INTEGRITY TESTING for GNSS SOLE MEANS

James L. Farrell

VIGIL, Inc.

Severna Pk, MD

Frank van Graas

Avionics Engineering Center

Ohio University

Athens Ohio

ABSTRACT

Within Working Group #5 of RTCA SC-159, a test plan has emerged for validating GNSS (GPS and GLONASS) receivers. The plan makes maximum usage of the approach documented for Supplemental Navigation, but Sole Means Navigation imposes additional factors to be addressed (*e.g.*, fault detection, exclusion decisions, and exclusion decision reset capability). Further differences arise from concentration on end-to-end testing and from firmly established confidence levels, prompted by uncertainty in probabilistic parameters themselves – a standard concept in statistics.

Pass/Fail criteria will be GO/NO-GO (as in Ref. 1) but, in addition, immediate rejection will result from "blind regions" (which have arisen in the past) or catastrophic errors. Sole Means tests are to be performed with dynamics relevant to the phase being certified. Probability scaling will enable realization of 99% confidence, even with less than 2000 total test runs. Runs will start near each location in Ref. 1, with the addition of a random displacement from nominal location, precluding usage of test coordinates that are known and/or integer-valued.

To aim toward 'no-hints' testing conditions, alarm and detection tests will be interwoven. Early detections will have restricted acceptance, and biases can be steps as well as ramps, appearing at unknown times (enhancing test efficiency, since pre-bias detection test periods qualify as bona fide alarm testing) at levels chosen to avoid inconclusive runs. This test plan description shows how ensembles will be generated, and validates the 99% confidence to be achieved.

_____

INTRODUCTION

Material herein was first discussed in mid-1993, and has since undergone extensive revisions – in response to written and verbal comments (gratefully acknowledged). Parameters herein are illustrative values, not immutable. This base line test procedure is not optimized (*e.g.*, replacement by sequential decisions could yield substantially shorter average on-line test durations without sacrificing performance).

As a guiding philosophy toward preparation of the test plan, paramount importance was attached to flight safety. At the same time, every effort was made to ensure that equipment would not be required to perform unrealistic tasks. The goal is a test plan that would not require the equipment to show more – *nor less* – capability than that needed in flight. In addition, it is in the best interest of SC159 that the plan be above reproach, so that other groups cannot find flaws. Beyond those overriding goals, the final plan will reflect the collective will of SC159. An effort was made to observe that collective will as now perceived. For example, GO/NO-GO PASS/FAIL criteria will be used throughout, based only on navigation solutions and external flags.

A central feature of this methodology is off-line test data generation, not for storage on disk or other media but for experiment design that can be fully substantiated, fully disclosed, and fully documented (*e.g.*, in libraries of test case error histories). All of the techniques cited here for SV error history generation represent readily verifiable public domain information, not subject to proprietary rights or other limitations. Responsible government agencies can designate any desired institution for generation and maintenance of this information.

Present plans are for all Sole Means tests to use *rf* signal inputs, with no off-line tests for only software. Signals will reflect dynamics for phases to be certified, including velocity, acceleration and lever arm effects, but not obstruction, nonrigidity, or directivity loss. To ensure that FDE capability is exercised, the meaning of marginal geometry has been broadened somewhat, and the biased SV will not necessarily always be the key satellite. Since error histories are to be generated off line, test cases can be run in any order (not necessarily the order in which they were computed).

Another feature adopted for Sole Means test is probability scaling, to allow realization of 99% confidence while still providing practical test set length (fewer test trials than in Ref. 1). As a result of this scaling step, alarm and missed detection events have large enough sample sizes so that no second or third chance is needed. In these tests, any catastrophic errors, blunders, or "blind regions" are grounds for immediate rejection.

The high confidence is obtained through higher threshold levels, which causes a modest sacrifice in availability. While any availability reduction is unfortunate, two observations are appropriate at this point:
- The loss is not as severe as might be expected by those who are unfamiliar with this procedure. The explanation lies in how the gaussian function hugs the asymptotic value (zero) at the tails; lower nominal threshold corresponds to larger normalized variance, spreading more underlying area – always normalized to unity – to regions beyond ± threshold that were almost empty.
- The gaussian property just cited is a blessing for test but a potential curse for operation: If satellite ranging errors are increased beyond the assumed 33.3-m RMS level, the resulting rise in internal alarm rates could compromise continuity and availability. For example, a 25 % increase would raise the internal alarm rate by two orders of magnitude, from 10^{-5} to 10^{-3}.

Finally, if altimeter aiding should be added, straightforward weighting methods already known open the door for generalizing this test approach to other augmentations.

APPROACH

The level of testing defined herein primarily concerns availability (observed from external alarm rate) and missed detection probability. The intent is to demonstrate high confidence (on the order of 99%) that these are within prescribed limits. By itself that conclusion does not prove satisfaction of overall requirements (*e.g.*, tunnel RNP, continuity). In combination with the rationale that links overall requirements to the prescribed limits just cited and to a 0.3 nmi containment warning level (CWL), however, the tests are adequate; thus no duplication of effort is implicit in this test plan.

Integrity performance will be judged from GO/NO-GO criteria throughout, based only on navigation solutions and <u>external</u> flags (*e.g.*, internal flags that may have affected the choice of SVs forming that solution will not influence the test scoring). The integrity alarm will be validated by multiple sets of test runs (described here in terms of verification of alarm, detection, and admissibility performance). Each run will consist of a sequence up to 10 minutes in duration wherein solutions are computed repeatedly at regular intervals in the presence of randomly generated measurement errors, *i.e.*, each test run is a Monte Carlo simulation trial. Each set of run sequences (*e.g.*, 1000 runs for ramp bias detection and 50 runs for step bias detection) will be referred to herein as a series. Test sequencing will not be limited to consecutive ordering. When trials can be run in random order, *i.e.*, with false alarm and detection trials arbitrarily intermingled, overall test chronology could contain a false alarm run followed by two detection runs followed by another false alarm run. This practice is called interweaving; the tests are <u>interwoven</u> (intertwined) randomly. In this scheme it would not be known during any individual run which type of test is currently being performed – just as an observed alarm cannot be ruled as surely true or surely false in flight. Further interweaving can also occur with tests for <u>inadmissibility</u> (*i.e.*, for space-time points having geometries that do not satisfy the algorithm's criteria for reliable failure detection in the phase of flight being certified). With no hint as to which case or which parameter set is instantaneously active, then, the equipment and its integrity algorithms would still have to satisfy all requirements.

It is acknowledged that interweaving imposes a cost burden, due to added test setup time. For this reason the possibility is allowed herein that "interwoven" sequences actually used for formal certification might include simplified ordering, $e.g.$, with chronologically arranged start times. Nevertheless, whether test series are run in consecutive or random order, the requirement for <u>capability</u> of satisfying randomly interwoven tests remains. That capability is implicit in any claim that certification tests have been performed successfully. With that understanding the word "interweave" and all its derivatives will hereafter imply the general interpretation, to include both random and systematic time orderings.

For test efficiency it is essential to keep the requisite number of runs at a practical level and, for validity, a high confidence for all conclusions to be drawn. These conflicting needs are met through probability scaling, whereby input noise levels are amplified to increase the expected number of events ($e.g.$, false alarms for availability runs; missed detections for detection runs) to be observed in test.

The amount of noise level amplification is developed in the APPENDIX, along with explanations of pertinent concepts. The APPENDIX then develops, in addition, a rationale for setting bias levels. For purposes of immediate discussion it suffices to state that, for biased SVs in variable geometries, nonstationarity is inherently present. The influence of that nonstationarity is minimized, while occurrence of both inconclusive runs ($e.g.$, biases causing neither CWL violation nor alarms) and runs of trivial significance ($e.g.$, detection of biases whose effects would be almost impossible to miss) can be minimized or eliminated, by proper experiment design as shown in the APPENDIX. Finally in connection with the "no‑hints" testing concept, biases will not occur at any consistent time. Aside from those intervals close to the end of a run (within the allowable time-to-alarm), biases can be initiated at essentially any time within a test run – just as they can occur at any unexpected time in flight. From this it clearly follows that alarms occurring before bias initiation are no different from false alarms during admissibility testing; their occurrence in detection tests must be added to the alarms observed during alarm runs.

To stress the equipment somewhat ($e.g.$, by simulating outages), the rf signal will often contain only a subset of visible SVs. Many runs will be conducted with only five or six. By responding to alarms only if they are external, these tests facilitate passage of availability checking when there are more than 5 SVs. If all availability testing were done with only 5 SVs, successful results would provide 99% confidence for alarm rate satisfaction under maximally stressful conditions. Though less stressful, these tests are still conservative; availability runs will have an abnormally high fraction of 5-SV cases.

Each test run will last until either an alarm occurs or a ten-minute interval (real time) has elapsed. For a good receiver ($i.e.$, one not causing premature cessation by failing early), total active time for all required tests is then on the order of 300 hours; to that must be added more time for repetitive initialization and other overhead tasks. The price is admittedly significant, but by no means unprecedented for avionics test – and not inappropriate when considering the critical importance of flight safety.

Tests described here apply to equipments employing a snapshot-type RAIM integrity algorithm that uses pseudorange measurements from GPS satellites either singly or in conjunction with signals from other satellites including GLONASS. Alternative algorithms ($e.g.$, not limited to snapshot) developed by manufacturers are of course required to satisfy all requirements, but are beyond the scope of this test plan. Manufacturers of equipment employing alternative algorithms would then develop alternative procedures, to demonstrate achievement of the same objectives defined herein.

SIMULATION TESTS

Throughout all descriptions herein it should be kept in mind that interweaving applies to all conditions under test, including 1000 runs with ramp biases (from which the pre-bias portions comprise about half of the admissibility test time, intrinsically interwoven), 50 runs with step biases, another 500 runs with no bias (bringing the total admissibility test duration to 10,000 minutes), plus 50 inadmissible geometries, forming an overall ensemble of 1600 runs. It is reiterated here that half of the active test time for admissibility is supplied by the pre-bias portion of the detection test runs.

Receiver / SV Geometry

Nominal geographic locations chosen in Ref. 1 will be used for Sole Means tests, with one departure: Coordinates listed are nominal only; to preclude restriction to initial positions that are known or limited to integer values, each test run will begin with an independent gaussian randomly selected vector (zero mean, 10 arc-second per coordinate standard deviation) displacement from the location shown. Start times, controlled by test series selection, can be used as a means of controlling the experiment design.

The GPS satellite configuration to be used in Sole Means testing will match the Optimized GPS-24 Satellite Constellation (See Ref. 2), with epoch at July 1, 1993. The GLONASS constellation will conform to Ref. 3.

Twenty alarm and forty detection sample runs of up to[§] ten minutes duration will be initiated for each location, with starting times controlled by test series selection. Start times do not all have to be distinct, *e.g.*, a pair of test runs can be identical except for their random number sequences. Whenever random selection yields insufficient SV's in view, start time will be reselected by another random draw. Conversely, when random selection yields HDOP values too low to be marginal, geometries can be made more challenging by withholding certain (*e.g.*, lower-elevation) SV signals from the *rf* input; this also facilitates maintaining marginal geometry through 10 minutes of continuous motion by receiver and all SVs. Points having inadmissible geometry will likewise be replaced but, as in Ref. 1, the best fifty of these will be set aside for inadmissibility testing (*i.e.*, the fifty that are closest to admissible; steps similar to those just described can ensure that at least fifty will be generated). There will be a total of 1000 space-time sample points representing initial conditions for dynamic test runs. Since each run - whether an alarm or a detection trial - has up to 10 minutes duration and independent samples have two-minute spacing, all SA samples will be independent. There will be 5000 total SA samples for availability testing.

_____

[§] An alarm can prematurely terminate a run.

Pseudorange Errors to be Simulated

For GPS signals, SA effects will be simulated as the sum of a zero mean normally distributed random constant with standard deviation of 23 m and a 2nd order Gauss-Markov process, 23 m RMS with 120-second autocorrelation time. Thus, although receiver outputs can be sampled at closely-spaced intervals (*e.g.*, at 1 Hz or more), the effective sampling rate for sequentially independent SA errors will be once every two minutes. The 23 m levels will be scaled to amplify expected values of missed detections and alarms in test. SA processes on all satellites will be statistically independent.

Because SA dominates the overall error budget, no effort is made to model other degradations (*e.g.*, inexactness of ephemerides or propagation correction) in simulated pseudorange errors. Effects of tracking-loop noise and all other receiver errors will be naturally included as *rf* signals are fed to the equipment under test.

A satellite malfunction will be simulated as a range error (50 step runs and 1000 ramp runs). Activation will be controlled by random number generation, with parameter setting (*e.g.*, ramp slope) as described in the APPENDIX. Random number generation will be controllable via starter values for specific generating routines, allowing a wide variety of repeatable test sequences that could be documented as test case libraries (libraries could also be generated and collected for flight path scenarios of interest).

Detection Test Runs

The procedures that follow provide extensive testing of poor detection geometries in non-precision approach (parameters will be adjusted for certification limited to other phases). Although correctness of FDE decisions will be evidenced by correct flag states only, this does not imply weakness of available evidence. False exclusions (occurring after FA) will hamper subsequent availability while wrong exclusions (occurring after true alarms) will do that and also cause more false alarms. More conclusively, acceptable nav error (*i.e.*, below CWL) in the presence of unacceptable bias (*i.e.*, sufficient to cause CWL violation) clearly indicates absence of the biased SV from the solution. When this occurs with only six SV's in view, evidence of an exclusion is complete.

Each test result will be compared vs. the outcome of a corresponding test using the base line algorithm, with the conditions
- these outcomes will be stored in a data base available to the test equipment,
- only the totals must meet or exceed base line performance; *e.g.*, the equipment can miss one detection that the base line caught and catch one that the base line missed,
- adequate availability of the base line is being verified independently.

Thus the plan is complete and can accommodate a general class of FDE algorithms.

Although interweaving holds throughout, the next section will address only the test cases with ramp bias (*e.g.*, step bias, alarm, and inadmissibility tests added to the interwoven ensemble are addressed separately).

Ramp Bias Runs
The 1000 space-time points will first be rank ordered, as in Ref. 1, at their initial locations. Also as in that reference, the best of these most difficult space-time points will be used in the fifty "inadmissibility" test cases, for which the integrity alarm must be annunciated, and remaining points are referred to as the "admissible geometries." Next, the least desirable detection geometries (*e.g.*, ten or twenty least desirable) from the list of admissible geometries are to be selected (the "marginal" geometries), for use with ramp bias simulations within the admissible cases. Though geometry variations inherently produce nonstationarity in detection statistics, the effect is minimized through bias level control (*e.g.*, an easier geometry with small bias can be either more challenging or less challenging than a more difficult geometry with larger bias; see APPENDIX).

For the detection sequences, a ramp bias will be introduced into one SV. In these 1000 ramp bias runs, the bias will be introduced into the most difficult-to-detect satellite unless the missed detection probability – as shown in the Appendix – exceeds 0.001 (in which case the second most difficult-to-detect SV will contain the bias). A Monte Carlo random trial run will be made with the bias initiated beyond the chosen space-time point. The run will continue until one of the following events occurs:

(1) The alarm is triggered before the bias is initiated (an "unacceptably early detection").
(2) The alarm is triggered after bias initiation but before containment warning level (CWL) is exceeded (an "acceptably early detection").
(3) The CWL is exceeded and the alarm is subsequently triggered within allowable time-to-alarm (a "timely detection").
(4) The CWL is exceeded but the alarm is not triggered within the specified time-to-alarm (a "miss").
(5) A ten-minute run ends with neither violation of the CWL nor an alarm ("inconclusive run").

The Monte Carlo run just described will be repeated for the total ensemble of 1000 trials, covering the full set of space-time points under consideration. The 1000-trial ensemble will include the different sample realizations of the various random processes and the deterministic pseudorange errors previously described.

Analysis of Ramp Bias Detection Results
Although all tests can be interwoven, conditions applicable to each run are of course available to the test apparatus; alarm, detection, and inadmissibility results will be separated for evaluation. Results of the 1000 Monte Carlo runs will be interpreted as follows: Because input random noise levels are to be scaled to produce an expected value of missed detections amplified by a factor of thirty (*i.e.*, to produce a probability of 0.03 instead of 0.001 for nominally set thresholds), an experimental count of 19 misses would allow the experimenter to state, with confidence near 0.99, that the unknown <u>actual</u> scaled missed detection probability is less than 0.03 – and thus the unknown <u>un</u>scaled missed detection probability with normal SA error levels would be less than 0.001. Therefore: Determine the total of (a) misses, (b) timely detections, (c) acceptably early detections, (d) unacceptably early detections, and (e) inconclusive runs. (The sum of a, b, c, d, and e must be 1000. The sum of a, b, and c will be denoted here as <u>S</u>; this specification is accurate if <u>S</u> is at least 990). In order for the equipment to pass the detection probability simulation test,
- The total number of misses must not exceed $19 \times S / 1000$, and
- External flags described in item (d) above are recognized as alarms appearing with no SV bias. All such improper external flags will thus be added to the alarm count.

Step Bias Runs

From the 1000 runs used for ramp bias simulation, fifty will be selected and subjected to the interweaving process. Steps can be inserted at virtually any instant after start time (*i.e.*, until ten minutes minus the time-to-alarm). Acceptance of a run will be contingent on the missed detection probability computation used for ramp tests, and the bias level control (APPENDIX) will ensure that retention of the failed SV causes CWL violation. Thus each run will produce either an unacceptably early detection, a timely detection, or a missed detection.

Analysis of Step Bias Detection Results

Step bias runs will be extracted from the interwoven ensemble for separate evaluation. Results of the fifty step bias runs will be interpreted thus: there can be no missed detections, and the number of unacceptably early detections must be added to the alarm count test described in the next section.

Alarm Rate Tests

Again, although alarm and detection sets will be interwoven, conditions applicable to each test run are of course available to the test apparatus; alarm test results will be separated for evaluation. Also, input random noise levels will be scaled to produce an amplified alarm probability of 0.030 for nominally set thresholds. An experimental count of 124 alarms would allow the experimenter to state, with 99% confidence, that the unknown actual scaled alarm probability is less than 0.030 – and thus the unknown unscaled alarm probability with normal SA error levels would be less than its allowable value.

The specified internal alarm rate is 10^{-5} alarms per sample, and it is assumed that independent samples are available every 2 minutes. In 10,000 minutes there are effectively 5000 alarm snapshots; this determines the scaling needed to produce the 0.03 alarm probability. For 99% confidence, allowable alarm count observed in test can be no greater than 124 (instead of 5000 × 0.03 = 150). Note that, during Sole Means alarm rate tests, it is not assumed that alarms due to bona fide, unannounced satellite malfunctions are "extremely rare" (as was assumed for Ref. 1); Sole Means alarm tests will be interwoven with detection tests.

NPA Simulations with No SV Failures

From the overall set of 1600 runs (which, it is recalled, include the marginal locations, with and without SV bias, plus inadmissible geometries), only the alarm runs with admissible geometries plus pre-bias segments of the detection runs are to be extracted for immediate consideration. Effectively these constitute 5000 snapshot random trials with no satellite failures. The combined effect of receiver and scaled SA noise will be assumed normally distributed with zero mean. Different noise samples will be used for each satellite being used in the test, and new sets of noise samples will be used for each trial at each admissible space-time point. In order to pass the alarm rate test, the following criteria must be met: The total number of external alarms for all admissible points, plus the number of improper external flags from the Detection Test, must not exceed 124 and, to assure no unusual bunching of alarms at any one space-time point, the number of external alarms for any one space-time point cannot exceed one.

The number of statistical samples in this test is 5000. The criteria for passing the test have been set such that equipment which satisfies the alarm rate test limit on one try will have a 99 percent confidence that the unknown actual scaled alarm probability is less than 0.030 – and thus the unknown unscaled alarm probability with normal SA error levels would be less than its allowable value. If the equipment fails this alarm test, then it must be modified to correct the alarm rate problem before re-testing.

It should not be concluded that passage of this availability test will be at all difficult. Only external alarms are observable and, in many of the test cases, the *rf* test signal will contain data from more than five SVs.

Inadmissibility Tests

The bench tests just described were devised to test for a variety of conditions using admissible geometries. Interwoven among those cases will be tests run for the phase of flight being certified, using inadmissible geometries. The purpose of these tests is to verify that the alarm is annunciated whenever an inadmissible set of geometries is presented to the receiver.

As described earlier, 50 inadmissible geometries are chosen for the phase of flight being certified, each being at the top of the ordering for that phase; *i.e.*, the best fifty within the respective inadmissible group. Simulator signals corresponding to each inadmissible geometry, containing the usual simulated SA noise but no bias, will be fed to the receiver. After the initial acquisition period and when the receiver is in the navigation mode, an integrity alarm is expected to be annunciated indicating that a valid assessment of system integrity cannot be made. The alarm must be annunciated within time-to-alarm for all 50 of these tests.

Test Setup and Conditions

History of simulated velocity vector and altitude will be typical of the phase of flight being tested. CWL violation via ramp ascension or step bias can occur at virtually any time, up to ten seconds before run completion (more generally, preceding the run completion by time-to-alarm). Aircraft motion is to be properly accounted for throughout the progress of the run. The GPS simulator will place the receiver at any of the specified locations, with the receiver equipment in the mode being certified.

In addition to simulated horizontal and vertical velocity, aircraft dynamics can include maneuver acceleration, speed variations, rotations, and lever arm kinematics in the presence of those rotations. Additional variations in conditions not now planned for these tests are deformations due to structural nonrigidity, intermittent SV sightline obstruction from aircraft surfaces, low SNR, and reduced antenna directivity.

Settings for Input Noise Scaling and Biases

Based on the RSS of two 23-meter gaussian inputs it is anticipated that designers can set thresholds sufficiently high to provide the 99% confidence coefficient – or higher (see APPENDIX). The amount of scaled input noise for alarm tests will enable the high-confidence to be reached with enlarged thresholds. It is recognized that this higher threshold will be a slightly inhibiting influence on detectability of bona-fide alarms – and <u>that</u> effect will be taken into account in defining the amount of simulated input bias for detection tests. Bias levels will be individually chosen to minimize both trivial and inconclusive runs (see APPENDIX).

Retest

It is recognized that the bench tests involve a finite, though – with the scaling – adequate, number of statistical samples. It is possible (although unlikely) for properly functioning equipment to fail the bench tests simply due to chance alone. Nevertheless, if the equipment being tested fails either the alarm or detection test, or any inadmissible geometry test trial, equipment must be modified to correct the problem before re-testing for Sole Means.

Regardless of performance in all tests described above, immediate rejection will occur for any equipment exhibiting any blind locations, blunders, or catastrophic errors.

CONCLUSION

A Sole Means integrity test plan has been described, for verification of GNSS receivers' fault detection/exclusion (FDE) and availability performance. All tests are end-to-end with *rf* signal inputs, and outputs of navigation solutions with external flags only. Tests will be conducted under 'no-hints' conditions, with initial locations known only nominally, and with step or ramp biases that can appear at any time; test runs with and without biases will be interwoven at random, so that it will not be known which case is instantaneously active. Some visible SV signals will be absent from the *rf* input.

Tests are devised to provide, under conservative conditions, 99% confidence – substantiated by analysis herein, based on accepted statistical practice. Mathematical support for scaling, RMS input error levels in test, illustrative threshold settings, bias level settings, and parity / χ^2 correspondence appears in the APPENDIX.

Failure to pass any part of the test combination, or occurrence of any catastrophic error during test, will render the equipment unsafe for Sole Means navigation in the National Airspace.

REFERENCES

1. "Minimum Operational Performance Standards for Airborne Supplemental Navigation Equipment Using Global Positioning System (GPS)", RTCA/DO-208, July 1991.
2. RTCA Paper No. 316-93/SC159-453 .
3. RTCA Paper No. 294-93/SC159-451 .

APPENDIX

Setting Parameters for 99% Confidence

In 10,000 minutes of bias-free test time there are 5000 opportunities for an alarm to be triggered by SA with 2-minute correlation time. Thus, with SA input noise scaled upward to provide a single run alarm probability of 0.03, a large group of 5000-sample test series would yield an average of (5000)•(0.03) = 150 alarms per series. If one series in fact yielded 150 alarms then, by definition of an average, the experimenter would be 50% confident that the actual (unknown) alarm rate over an infinite number of tests would not exceed 0.03. Because the scaling derived below depends directly on the two alarm rates (10^{-5} and 0.03), there would also be 50% confidence that the actual *un*scaled alarm rate would not exceed 10^{-5}.

Immediately the following question arises: If a set of availability tests produced k alarms, then the experimenter could be 99% confident that the actual unknown alarm rate would not exceed what level? The answer is based on the last equation of Ref. 1; if the binomial/gaussian ratio can be replaced by unity for large abscissa values (admittedly optimistic), the equation would then yield a scaled alarm rate of

$$P_C = \frac{k + 2.32635 \sqrt{k\,(1 - k/5000)}}{5000}$$

Since this relation is based on a gaussian approximation to the binomial with mean value k and variance $k(1 - k/5000)$, it is accurate only for the larger abscissa values (*e.g.*, beyond ten). The optimistic simplification just mentioned could be undone by a small percentage reduction in the number of observed alarms allowed. The equation was programmed; the accompanying plot shows the results.
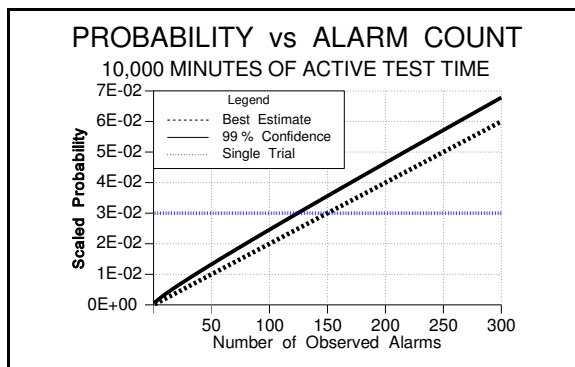

PROBABILITY vs ALARM COUNT
10,000 MINUTES OF ACTIVE TEST TIME

Figure 1: Scaled Alarm Probability

It is seen that essentially 124 alarms can be permitted (instead of 150). Actually this allowable alarm count is imprecise, for two reasons:

• the aforementioned gaussian approximation – *i.e.*, absence of a multiplying factor based on Fig. 3 of Ref. 1. Fortunately, at an abscissa of 124, the effect is slight.

• the denominator *5000* in the equation just shown. If alarms can cause a test run to end prematurely in 124 cases, and if the occurrence of CWL violation is equally likely for all bias-free Monte Carlo samples, the number would be short by 310, since
$[5 \bullet (1000 - 124) + 2.5\,(124) \; = \; 4690]$.

This latter item will now be offset by adjustment: Recall the intrinsically interwoven alarm test samples coming from detection tests. Thus far it has been loosely implied that about 2500 samples (half of the total) would come from that source, consistent with the practice of counting pre-bias detections as alarms. A more detailed look now prescribes the number thus: Rather than two halves of the requisite 5000 coming from 500 availability runs and 1000 ramp tests, there is a needed increase of 310 (as just explained), and also a reduction by the samples contributed from half of $5 \bullet (50)$, the pre-bias time from step input runs. The result, 2685, prescribes an average pre-bias fraction of $2685 / 5000 = 0.537$, slightly more than halfway through the ramp runs.

A similar program, for missed detections, produces an allowable miss count of essentially 19 (instead of 1000 • 0.03 = 30) from the following plot:
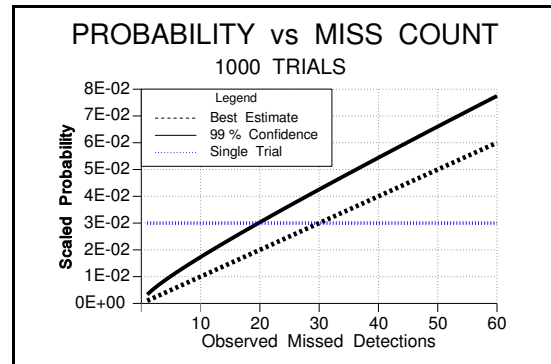

PROBABILITY vs MISS COUNT
1000 TRIALS

Figure 2: Scaled Missed Detection Probability

Note that the discussion after the alarm rate plot is inapplicable here; an allowable missed detection count of 19 is not permissive, and additional chances for missed detections from other test runs need not be considered. The dominant error (SA) causing CWL violation to be missed will undergo little change during time-to-alarm – and the value of the test statistic at that instant (plus a brief interval thereafter, *e.g.*, 10 sec) determines the outcome. A series of 1000 runs with bias will produce no less than 1000 chances for a missed detection to occur.

It remains to show how a designer, armed with the above information, might set a threshold to achieve passage of the 99% confidence test. To demonstrate that both χ^2 and parity produce the same value for alarms with unbiased SVs: The χ^2 pdf for 1 degree of freedom,

$$f(u) = \frac{1}{\sqrt{2}\ \Gamma(1/2)}\frac{e^{-u/2}}{\sqrt{u}} = \frac{1}{\sqrt{2\pi}}\frac{e^{-u/2}}{\sqrt{u}}$$

corresponds to the square of normalized gaussian random variable ($u = x^2/\sigma^2$), *e.g.*, in Table 1 of Ref. 3, the threshold for 10^{-6} is *23.93*. Before Eq (14) *of* Ref. 4, it is seen that $(1/\sigma) \times ($ Threshold for $10^{-6}) = 3.45892\ \sqrt{2} = \sqrt{23.93}$. Not at all surprisingly, the probability that a normalized variable exceeds that threshold in absolute value is identical to the probability that the square of that variable exceeds the square of that threshold. Therefore, regardless of which method is adopted, the parity analysis can be used here. Thus the nominal threshold is

$$T_{D0} = \sigma\ \sqrt{2}\ \text{erfc}^{-1}\ (10^{-5})$$

so the scaled RMS error used for test will be

$$S_A = T_{D0}\ /\ [\sqrt{2}\ \text{erfc}^{-1}\ (0.03)]$$

To design conservatively (*e.g.*, for 0.02 instead of 0.03), a higher threshold T_{D1} could be used;

$$T_{D1} = S_A\ \sqrt{2}\ \text{erfc}^{-1}\ (0.02)$$

$$= T_{D0}\ \text{erfc}^{-1}\ (0.02)\ /\ [\text{erfc}^{-1}\ (0.03)]$$

corresponding to a 7.2% increase in threshold. With 5000 independent SA samples this gives a mean of (5000) • (0.02) or 100 (instead of 150) and variance $(100) \cdot (1 - 0.02) = (9.9)^2$. Thus the allowable alarm count (124) would be a 2.424-sigma point, giving a probability of about 0.985 for passing the alarm rate requirement –

even if all availability testing were conducted with only 5 SVs. Because much of the testing will be done with 6 or more SVs and because only external alarms are observed, a well-designed receiver should encounter no difficulty whatever.

The foregoing analysis is for one redundant SV. Growth to include more measurements will obviously call for generalization of the expression for the threshold. The base line algorithm shows the requisite modifications: for $m - n$ degrees of freedom the nominal threshold is

$$T_{D0} = \sigma\ \sqrt{2}\ \text{erfc}^{-1}\ \{\ 10^{-5}/(m-n)\ \}$$

and, to scale the RMS error for a total internal alarm probability of 0.03, this is set equal to

$$T_{D0} = S_A\ \sqrt{2}\ \text{erfc}^{-1}\ \{\ 0.03/(m-n)\}$$

so the general relation for scaled RMS error is

$$S_A = \sigma\ \frac{\text{erfc}^{-1}\ [\ 10^{-5}/(m-n)]}{\text{erfc}^{-1}\ [\ 0.03/(m-n)]}$$

and a threshold conservatively set for 0.02 instead of 0.03 is again moderately increased (*e.g.*, by a factor of *1.83/1.72*, producing a *6.4%* increase for two redundant SVs);

$$T_{D1} = T_{D0}\ \frac{\text{erfc}^{-1}\ [\ 0.02/(m-n)]}{\text{erfc}^{-1}\ [\ 0.03/(m-n)]}$$

The price to be paid for the high confidence to be realized in test, then, is a protection radius increase (typically on the order of 12 meters in the example just shown), which entails a modest reduction in availability. While any sacrifice in availability is unfortunate, two observations are appropriate at this point:

- The loss is not as severe as might be expected by those who are unfamiliar with this procedure. The explanation lies in how the gaussian function hugs the asymptotic value (zero) at the tails; lower nominal threshold corresponds to larger normalized variance, spreading more underlying area – always normalized to unity – to regions beyond ± threshold that were almost empty.
- The alternative of low test confidence is untenable. Of all design characteristics that warrant credibility, integrity would seem to command very high priority.

Bias Levels for Test

The preceding portions of this Appendix establish the case for using confidence levels and illustrate how noise levels and thresholds can be set. This section sets bias levels to be

- big enough to cause CWL violation (otherwise a test could trigger false alarms or be inconclusive)

but - not so large that it would be detected even with an inordinately high threshold setting.

To ensure the first of the above conditions, a posteriori navigation error $\delta \mathbf{x}$ is related to measurement errors $\delta \mathbf{y}$ by the usual expression

$$\delta \mathbf{y} = - \mathbf{H} \, \delta \mathbf{x}$$

wherein $\mathbf{y}$ and $\mathbf{x}$ are $m \times 1$ and $n \times 1$ column vectors, respectively, while $\mathbf{H}$ is the $m \times n$ matrix decomposed into factors that are orthogonal ($\mathbf{Q}$) and upper triangular ($\mathbf{R}$); the factors are further partitioned into submatrices $\mathbf{Q}_x$ ($m \times n$), $\mathbf{Q}_p$ ($m \times m-n$), $\mathbf{R}_x$ ($n \times n$), and a null matrix $\mathbf{O}$ ($m-n \times n$);

$$\mathbf{H} = [\mathbf{Q_x} \vdots \mathbf{Q_p}] \begin{bmatrix} \mathbf{R_x} \\ \dots \\ \mathbf{O} \end{bmatrix}$$

Observation errors are subdivided into the $m \times 1$ random error vector $\boldsymbol{\epsilon}$ and a scalar bias b acting on the jth SV; with $\mathbf{1}_j$ representing the jth column of the $m \times m$ identity matrix,

$$\delta \mathbf{y} = \boldsymbol{\epsilon} + \mathbf{1}_j b \; ; \quad \mathbf{Q}_x^{\mathsf{T}} \, \delta \mathbf{y} = \mathbf{Q}_x^{\mathsf{T}} \boldsymbol{\epsilon} + \mathbf{Q}_x^{\mathsf{T}} \mathbf{1}_j b$$

or, since $\mathbf{Q}_x^{\mathsf{T}} \, \delta \mathbf{y} = - \mathbf{R}_x \, \delta \mathbf{x}$ by reason of the matrix partitioning,

$$\delta \mathbf{x} = - \mathbf{R}_x^{-1} \mathbf{Q}_x^{\mathsf{T}} \boldsymbol{\epsilon} - \mathbf{R}_x^{-1} \mathbf{Q}_x^{\mathsf{T}} \mathbf{1}_j b$$

With a matrix $\mathbf{C}$ extracting just the horizontal component $\delta \mathbf{x}_H$ of position error,

$$\delta \mathbf{x}_H = - \mathbf{C} \, \mathbf{R}_x^{-1} \, \mathbf{Q}_x^{\mathsf{T}} \, \boldsymbol{\epsilon} - \mathbf{C} \, \mathbf{R}_x^{-1} \, \mathbf{Q}_x^{\mathsf{T}} \, \mathbf{1}_j \, b$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \dot{\mathbf{O}} \\ \mathbf{0} & \mathbf{1} & \vdots \end{bmatrix},$$

each side can be premultiplied by its transpose to yield

$$\delta \mathbf{x}_H^{\mathsf{T}} \, \delta \mathbf{x}_H = \boldsymbol{\epsilon}^{\mathsf{T}} \mathbf{Q}_x \mathbf{R}_x^{-\mathsf{T}} \mathbf{C}^{\mathsf{T}} \mathbf{C} \, \mathbf{R}_x^{-1} \mathbf{Q}_x^{\mathsf{T}} \boldsymbol{\epsilon}$$

$$- 2 \boldsymbol{\epsilon}^{\mathsf{T}} \mathbf{Q}_x \mathbf{R}_x^{-\mathsf{T}} \mathbf{C}^{\mathsf{T}} \mathbf{C} \, \mathbf{R}_x^{-1} \mathbf{Q}_x^{\mathsf{T}} \mathbf{1}_j b$$

$$+ \mathbf{1}_j^{\mathsf{T}} \mathbf{Q}_x \mathbf{R}_x^{-\mathsf{T}} \mathbf{C}^{\mathsf{T}} \mathbf{C} \, \mathbf{R}_x^{-1} \mathbf{Q}_x^{\mathsf{T}} \mathbf{1}_j b^2$$

When the left side of this expression is chosen to exceed (CWL^2) by a modest amount (design margin), the resulting quadratic equation can be solved for the requisite bias level. This is the value to be reached by a step or ramp in order to prevent an inconclusive run. It must now be assured that the test does not demand unrealistic performance from the receiver in extreme geometries; this is verified by substituting the resulting bias level into the expression for Missed Detection Probability,

$$\frac{1}{2} \operatorname{erfc} \left\{ \frac{b \, |\mathbf{m_j}| - T_D}{\sigma \sqrt{2}} \right\}$$

and allowing this test case to be included only if the result – even with the enlarged threshold – is less than 0.001 (corresponding to a scaled value of 0.03 in these tests). It should be noted that this test design approach is superior to methods relying on protection radii and minimum detectable biases; here we exploit knowledge of specific random sample values (ϵ).

REFERENCES

1. Farrell and van Graas, "Statistical Validation for GPS Integrity Test," *IONJ* v 39 n 2, 1992, p. 205.

2. Mood and Graybill, *Introduction to the Theory of Statistics*, McGraw-Hill, 1963.

3. Sturza, *"Navigation System Integrity Monitoring Using Redundant Measurements,"* *IONJ* v 35 n 4, 1988, p. 487.

4. vanGraas and Farrell, "Receiver Autonomous Integrity Monitoring (RAIM): Techniques, Performance, and Potential," *ION 47^h Annual Meeting Proceedings*, Williamsburg VA, June 1991, *pp.* 421 - 428.