ_y^{-1}A)^{-1}A^TQ_y^{-1}$, $\sigma_i = \sqrt{Q_{yy}} = \sigma_y$, and where the subscripts in square brackets indicate the indexes of the matrix elements. The Vertical Protection Level (VPL) is computed as:

$$VPL \equiv \max_i \left( V_{\text{slope}} \right) k + k_{MD}\sigma_{s_3} \quad i = 1, 2, \ldots, m$$

where $k$ and $k_{MD}$ are obtained as:

$$k = \sqrt{\text{inv-}\chi^2_{m-n}(P_{FA})}; \quad k_{MD} = \Psi^{-1}\left( \frac{P_{HMI}}{m p} \right)$$

with $\text{inv-}\chi^2_{m-n}(\cdot, m-n)$ the inverse of a central $\chi^2$ CDF function with $m-n$ degrees of freedom, $\Psi(\cdot)$ the tail probability of the cumulative distribution function of a zero mean unit Gaussian distribution, and $p$ the a-priori probability of hazardous fault in one satellite. The above formulas for the VPL are based on the following expression of the integrity risk under an alternative hypothesis:

$$P_{HMI|H_i} = P_{MD_i} \cdot P(\bar{x}_3 - x_3 > VAL|H_i)$$

which assumes that an integrity event corresponds to the simultaneous occurrence of an MD and a positioning error larger than the Vertical AL (VAL), and is justified by the fact that test

Figure 3. Representation of the weighted RAIM’s $V_{\text{slope}}$ concept.
statistic and positioning error are uncorrelated. The VPL is the measure of the observation model strength: if VPL > VAL, integrity is not available for the geometry considered.

4.5. Real time availability

The real time availability assessment is performed if the model strength assessment was passed successfully (VPL < VAL). At each epoch, once the observations are collected, the WSSE is computed and compared with the threshold. As in standard hypothesis testing, we have the following decision rule:

\[
\text{If WSSE} > k, \text{ reject the fault-free hypothesis and declare Alert (20)}
\]

else standard operations continue.

4.6. ARAIM

The Weighted RAIM presented in the previous section was developed for the single GPS constellation and has been found generally suboptimal, even though presenting a very practical and efficient approach. An enhanced approached, known as ARAIM, provides the following improvements [14]:

- in addition to single satellite faults, multi-dimensional faults (affecting multiple satellites at a time) are accounted for [14, 18];
- the potential of the multi-constellation GNSS is fully exploited, instead of GPS only [14, 18];
- rather than using only single-frequency observations, use of dual-frequency observations, to remove the first order ionospheric delay, is foreseen, [14, 18];
- a proof of safety is given [14, 18]. Weighted RAIM is not proven to be always conservative;
- different statistical tests, more tailored to detecting faults that have sensible impact on the position estimate [19], are employed.

The basic concepts of ARAIM are here outlined. For more details, see [14, 18, 19]. Figure 4 shows a block diagram representation of the ARAIM algorithm. From a statistical point of view, ARAIM is based on the following concepts:

- Multiple Hypothesis approach with a-priori probabilities: the system is supposed to be in one out of a set of different possible states described by multiple hypotheses, to each of which is assigned an a-priori probability of occurrence (Section 4.2). The \( P_{\text{HMI}} \) is computed by the sum of the \( P_{\text{HMI}} \) under the different hypotheses, weighted on the base of their prior probabilities.
- Solution Separation (SS) as test statistics: to discriminate between hypotheses, to eventually exclude faulty measurements, the difference between the position solutions under the different alternative hypotheses and the null hypothesis is computed and used as a test statistic. For each alternative hypothesis considered a difference vector (SS) is computed and a test is run for each of the position components of the vector.
If one characterizes each alternative hypothesis by a different subscript $i$, the $i$th Solution Separation vector can be written as:

$$TSS_i = b x_i - b x_0$$

where $b x_0$ and $b x_i$ are the position solutions obtained employing the null and the alternative model respectively, i.e.:

$$b x_0 = A^T Q_{y}^{-1} A^{-1} A^T Q_{y}^{-1} y = S_y$$

$$b x_i = A^T Q_{y}^{-1} A^{-1} A^T Q_{y}^{-1} y_i = S_i y$$

where $Q_{y}^{-1}$ is obtained from $Q_{y}^{-1}$ replacing the diagonal elements corresponding to the faulty satellites in hypothesis $H_i$ with 0 (this means giving zero weight to such observations). In practice, these tests have similar performance to the UMPI tests (see [20, 21]).

4.7. Model strength assessment

The PLs are computed on the basis of the model strength (satellite geometry and stochastic model), and compared to the AL to determine the integrity availability. The computation of the PLs is based on an iterative procedure: the PLs are determined in such a way that the sum of the $P_{HMI}$ under each alternative hypothesis is equal to the full $P_{HMI}$ requirement:

$$P_{HMI} = \sum_i N_i P_{HMI} = \sum_i N_i \left[ P_{MD} \cdot P(\bar{x} - x \notin \Omega_{AL}[H_i]) \right]$$

Figure 4. ARAIM baseline architecture. The algorithm checks the coherency of the observations by means of the solution separation tests, evaluates the possibility of excluding corrupted observations with exclusion specific tests, and computes the PLs. Integrity is guaranteed if PLs < ALs.
where in the last equality the relation (19) is applied. As a result, the VPL must satisfy the following equation [18]:

\[
2\Psi\left(\frac{VPL}{\sigma_{x,3}}\right) + \sum_{i=2}^{N} P_i\Psi\left(\frac{VPL - k_{i,3}}{\sigma_{x,1}}\right) = \zeta \Psi_{\text{HMI}}
\]  

(24)

where \(k_{i,3}\) is the test threshold for the \(i\)th SS test, 3rd component (vertical), and \(\zeta\) is the fraction of the full \(\Psi_{\text{HMI}}\) allocated to the vertical direction (\(0 < \zeta < 1\)). The HPL instead is computed with \(\text{HPL} = \sqrt{\text{PL}^2_{x_1} + \text{PL}^2_{x_2}}\), where \(\text{PL}^2_{x_1}\) and \(\text{PL}^2_{x_2}\), the PLs for the two horizontal components, are computed with formulas equivalent to (24). The thresholds \(k_{i,j}\) are computed with

\[
k_{i,j} = -\Psi^{-1}\left(\frac{P_{\text{FA}}}{4N_n}\right)\sigma_{\Phi,\hat{x}_{i,j}}
\]  

(25)

where \(\sigma_{\Phi,\hat{x}_{i,j}}\) is the standard deviation of the corresponding SS test statistic (see [18]). If any PL \(\text{PL} > AL\), integrity is not available for the geometry considered.

4.8. Real time availability

In ARAIM the testing is subdivided in two steps:

1. Detection tests: the SS tests are computed and compared to their thresholds; if none of the tests exceeds the threshold, the fault-free hypothesis is confirmed and standard operations continue, otherwise the algorithm proceeds to the next step.

2. Exclusion confirmation tests: extra tests are run to determine if it is safe to exclude some observations and continue to provide navigation service. These tests are meant to minimize the risk of wrong identification. More details are given in [18].

After detection and eventual exclusion of observations, the PLs are re-computed (as post-observations PLs) and compared with the thresholds. If any \(\text{PL}_{\text{post}} > AL\), an Alert is raised.

5. The ITS challenge

As mentioned in Section 1.2, when moving from aviation to land applications, a number of issues have to be taken into account in the context of integrity monitoring. The main two issues are:

- Positioning has to be performed in urban environment: additional vulnerabilities are to be taken into account, i.e., multipath, NLOS, interference and spoofing.

- Higher precision/smaller PLs are required: this may lead to the use of precise positioning techniques (PPP, RTK) with their additional vulnerabilities, as well as additional navigation sensors/technologies (INS, V2I and V2I communication, camera, etc.).

The main assumptions on which the FDE procedures and RAIM algorithms described so far rely on are (Section 4.2): 1) linear estimation problem, 2) Gaussianity of the observables and 3)
mean-shift model for the anomalies. In land applications, these assumptions are likely to hold, though multipath and NLOS may challenge the second one, while the large number of observations available and vulnerabilities increases the computational complexity of FDE procedures. These aspects are addressed in more detail in the following.

5.1. Urban environment: multipath, NLOS and interference

Multipath is the most significant source of measurement errors in ITS applications, as it is dependent on the environment surrounding the antenna and is especially intense in dense urban areas. Buildings and other obstacles degrade the signal reception in three ways: 1) signals are completely blocked and unavailable for positioning, 2) signals are blocked in their direct path, but are still received via a reflected path, with the NLOS reception, 3) both direct Line-Of-Sight (LOS) and reflected signals are received, i.e., the case of multipath. NLOS code signals can exhibit positive ranging errors of tens of meters magnitude in dense urban areas.

Numerous innovative techniques have been developed in the recent years to address the multipath and NLOS threats in urban environment. Interest was raised by 3D-map-aided (3DMA) GNSS, a range of different techniques that use 3D mapping data to improve GNSS positioning accuracy in dense urban areas. 3D models of the buildings can be used to predict which signals are blocked and which are directly visible at any location [22, 23]. A technique that determines position by comparing the measured signal availability and strength with predictions made using a 3D city model over a range of candidate positions is known as the shadow matching technique [24]. Such techniques may possibly be integrated with RAIM algorithms for ITS in the near future.

5.2. Precise positioning techniques and multi-sensor integration

The use of precise positioning techniques rather than SPP and the need of integration with other sensors bring a number of complications to the IM for land applications. Some of the main challenges are summarized in the following.

5.2.1. PPP and RTK: Carrier phase observations vulnerabilities

Precise positioning techniques employ carrier phase observations next to code observations. Even though the estimation problem is characterized by a much larger number of observations and unknowns to solve for, it is still a linear estimation problem. The same hypothesis testing theory applies, and therefore, the same RAIM concepts developed for aviation can be implemented, with appropriate adjustments. However, one drawback of the ARAIM is the associated heavy computational burden, due to the need of running a test for each possible combination of simultaneously biased observations. When multi-systems, multi-frequency and carrier phase based positioning is in use, the total number of combinations of possibly biased observations increases dramatically — so does the computational load for the algorithm. It is thus possible that the current ARAIM approach will not be optimal.

Another issue is constituted by the additional vulnerabilities that affect the precise positioning techniques, mainly carrier phase multipath and cycle-slips. Multipath affects carrier-phase observations with the same mechanism as code observations [25]. Carrier-phase multipath is
one of the critical elements in determining the Time to Ambiguity Resolution (TAR), which can become of concern in regards of timeliness requirements. Furthermore, cycle-slips, which constitute the main RTK-specific threat, as they can cause wrong ambiguity fixing and result in large errors in the positioning, require specific additional monitoring. There is a vast literature on cycle-slip detection, e.g., by [26, 27]. Most cycle-slip detection methods are based on hypothesis testing, but exploit the multi-epoch data processing to increase their detection power.

5.2.2. Multi-sensor integration and recursive data processing

Use of multiple sensors for navigation means that extra observations shall be integrated with the GNSS observations. If the extra observations are linear in the unknown parameters, they can be simply stacked together in the same linear estimation problem. Integration with INS is a complex problem on which a large literature exist [28]. Finally, while the focus of this chapter was only on snapshot RAIM (single epoch), RAIM techniques for multi-epoch recursive data processing are under development [29].

5.2.3. Cooperative integrity monitoring (CIM) concept

Section 2.4 shows that an integrity assessment can be made before the observations are taken, when only satellite geometry and environment are known or partially known. New IM concepts intend to exploit the fact that satellite geometry and satellite visibility can be reasonably predicted at any time and location (for instance with the use of city models), and that the same observability conditions repeat periodically over time. Beside the environment nearby the receiver in its nominal conditions, these new concepts plan to exploit also the potentialities offered by a Vehicular Ad-hoc Network (VANET) infrastructure [30]. The potential availability of multiple observations of GNSS signals, taken by different vehicles participating to a VANET, can be shared and combined in order to implement a collaborative spatial/temporal characterization and prediction of the local degradation of the GNSS signals.

6. An example

In this section, the results of a first attempt to perform IM in urban environment employing the RTK positioning method with a short baseline, and applying a prototype ARAIM algorithm, are shown. Such results are only indicative, since most of the assumptions behind the use of ARAIM in an RTK set-up are yet to be justified.

A kinematic test is conducted for practical demonstration of IM for ITS. A small vehicle is fitted with a Trimble multi-GNSS geodetic receiver and a survey-grade antenna. The test is carried out in a dense urban area in Tokyo, Japan. The RTK system uses GPS, GLONASS and BeiDou dual-frequency observations with a sampling rate of 10 Hz. A prototype RAIM algorithm derived from ARAIM is implemented. Due to the lack of common standards, the PLs are computed in the test using different values of $P_{HMI}$ ranging from $10^{-3}$ to $10^{-6}$ in order to track empirically the impact of $P_{HMI}$ on the obtained results. A false alert probability ($P_{FA}$) of 0.01 is applied.
Figure 5 shows the PL for the along-track and cross-track directions (shown as PL\(_{AT}\) and PL\(_{CT}\)) and the absolute values of the positioning errors along these directions (denoted as err\(_{AT}\) and err\(_{CT}\)) using an integrity risk of 10\(^{-4}\) and 10\(^{-6}\) as examples. The figure shows that the RTK with correct ambiguity fixing gives positioning errors within a few centimeters. The average absolute value of the AT and CT positioning errors are 0.058 and 0.054 m, respectively. The FDE method detected 15 code observations with severe irregularities, which are attributed to high multipath in this environment. These observations were excluded from further processing. There were a few cases where the ambiguity fixing seemed to be incorrect by one or two cycles, which were not detected by the FDE procedure. However, the PL adapted to these situations and bounded this error as illustrated in the Figure 5. Inspection of the Figure also shows that when using RTK with correct ambiguity fixing, an Alert Limit (AL) can be safely chosen as 1 m. The sub-decimeter positioning accuracy of RTK is bounded by a tight protection level. The positioning errors in the test were always bounded by the PLs, and PLs < ALs for the whole period, with an integrity availability of 100%. The medians of the PL for the AT and CT using different integrity risk (\(P_{HMI}\)) values are given in the Table 1. Both table and Figure 5 show that the PLs increase with the decrease of the allowed integrity risk.

<table>
<thead>
<tr>
<th>Integrity risk ((P_{HMI}))</th>
<th>(10^{-3})</th>
<th>(10^{-4})</th>
<th>(10^{-5})</th>
<th>(10^{-6})</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL(_{AT})</td>
<td>0.176</td>
<td>0.197</td>
<td>0.215</td>
<td>0.232</td>
</tr>
<tr>
<td>PL(_{CT})</td>
<td>0.148</td>
<td>0.164</td>
<td>0.177</td>
<td>0.188</td>
</tr>
</tbody>
</table>

Table 1. Median PL\(_{AT}\) & PL\(_{CT}\) in meters for different values of integrity risk \((P_{HMI})\).
7. Concluding remarks

In this chapter the concepts of integrity and IM have been introduced, and the main RAIM methods currently in use or under development have been presented. As these methods were developed in the aviation context, their adoption in land applications has been discussed. The positioning methods used in land applications still satisfy the assumptions made by current RAIM algorithms, though great care shall be taken in addressing the larger number of vulnerabilities affecting the positioning system, in particular multipath and the carrier-phase specific vulnerabilities. Some preliminary but promising results of the application of a RAIM algorithm in urban environment were shown. Further research and practical experiments are necessary to strengthen the confidence in the models.

Acknowledgements

Dr. Nobuaki Kubo from Tokyo University of Marine Sciences and Technology is acknowledged for providing the RTK kinematic test data in Japan.

Author details

Davide Imparato¹, Ahmed El-Mowafy¹* and Chris Rizos²

*Address all correspondence to: a.el-mowafy@curtin.edu.au

1 Curtin University, Bentley, WA, Australia
2 UNSW, Kensington, NSW, Australia

References


[10] Teunissen PJG, Simons DG, Tiberius CCJM. Probability and Observation Theory. Faculty of Aerospace Engineering, Delft University of Technology: Delft University; 2005


[22] Suh Y, Shibasaki R. Evaluation of satellite-based navigation services in complex urban environments using a three-dimensional GIS. IEICE. 2007;E-90(B):1816-1825


