# Testing Continuity and Integrity for Advanced RAIM and Similar Systems

Juan Blanch, Todd Walter, *Stanford University*

Fan Liu, *FAA*

**ABSTRACT**

Advanced RAIM is an extension of RAIM that will integrate up to four core constellations (GPS, Galileo, BeiDou, and GLONASS), and signals in two frequencies (L1 and L5). The performance standards and an example Horizontal ARAIM user algorithm, which will be included in the future release of the DFMC SBAS Minimum Operational Performance Standards (MOPS) (ED-259B/DO-401A), are currently being developed. A key part of these standards is the set of offline tests designed to demonstrate compliance with the integrity and continuity requirements. In this paper, we go over the key aspects of the tests described in the draft DFMC SBAS MOPS and provide a rationale or justification for each one of them.

**INTRODUCTION**

GPS L1 based Receiver Autonomous Integrity Monitoring (RAIM) has provided worldwide horizontal guidance to hundreds of thousands of aircraft since the mid 1990's. However, because it only uses one constellation and one signal in one frequency, it does have some limitations. For example, its performance is very sensitive to the geometric strength of the GPS constellation. The use of a model to correct the ionospheric delay also places a high floor on the pseudorange error bounds [1,2,3,4,5].

Advanced RAIM is an extension of RAIM that will integrate up to four core constellations (GPS, Galileo, BeiDou, and GLONASS), and signals in two frequencies (L1 and L5). The use of at least an additional constellation is expected to greatly mitigate both the geometric weakness of RAIM and the additional frequency will mitigate the errors due to the residual ionospheric delay. As with RAIM, SBAS, and GBAS, the performance of ARAIM must be standardized before it can be used by civil aviation [6].

The standards for Horizontal ARAIM, which will be included in the next version of the Dual-Frequency Multi-Constellation (DFMC) Satellite-Based Augmentation System (SBAS) Minimum Operational Performance Standards (an initial version [10] is already available), are currently being developed. A key part of these standards is the set of offline tests designed to demonstrate compliance with the integrity and continuity requirements.

To evaluate an ARAIM algorithm, the key aspects of testing include scaling up the PHMI for testing, determining the number of trials and passing criteria in order to test the modified PHMI with sufficient confidence, defining the faults to be tested, which includes the probability of each fault mode, fault bias, and fault duration, and defining the simulated nominal noise and nominal bias. In this paper, we go over these key aspects and provide a rationale or justification for each one of them.

The rationales for the following topics are presented:

1) Scaled PHMI for testing. Demonstrating that PHMI is less than $10^{-7}$ per hour would require an impractically large number of trials. The PHMI is scaled up based on the ISD and total probability of all fault modes.
2) Number of trials and passing criteria. The number of trials must be sufficient to ensure that an adequate algorithm passes the tests with high probability and that an inadequate one fails with high probability. We will derive the number of trials and passing criteria based on the scaled PHMI using the same method to accept a good design and reject a bad design as existing RAIM testing
3) The probability of each fault mode. The set of simulated fault modes is determined by the fault rate and the mean fault duration included or implied by the Integrity Support Data (ISD).
4) Fault bias magnitude and temporal profile. The fault magnitude is designed to approximate a worst-case scenario. This is a key aspect of these tests, because ARAIM will need to protect against worst-case multiple simultaneous faults (like for example a constellation-wide fault [7]). The fault bias is simulated as a ramp. This is to ensure that the test is not missing the worst-case magnitude for a given algorithm.
5) Fault Duration. The fault duration is based on the mean fault duration of the simulated faults. However, it must be adjusted to account for the limited duration of each trial.
6) Nominal noise. The simulated nominal noises are correlated and still yield enough effective samples during the exposure time of $T_{EXP}$=1 hour for each trial.
7) Nominal bias. When nominal bias specified in ISD is non-zero, it is simulated to maximize the impact on the user error in order to stress test the ARAIM algorithm.

For the elements that have been documented elsewhere, we will refer readers to the appropriate references. We will focus on elements that have not been documented before.

Although not included in the MOPS, we also describe a necessary condition on the Protection Level. Specifically, we include a lower bound on the achievable PL [8]. It may be used either as a sanity check on the formulation of the PL or its implementation.

Our purpose with this paper is to ensure complete transparency of the tests used to evaluate and ensure that the integrity and continuity requirements of aviation GNSS receivers are met. We welcome the feedback of the community to improve these tests either by making them more computationally efficient or more reliable.

## SCALED PHMI FOR TESTING

The goal of the tests is to ensure that the probability of hazardously misleading information (PHMI) is met by a particular ARAIM user algorithm. The PHMI is the probability that the position error exceeds the protection level (PL) and there is no alert, that is, the receiver assumes that the PL is correct.

At the outset, we can see that testing this requirement with sufficient confidence is potentially problematic, because the required PHMI is $10^{-7}$ per hour. Without further assumptions, demonstrating that this probability is met would require an impractically large number of trials (see next section for some examples).

To reduce the number of trials, we must scale up the probability we are evaluating. This can be done as follows. We start by writing the total probability for the PHMI with the explicit contribution of each fault hypothesis $H_k$ (with $H_0$ being the fault-free hypothesis):

$$P(HMI) = P(HMI \mid H_0)P(H_0) + \sum_{k \geq 1} P(HMI \mid H_k)P(H_k) \qquad (1)$$

We now assume that the contribution of the fault-free hypothesis is negligible. In that case, we can write that:

$$P(HMI) = \sum_{k \geq 1} P(HMI \mid H_k)P(H_k) \qquad (2)$$

If we now write:

$$\frac{P(HMI)}{P_{norm}} = \sum_{k \geq 1} P(HMI \mid H_k)\frac{P(H_k)}{P_{norm}} \qquad (3)$$

We can see that we can evaluate $\frac{P(HMI)}{P_{norm}}$ by scaling up the probabilities of fault from $P(H_k)$ to $\frac{P(H_k)}{P_{norm}}$. The only requirement on $P_{norm}$ is that:

$$\sum_{k \neq 0} P(H_k) \leq P_{norm} \qquad (4)$$

For the default Integrity Support Data (ISD) and assuming no more than 12 GPS satellites and 12 Galileo satellites in the navigation solution, $P_{norm} = 1.2 \times 10^{-3}$ is adequate. It leads to the requirement on the scaled PHMI:

$$\frac{P(HMI)}{P_{norm}} \leq 8.3 \times 10^{-5}$$

**NUMBER OF TRIALS AND PASSING CRITERIA**

A trial is defined as:

1) Drawing a fault mode from the possible set of fault modes according to the scaled probabilities defined above in Eq. (3)
2) Simulating pseudoranges with the injected fault for a period of one hour
3) Computing the position solution and the PL at each step

A failure occurs when the position solution error exceeds the PL without being detected at any time during the trial.

The necessary number of trials is determined using the following two probabilities:

$P_\alpha$, the probability that compliant equipment (i.e. missed alert probability of $p = 1\times10^{-7}/P_{norm}$) does not pass the test (the objective is to keep this probability to be less than 1%); and

$P_\beta$, the probability that equipment with a missed alert probability of $2 \cdot p = 2\times10^{-7}/P_{norm}$ passes the test (the objective is to keep this probability to be less than 1%).

In the Appendix, we show that a good approximation of the number of trials $N_{trials}$ and the failure threshold $N_{max}$ is given by

$$N_{trials} = \left\lceil \frac{\left(\sqrt{2}Q^{-1}(P_\beta) + Q^{-1}(P_\alpha)\right)^2}{p} \right\rceil \quad (5)$$

$$N_{max} = \left\lfloor \left(\sqrt{2}Q^{-1}(P_\beta) + Q^{-1}(P_\alpha)\right)\left(\sqrt{2}Q^{-1}(P_\beta) + 2Q^{-1}(P_\alpha)\right) \right\rfloor \quad (6)$$

where $Q$ is the complementary cdf of a normal distribution.

In our case, we have:

$$N_{trials} = \left\lceil \frac{P_{norm}\left(\sqrt{2}Q^{-1}(P_\beta) + Q^{-1}(P_\alpha)\right)^2}{PHMI} \right\rceil$$

For $P_{norm} = 1.2 \times 10^{-3}$, we obtain approximately 380000 trials and a corresponding failure threshold $N_{max}$ of 45. If the probabilities are not scaled up, i.e., P(HMI) is $10^{-7}$, the required number of trials would be more than $3 \times 10^8$, which is very impractical.

The trials are distributed among 40 different geometries representing a range of PLs, with single and dual constellation, minimum and maximum number of satellites used in the navigation solution, etc.

## FAULT MODES

The fault mode probabilities $P(H_k)$ are computed based on the parameters included in the ISD as follows:

$$P(H_k) = R_k \cdot (MFD + T_{EXP}) \tag{7}$$

where

$R_k$ is the fault rate

$MFD$ is the mean fault duration

$T_{EXP}$ is the exposure time (1 hour)

We note that for dual fault events, the rate and the mean fault duration are computed using the formulas described in [9]. For example, for the rate of a dual fault we have:

$$R = R_1 \cdot R_2 \cdot (MFD_1 + MFD_2) \tag{8}$$

The mean fault duration is given by

$$MFD = \left( \frac{1}{MFD_1} + \frac{1}{MFD_2} \right) \tag{9}$$

## FAULT BIAS MAGNITUDE AND TEMPORAL PROFILE

This is a critical part of the test, because the response to the faults is very dependent on the relative magnitude of the fault biases (especially when considering multiple faults, as in a constellation-wide fault). Since the goal of the test is to stress the ARAIM solution and PL, we want to inject a fault that is worst-case in a certain sense. A method to define a near-worst case was described in [7]. This method defines fault magnitudes that minimize the norm of the biases in the parity space while keeping the error in the weakest direction constant.

However, as opposed to the fixed bias proposed in [7], the test will use a ramp as a scaling factor. The final value is set to result in 1.5 times the PL. It was determined that a ramp provided a fault that could be near-worst case for a larger set of possible algorithm designs.

### FAULT DURATION

As mentioned above, each trial represents an exposure time of $T_{EXP}$=1 hour. This does not mean that the fault is active during the whole exposure window. Instead, we must consider that the fault could have started before the window. For this reason, we will assume that the length of the fault $T_F$ is given by the average overlap between the exposure window and the fault duration MFD.

$$T_F = \left( \frac{1}{MFD} + \frac{1}{T_{EXP}} \right)^{-1} \tag{10}$$

### NOMINAL ERROR MODEL

The tests are run by simulating Gaussian noise corresponding to the dual frequency ionospheric free error models. These error models are fully specified in the draft DFMC MOPS [10]. The errors are generated by a first order Gauss-Markov process with a 4 s time constant. The time constant was chosen to be close to a worst case in the sense that:

- It is high enough to lead to multiple effectively independent exposures during the exposure window
- it is not high enough to lead to many exposures within the time to alert, which would allow the user to reduce the noise in the detection statistics in a way that would not be available in a real scenario

### NOMINAL BIASES

The ISD includes an upper bound on nominal biases, defined by $b_{nom}$. These are biases that potentially affect all measurements. In the default ISD, these biases are set to zero, but they could be set to larger values in future broadcast ISDs (through the Integrity Support Message). For this reason, the test must ensure that algorithms protect against these biases.

Following the same principle as for the fault biases, we define a set of biases that maximizes the impact on the user error, at least in an approximate way. To achieve this property, we set:

$$b_{nom,i} = sign\left(s_i^{(k)}\right) \cdot b_{nom,max,i} \tag{11}$$

where

$s_i^{(k)}$ are the coefficients of the subset solution estimator [7] in the coordinate that maximizes the uncertainty

$b_{nom,max,i}$ is the maximum magnitude of nominal biases (as specified in the ISD)

The rationales for this choice is given in Appendix B.

**LOWER BOUND ON PROTECTION LEVEL**

For any estimator of the position *f* and region Ω for the measurement residuals such that

$$\max_{b_i \in R^{q_i}} P\left(\left|f(y) - e^T x\right| > L, y \in \Omega\right) \le \mu$$

(integrity requirement) and

$$P\left(y \notin \Omega\right) \le \alpha$$

We must have

$$\sqrt{2} Q^{-1}\left(\sqrt{\frac{\mu + \alpha}{\max\left(p_i, p_j\right)}}\right)\sigma_{ij} \le L \tag{12}$$

where

$\sigma_{ij}$ is the standard deviation of the least squares solution that is fault tolerant to the fault hypothesis *i* and *j* [8].

$p_i$ is the state probability of fault mode *i*

It is significant, (and perhaps counterintuitive) that the lower achievable PL is related to the filter fault tolerant to the combination of two faults rather than the fault tolerant filter to one fault only.

This lower bound can be used as a sanity check on the guaranteed protection level after a potential exclusion. If the algorithm outputs a PL that is smaller than the bound in Eq. (12), it is likely to be too low to guarantee integrity.

**REFERENCES**


[1] Working Group C, ARAIM Technical Subgroup, Milestone 3 Report, February 26, 2016.

[2] Working Group C, ARAIM Technical Subgroup, Milestone 2 Report, Issue 1.0, February 11, 2015.

[3] Blanch, Juan, Walter, Todd, Enge, Per, Burns, Jason, Mabilleau, Mikael, Martini, Ilaria, Boyero, Juan Pablo, Berz, Gerhard, "A Proposed Concept of Operations for Advanced Receiver Autonomous Integrity Monitoring," *Proceedings of the 31st International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2018)*, Miami, Florida, September 2018, pp. 1084-1090.

[4] ARAIM CONOPS, posted in ICAO NSP as JWGs/3 WP38 available upon demand

[5] Draft of ARAIM SARPS, ICAO NSP Working Paper, posted as JWGs/4 IP32

[6] Blanch et al. "Development of the ARAIM MOPS" *Proceedings of the 32nd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2019)*, Miami, Florida, September 2019

[7] Blanch, Juan, Walter, Todd, "Stress Testing Advanced RAIM Airborne Algorithms," *Proceedings of the 2020 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2020, pp. 421-439. https://doi.org/10.33012/2020.17153

[8] Blanch, Juan, Lai, Yu-Fang, Walter, Todd, "Integrity-Based Estimators for Precise Point Positioning and Other Applications," *Proceedings of the ION 2024 Pacific PNT Meeting*, Honolulu, Hawaii, April 2024, pp. 477-490. https://doi.org/10.33012/2024.19670

[9] Milner, Carl, Pervan, Boris, Blanch, Juan, Joerger, Mathieu, "Evaluating Integrity and Continuity Over Time in Advanced RAIM," 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), Portland, Oregon, April 2020, pp. 502-514.

[10] EUROCAE ED-259A / RTCA DO-401, Minimum Operational Performance Standard For Dual-Frequency Multi-Constellation Satellite-Based Augmentation System Airborne Equipment, January 2024

**ACKNOWLEDGEMENTS**

This work was funded by the the FAA Satellite Navigation Team.

**APPENDIX A: NUMBER OF TRIALS**

We want to verify a probability *p* with *n* trials for a Boolean outcome (success of failure). The probability *p* is probability of failure. The *n* trials can be considered independent events.

*Probability of not passing the test given that the probability is p is correct*

The total number of failures X follows a binomial distribution $B(n, p)$, where *p* is the probability of failing a single trial and *n* is the number of independent trials. The probability that there are exactly *k* failures is given by

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The mean and standard deviation of this distribution are given by

$$E(X) = n \cdot p$$

$$var(X) = n \cdot p$$

For large values of n, the CDF of this probability, i.e., the probability of having less than and equal to *k* failures, can be well approximated by the Gaussian distribution $N\left(n \cdot p, \sqrt{n \cdot p}\right)$. For a test having *n* trials and passing threshold of *T*, the probability of passing the test is given by:

$$P(X \leq T) = Q\left(\frac{E(X) - T}{std(X)}\right)$$

Let us note $P_{fa}$ the required probability that the test does not pass given that the probability of failing a single trial is *p*. We must have

$$P(X \leq T) \geq 1 - P_{fa}$$

Using the Gaussian approximation, this means that

$$P(X \leq T) = Q\left(\frac{E(X) - T}{std(X)}\right) = Q\left(\frac{n \cdot p - T}{\sqrt{n \cdot p}}\right) \geq 1 - P_{fa}$$

$$T \geq n \cdot p + Q^{-1}\left(P_{fa}\right)\sqrt{n \cdot p} \qquad \text{(A.1)}$$

*Probability of passing the test given that the probability is* $2 \cdot p$

Let us note $P_{md}$ the probability of passing the test given that the probability of failing a single trial is $2 \cdot p$. We want

$$P(X \le T) \le P_{md}$$

The Gaussian approximation yields

$$Q\left(\frac{E(X)-T}{std(X)}\right) = Q\left(\frac{2 \cdot n \cdot p - T}{\sqrt{2 \cdot n \cdot p}}\right) \le P_{md}$$

This is equivalent to

$$\frac{2 \cdot n \cdot p - T}{\sqrt{2 \cdot n \cdot p}} \ge Q^{-1}(P_{md})$$

Or

$$2 \cdot n \cdot p - Q^{-1}(P_{md})\sqrt{2 \cdot n \cdot p} \ge T \tag{A.2}$$

Combining (A.1) and (A.2), we get the condition

$$2 \cdot n \cdot p - Q^{-1}(P_{md})\sqrt{2 \cdot n \cdot p} \ge n \cdot p + Q^{-1}(P_{fa})\sqrt{n \cdot p}$$

Or

$$n \ge \frac{\left(\sqrt{2}Q^{-1}(P_{md}) + Q^{-1}(P_{fa})\right)^2}{p} \tag{A.3}$$

The minimum number of samples is

$$n = \left\lceil \frac{\left(\sqrt{2}Q^{-1}(P_{md}) + Q^{-1}(P_{fa})\right)^2}{p} \right\rceil \tag{A.4}$$

The associated detection threshold is

$$\begin{aligned} T &= \left(\sqrt{2}Q^{-1}(P_{md}) + Q^{-1}(P_{fa})\right)^2 + Q^{-1}(P_{fa})\left(\sqrt{2}Q^{-1}(P_{md}) + Q^{-1}(P_{fa})\right) \\ &= \left(\sqrt{2}Q^{-1}(P_{md}) + Q^{-1}(P_{fa})\right)\left(\sqrt{2}Q^{-1}(P_{md}) + 2Q^{-1}(P_{fa})\right) \end{aligned} \tag{A.5}$$

## APPENDIX B. RATIONALE FOR $B_{NOM}$

The rationale for $b_{nom}$ is based on maximizing the probability of missed detection using a solution separation test.

The probability of missed detection is given by

$$P\left(\left|s^T(\varepsilon + b_{nom}) + b_f - s^{(k)T}(\varepsilon + b_{nom})\right| \le T, \left|s^T(\varepsilon + b_{nom}) + b_f\right| > L\right) \qquad \text{(B.1)}$$

where

$\varepsilon$ is the nominal error before adding nominal biases

$b_f$ is the injected fault in the position domain (one coordinate)

$b_{nom}$ is the vector of nominal biases

*T* is the detection threshold

*L* is the error bound (the PL)

We want to maximize the expression

$$\max_{b_f} P\left(\left|s^T(\varepsilon + b_{nom}) + b_f - s^{(k)T}(\varepsilon + b_{nom})\right| \le T, \left|s^T(\varepsilon + b_{nom}) + b_f\right| > L\right) \qquad \text{(B.2)}$$

Note that the maximum here is over the fault bias $b_f$.

We perform the change of variables $s^T b_{nom} + b_f \rightarrow b_f$. So that expression (B.2) becomes

$$\max_{b_f} P\left(\left|\left(s - s^{(k)}\right)^T \varepsilon + b_f - s^{(k)T} b_{nom}\right| \le T, \left|s^T \varepsilon + b_f\right| > L\right) \qquad \text{(B.3)}$$

Under this parametrization, the bias affecting the test statistic is $b_f - s^{(k)T} b_{nom}$ (for a bias $b_f$ affecting the position). The worst case probability will then be obtained by minimizing $b_f - s^{(k)T} b_{nom}$. For a positive $b_f$ (and sufficiently large), this will occur for:

$$b_{nom,i} = sign\left(s_i^{(k)}\right) \cdot b_{nom,max,i} \qquad \text{(B.4)}$$