A dynamic sampling scheme for GPS integrity assessment

S. Feng and W. Ochieng
Department of Civil and Environmental Engineering
Imperial College London
London, UK

D. Walsh and R. Ioannides
CAA Institute of Satellite Navigation, University of Leeds
Leeds, UK

ABSTRACT

The Global Positioning System (GPS) is already being used for certain aviation applications and some safety critical air traffic services will be based on GPS. These air traffic services must achieve allowable levels of safety before they can be accepted. For this to occur, GPS based navigation systems must achieve a defined level of performance specified in terms of accuracy, integrity, continuity and availability. This must be determined by various analysis techniques including failure mode and effects analysis (FMEA) and integrity assessment. Because of the high percentile requirements placed on integrity (as the parameter most directly related safety), it is unfeasible to measure system performance by demonstration (field trial). Realistic simulation informed by some field experience is usually employed. However, the current simulation-based approaches for receiver autonomous integrity monitoring (RAIM) performance assessment have a number of weaknesses including the use of coarse (large) spatial and temporal sampling intervals, loose definitions of error and geometric correlations, a lack of sampling of all geometries and the inability to account for critical points due to uncorrelated factors.

This paper proposes a dynamic sampling method that takes account of these weaknesses, identifying dynamically only the required points for integrity performance assessment. Comprehensive simulations carried out to test the proposed approach for a single point, an area, and a non-precise approach (NPA) flight path to Gatwick airport in the United Kingdom show that the method can be effective in capturing all the points enabling a robust and reliable assessment of system integrity.

NOMENCLATURE

%	percentage
$\cup$	union of sets
ε	vector of measurement noise
$\mathbf{G}$	observation (measurement) matrix
i	integer variant
I	identity matrix
m	metre
$\mathbf{P}$	intermediate matrix related to position error
PE_H	horizontal position error
s	second
S	whole sample set
S_C, S_i, S_G	sample set
$SLOPE_{MAX}$	maximum slope value
SSE	sum of squared errors

Paper No. 3003. Manuscript received 8 August 2005, revised version received 9 November 2005, accepted 12 December 2005.

T	intermediate matrix related to test statistic
w	vector of residuals
x	vector of the unknown parameters
$\hat{\mathbf{x}}$	estimation of x
y	vector of measurements
$\hat{\mathbf{y}}$	estimation of y

Subscripts

i	integer variant
i j	the row column element of matrix

Superscript

T	transpose of matrix or vector

1.0 INTRODUCTION

Receiver autonomous integrity monitoring (RAIM) is a receiver-based scheme to provide timely and valid alerts to users when a satellite navigation system should not be used for the intended operation (e.g. support the navigation requirements for a phase of flight). To perform RAIM, certain conditions need to be satisfied, such as the existence of a minimum number of visible satellites and an acceptable geometric configuration of the visible satellites. RAIM availability assessment is the term used for the process to determine whether the conditions to perform RAIM exist. If RAIM is available then failure detection and exclusion processes follow.

To assess the performance of the signal-in-space (SIS) or to predict RAIM availability, a certain number of sample data points are required. This depends on a number of factors including the requirements placed on integrity expressed in terms of the risk, time-to-alert and alert limit, and system and operational elements. The method commonly used is to overlay a grid over an area of interest and to search the nodes over time. Example spatial and temporal intervals that have been used in RAIM availability assessment include: five degrees and five minutes[1], three degrees and five minutes[2,3], while Eurocontrol employs 0·25 degrees and 2·5 minutes in the AUGUR RAIM availability software[4].

The determination of the spatial and sampling intervals is based on the correlations associated with geometry and measurement error both in the space and time domains. Hence, the characteristics at a point over time and an area over time are approximated by the relevant proximate sample points. Clearly the accuracy of the estimation of the characteristics between the sample points depends on the degree of correlation (expressed as a coefficient) and the sampling interval over space and time. Higher correlation and smaller sampling intervals give less characteristic approximation error. The sampling intervals quoted above are relatively large constant values (based on loose definitions of correlation) and are therefore susceptible to inaccurate determination of RAIM availability. Smaller sampling intervals can be used in order not to miss a critical point or region (e.g. a RAIM hole) because most of the factors that affect error correlation (e.g. ionosphere error) are physical characteristics beyond human control. However, such intervals result in a large number of sample points requiring increased computational resources (this could also give erroneous results if correlation is not accounted for properly). Therefore, grid-based search methods always have to consider the trade-off between accuracy and computational load.

The assessment of a system's capability to support aircraft navigation requires a large number of samples. This is because of the high percentile requirements of safety critical operations such as aviation. The sample points on a flight path can be obtained by multiplying the velocity of an aircraft with time points at regular intervals. Thus, the temporal sample points are converted to spatial sample points. Taking a ten-minute non-precision approach (NPA) phase flight as an example, an aircraft could initiate NPA at any time during the day. Hence, to experience all satellite geometries (say at a one-second interval), a very large number of sampling points are required for a sidereal day. Furthermore, when the impact of random errors is considered, a significant number of simulations are required at each sampling point for a statistically robust assessment. As a result, a significant processing resource is required just for a ten-minute phase of flight.

Following the above example, a key issue in performance assessment is the size of the sampling intervals. Many documents that recommend standards and practices consider both the unfeasibility of the use of very small intervals and their necessity. As a result they recommend constant and relatively large sampling intervals (in space and time). Clearly the balance between grid size and sample size should be such that over-sampling is avoided while maintaining an acceptable level of accuracy of capturing all the required points. However, constant sampling intervals do not facilitate the achievement of a good balance since the correlation coefficients discussed above, are not necessarily constant over space and time. Furthermore, a constant sampling interval has weaknesses when the dynamics of a vehicle are taken into account. Even if satellite geometry changes relatively slowly, aircraft attitude could change very fast during certain time intervals which may cause frequent sudden changes in the number of visible satellites. The idea then is to find the critical spatial-temporal points (or boundary), which lie(s) between two correlated segments (or regions).

This paper proposes a dynamic sampling scheme based on a number of critical points, and the spatial-temporal correlation characteristics of both the satellite geometry and the measurement errors. The aim of the proposed scheme is to reduce the sample size without the potential risk of missing the required sampling points.

Section 2 presents the key factors that influence the determination of sampling intervals. The dynamic sampling scheme and procedure are presented in Section 3. Simulation results are presented in Section 4. The paper is concluded in Section 5.

2.0 KEY FACTORS IN THE DETERMINATION OF SAMPLING INTERVALS

To determine the sampling intervals, a number of key factors must be identified and analysed. These are:

- User requirements in terms of required navigation performance (RNP)
- Measurement error correlation
- Geometry correlation which also includes kinematics of the receiver platform
- System failure modes and effects

The relationships of these factors and the RAIM algorithm are shown in Fig. 1, where failure modes and effects analysis (FMEA) and error characteristics analysis are considered as parts of RAIM.

The basic approach to RAIM with respect to failure detection performance involves three parameters: a test statistic, a decision threshold, and horizontal and vertical protection levels (HPL/VPL). The decision whether to declare the presence of a failure requires a test statistic (an observed quantity) and the decision threshold to compare the test statistic against. The decision threshold is chosen on the basis of statistical characteristics of the test statistic and is dependent either on a false alert rate or missed detection probability. A protection level is an upper bound that a position error (a derived quantity) must not exceed without being detected with a given probability (integrity risk). The protection level is a function of geometry and minimum detectable bias (MDB) which is derived from the threshold. It is also used to determine if the right conditions exist to execute a RAIM detection algorithm.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

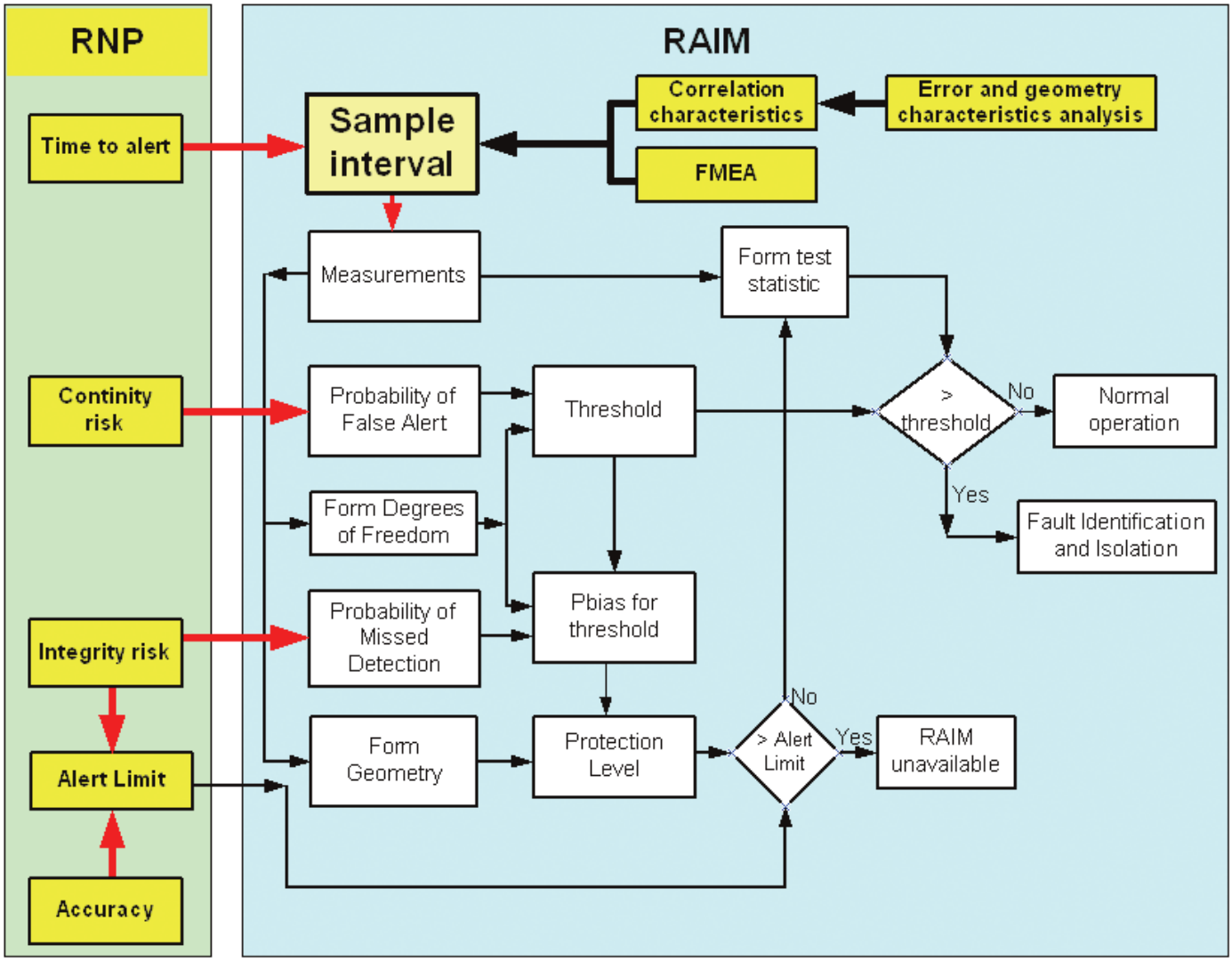

Figure 1. Mapping of RNP to RAIM algorithm.

The RNP parameters are the driver for RAIM, which can be used to derive the necessary input (including probabilities of false alert and missed detection) to RAIM algorithms. The measurement sample points used must also take into account possible system failures and correlations due to measurement errors and geometry.

2.1 Required navigation performance

Required navigation performance is the driver for RAIM as shown in Fig. 1. The RNP-RNAV (area navigation) is specified in terms of four parameters: accuracy, integrity, continuity and availability. Accuracy is defined as the degree of conformance of an estimated or measured position at a given time to a defined reference value. Ideally, this reference value should be a true value, if known, or some agreed-upon standard value. Integrity is defined as a measure of trust, which can be placed in the correctness of information supplied by the total system. Integrity includes the ability of a system to provide timely and valid warnings to the user. From the viewpoint of statistics, Integrity risk (loss of integrity) is defined as the probability that a user will experience a position error (PE) larger than the alert limit (AL) without an alert being raised within the specified time-to-alert (TTA) at any instant in time and at any location in the coverage area. Continuity of a navigation system is its capability to perform its function without unscheduled interruptions during the intended period of operation (POP) assuming that the system was operational at the start. It relates to the capability of the navigation system to provide a navigation output with the specified level of accuracy and integrity throughout the POP. Availability is defined as the percentage of time during which the service is available taking into account all outages.

The RNP parameter that is linked directly to safety is integrity. Integrity risk, alert limit, and time to alert are the three parameters that describe integrity, which are derived from the target level safety (TLS). The probability of missed detection in RAIM is derived from the integrity risk. Note that the TTA is directly linked to the temporal sampling interval, which should be less than the TTA. Accuracy is another factor related to the correlation factors described in Sections 2.2 and 2.3.

2.2 Measurement error correlation

GPS accuracy and integrity are mainly affected by satellite geometry and measurement errors. The main measurement error sources under

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

nominal conditions after the removal of selective availability (SA) are

- Satellite clock
- Satellite ephemeris
- Ionosphere
- Troposphere
- Multipath and noise

In general, different errors will have different correlation characteristics both in the spatial and temporal domains.

Satellite Clock Error: Each GPS satellite broadcasts a navigation data message containing a model to for the estimation of its clock offset from GPS time. This is used to correct the measured range (e.g. pseudorange) from the satellite to the receiver. Under fault free conditions, the satellite clock error after correction tends to vary slowly with time. The de-correlation time is the same as the update time interval (normally two hours) of the broadcast navigation data. In the spatial domain, all users within the coverage of a satellite have the same clock error. Therefore, the satellite clock error is completely correlated within the satellite coverage area.

Satellite Orbit Error: Each GPS satellite broadcasts a navigation ephemeris containing a prediction of its Keplerian orbital parameters. The satellite ephemeris data enables the determination of the satellite position and velocity. Any difference between the satellite's calculated state (position and velocity) and the true state is a potential source of orbit error. Under fault free conditions, in the temporal domain the broadcast ephemeris does not change between updates. Therefore, the de-correlation time is the same as the broadcast ephemeris update interval (normally two hours). In the spatial domain, the satellite orbit error projects onto the receiver range measurements differently at different sample points because the satellite orbit error is three dimensional while the range measurement is one dimensional. The de-correlation effect in pseudorange over distance on the Earth is less than 0·05m/100km[5].

Ionospheric Error: As the GPS signals propagate through the ionosphere, the signals are delayed by charged particles. The density of the charged particles, and therefore the delay, varies with location, time-of-day, angle of transmission through the ionosphere and solar activity. This causes errors in range measurements. In the temporal domain, the de-correlation effect of ionospheric error is less than 0·02ms^{-1}. In the spatial domain, the de-correlation effect of ionospheric error is less than 0·2m / 100km[5].

Tropospheric Error: GPS signals are also delayed as they propagate through the troposphere. This delay depends on the temperature, humidity, atmospheric pressure and the angle of transmission through the atmosphere. This delay causes errors in range measurements.

The tropospheric error is generally decomposed into two components:

- a dry component, which is a function of surface pressure and temperature, accounting for about 80% to 90% of the total delay;
- a wet component, which is a function of the distribution of water vapour, accounting for only 10% to 20% of the delay.

The dry component can be modelled accurately and corrected. However, the wet component is difficult to model accurately. In the temporal domain, this error does not suffer significant variations in timescales of a few minutes[6]. In the spatial domain, the decorrelation effect of the tropospheric error is less than 0·2m/100km[5].

Multipath and noise: Multipath is a satellite emitted signal arriving at the receiver antenna via more than one path. It is mainly caused by reflecting surfaces near the receiver. The received signals have relative offsets. Due to the arbitrary geometry of the reflecting surfaces, the correlation is very weak in both temporal and spatial domains. The GPS measurements also suffer from random observation noise, which is largely independent in both the temporal and spatial domains.

In summary, measurement errors could be either independent or correlated (over distance, time, or both). The independent error does not affect the determination of the sampling interval, while the correlated error will affect the determination of the sampling interval depending on the correlation coefficient.

2.3 Geometry correlation

Satellite geometry is the spatial distribution of visible satellites relative to the position of a GPS receiver. There are two ways to measure the geometry. One is through the so-called dilution of precision (DOP) parameters. Of the DOPs, the most common is HDOP (horizontal dilution of precision) which can give an indication of horizontal positioning accuracy if scaled with the measurement error. It plays an important role in GPS accuracy and integrity monitoring as well. The other way is to measure the relative geometry by calculating the 'slope' for each satellite. The 'slope' concept plays a key role in autonomous integrity monitoring.

The basic relationship between measurements and position is given by:

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \varepsilon \qquad \ldots (1)$$

where

- $\mathbf{y}$ is the difference between the actual measured parameter (range, pseudorange) and the predicted parameter based on a nominal user position and clock bias. It is a n × 1 vector. n is the number of measurements.
- $\mathbf{G}$ is the design matrix (i.e. the observation matrix in a local horizontal coordinates system, e.g. ENU: East-North-Up).
- $\mathbf{x}$ is the vector of the unknown parameters (i.e. three components of true position plus a userb clock bias).
- $\varepsilon = (\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n)^T$ is the measurement error vector caused by receiver noise, wave propagation, ephemeris etc.

In RAIM algorithms, the square root of the sum of squared errors (SSE) is taken as a test statistic, which can be calculated by the least squares or parity methods.

The least squares solution for the estimation of $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{y} = \mathbf{P}\mathbf{y} \qquad \ldots (2)$$

where $\mathbf{P} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}$.

The residuals can then be obtained as:

$$\mathbf{w} = \mathbf{y} - \hat{\mathbf{y}} = [\mathrm{I} - \mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G}^{-1})^{-1}\mathbf{G}^{\mathrm{T}}]\varepsilon = \mathbf{T}\varepsilon \qquad \ldots (3)$$

where $\mathbf{T} = I - \mathbf{G}\left(\mathrm{G}^{\mathrm{T}}\mathrm{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}$.

Then the SSE can be obtained from the residuals by:

$$SSE = \mathbf{w}^{\mathrm{T}}\mathbf{w} = \varepsilon^{\mathrm{T}}\mathbf{T}^2\varepsilon = \varepsilon^{\mathrm{T}}\mathbf{T}\varepsilon \qquad \ldots (4)$$

The ratio of the horizontal position error (PE_H) to the test statistic, referred to as *SLOPE*, is calculated from:

$$SLOPE = \frac{PE_H}{\sqrt{SSE}} \qquad \ldots (5)$$

The *SLOPE* can be used to project the test statistic to the position domain (position error). Under the assumption that a bias exists only in the ith satellite measurement and that the others are free of noise, the sensitivity of the horizontal position error due to the bias of the ith satellite i.e. *SLOPE(i)* can be expressed as:

$$SLOPE(i) = \sqrt{(P_{1i}^2 + P_{2i}^2)(n-4)/T_{ii}} \qquad \ldots (6)$$

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

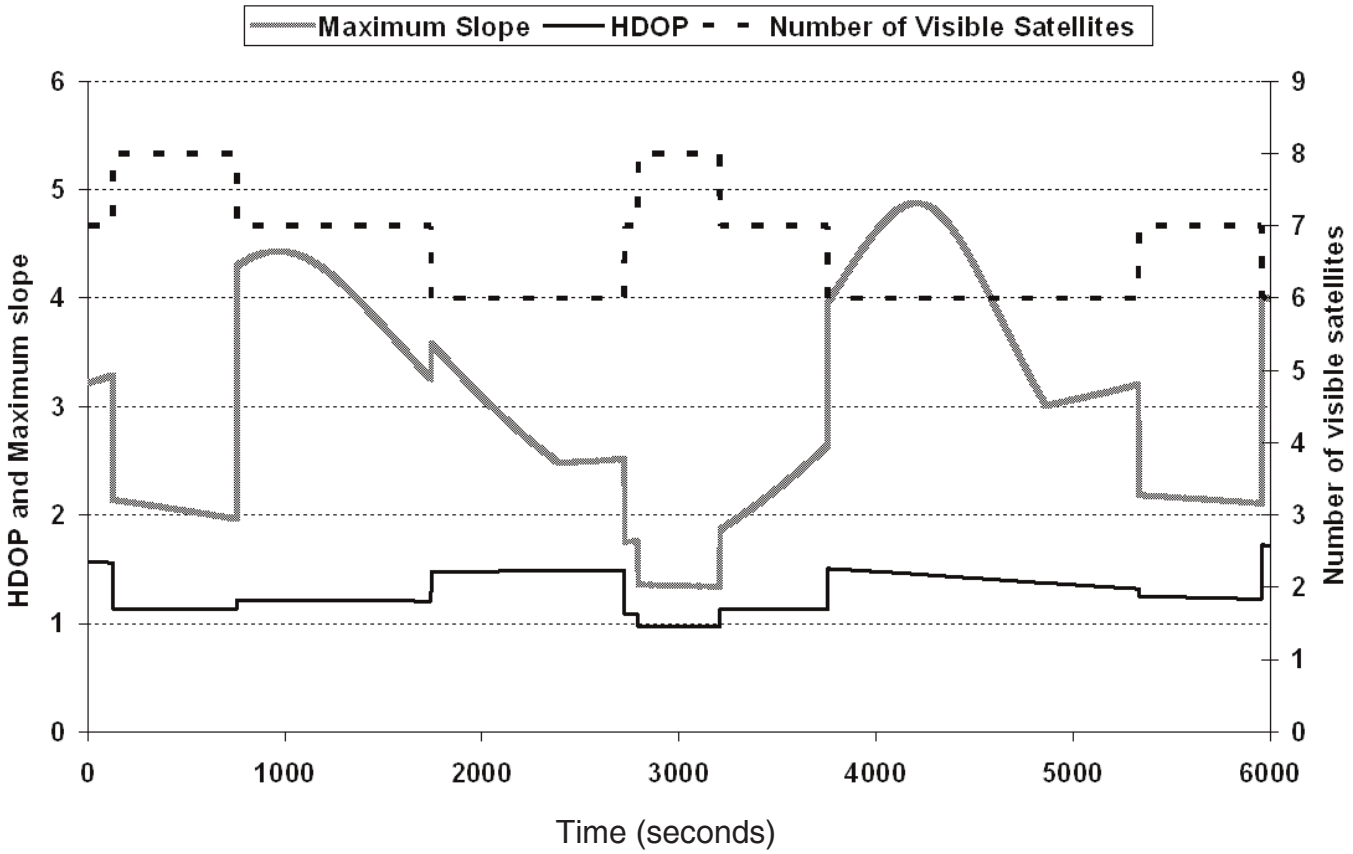

Figure 2. Relationship between geometry measures and number of visible satellites.

where T_{ij} denotes the ith element of the diagonal matrix **T**. The maximum slope, $SLOPE_{MAX}$, can then be determined as:

$$SLOPE_{MAX} = Max\left(SLOPE(i)\right) \quad \ldots (7)$$

Note that *SLOPE(i)* is purely a geometry factor. The satellite with the largest slope $SLOPE_{MAX}$ is the most difficult to detect for a given position error, as it yields the smallest test statistic (i.e. it has the highest influence on the position error). However, in reality the actual *SLOPE* value also contains noise from the lower slope satellites.

Two key factors dominate the geometry measures in either the DOP or *SLOPE*. One is the number of visible satellites and the other is the spatial distribution of these satellites. A change in the number of visible satellites causes a sudden jump in both HDOP and $SLOPE_{MAX}$ values, while slow changes occur during the interval when the number of visible satellites is constant. Figure 2 demonstrates the relationship between the geometry measures and the number of visible satellites for a simulated scenario where a receiver is on an aircraft travelling at 500 miles per hour. Even at this high speed, the HDOP changes very slowly in between the changes in the number of visible satellites. The maximum slope changes relatively faster, however its values are still correlated.

The points when (or where) the number of satellites change are the critical points. Neither the DOP nor *SLOPE* is correlated when their values undergo sudden changes due to the change in the number of visible satellites. Larger changes in the number of satellites will cause larger jumps in the values of the DOP and *SLOPE*. In the temporal domain, if a user is at the border of a satellite's coverage area, the satellite geometry with respect to the user may change causing a sudden jump in the DOP and *SLOPE* at the next epoch due to satellite motion. An extreme example is shown in Fig. 3(a) where coverage area boundaries of three satellites intersect at one point. The number n is the number of visible satellites at this point. At the next epoch, if these satellites move in the indicated direction, the point will lose three visible satellites as shown in Fig. 3(b), causing a big jump in the DOP and *SLOPE*. In the spatial domain, if a user is at the point of the intersection of the coverage area boundaries of several satellites, the DOP and *SLOPE* values may undergo a significant jump depending on the satellite motion. The satellite coverage boundaries are also the boundaries across which the geometry factors jump. Figures 3(a) and 3(b) also illustrate the changes in the number of satellites in the spatial domain.

Both the DOP and *SLOPE* change slowly in between sudden jumps, which indicates a certain degree of correlation. One theory is that the correlation time is approximately one hour, based on the average length of time satellites are visible to the user[7]. However, the value changes faster than the DOP value and therefore the correlation time should be of the order of a few minutes.

The critical points where the geometry jumps occur are the boundaries of correlation. These critical points must not be missed in RAIM assessment.

2.4 System failure mode and effects

The discussion in Sections 2.2 and 2.3 are based on the assumption of normal conditions where there are no failures. However, this is not always the case when considering the possible failures of GPS.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

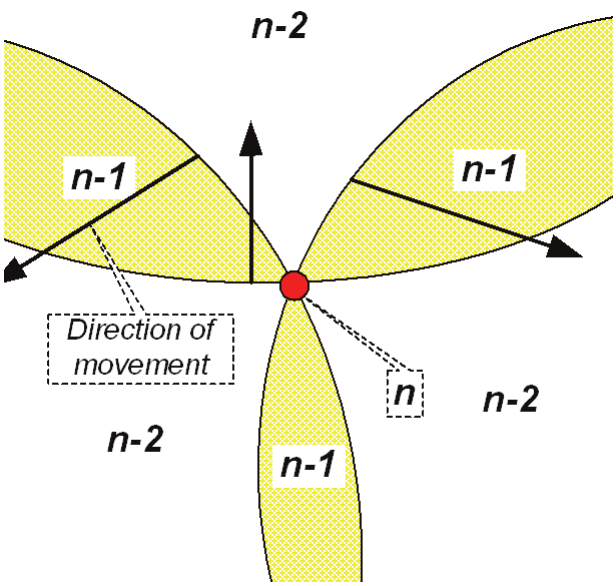

Figure 3(a). Number of visible satellite on and around a critical point.

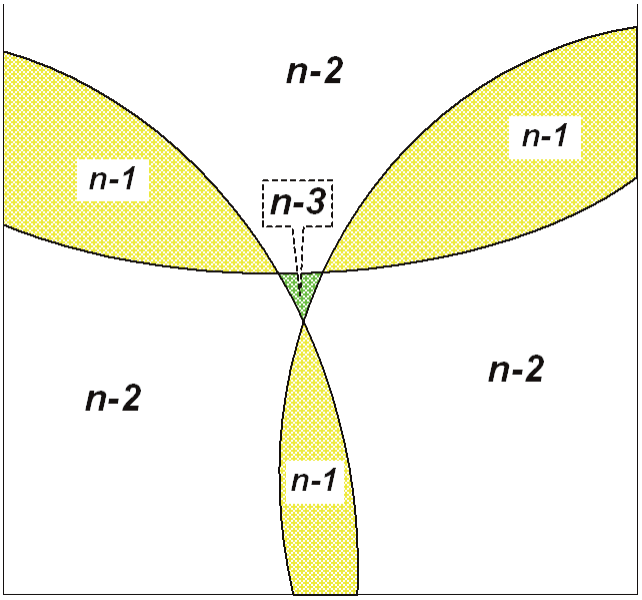

Figure 3(b). Number of visible satellite around a critical point at the next epoch.

GPS is a complex system consisting of three segments: space, control and user. There is potential for failure at any one of a number of stages, from the production of data messages and their upload to the GPS satellites, to their transmission, reception and processing within the user receiving equipment. Hence, failures could occur at system, operational environment and user receiver levels[8].

System level failures are those that occur within the space segment, the control segment and the interface between the two (i.e. data transmission). Such failures, for example, due to weaknesses in satellite design and algorithms within the master control station (MCS) environment, mainly result in excessive range errors. System level failures are usually listed under six categories: those related to erroneous clock behaviour, incorrect modelling and malfunction of the MCS, satellite payload performance, space vehicle performance and radio frequency (RF) signal performance.

Failures in the operational environment are mainly due to interference (intended and unintended) and the effects of the media along the signal path. The primary signal characteristic that makes GPS vulnerable to interference is the low power of the signal. A receiver can lose lock on a satellite due to an interfering signal that is only a few orders of magnitude stronger than the GPS signal. The intervening media between the satellite and the antenna also affects signal propagation. This includes the effects of ionosphere, troposphere and multipath.

User receiver failures relate to the end user and the end-user equipment, i.e. receiver hardware and software. Failures related to humans include the lack of adequate training, over-reliance on a single navigation system, etc. It is important to state that receivers for use with GPS for safety critical applications such as aviation must be certified to meet the minimum standards as specified by the relevant authorities.

GPS is designed to be fault tolerant – most potential failures are either handled before they manifest themselves or their effects are compensated for by the system. However, the system is susceptible to three types of failure: Insidious, Catastrophic, and Short-term transients[9]. Insidious failures are long-term (a day or more to manifest themselves) performance deviations, and do not propagate very quickly. Insidious failures are typically due to a problem in the ephemeris state estimation process. Catastrophic failures are almost instantaneous failures. These failures (e.g. satellite clock jump), in general, result in a rapid ranging signal error growth – range errors can grow to several thousand meters in a very short period of time. Short-term transients can last a few seconds and are characterised either by a short transition into and out of non-standard code, or by rapid changes in the phase of the ranging signal. Under this type of failure, receivers can experience a loss of signal availability during the time of the outage. In most cases, the effect on the receiver is far longer than the duration of outage of the satellite. The other effect is that receivers can exhibit large code and phase jumps at the onset of the outages. One example of this type of failure is a six-second outage for PRN 19 on 2 November 1998, experienced by a six- channel single-frequency receiver.

The characteristics of these failures indicate that the insidious failures have a long correlation time and large correlation distance. However, the catastrophic and transient failures are almost uncorrelated in the temporal domain, but highly correlated in the spatial domain. The sampling interval should be as short as possible in real time applications in order to catch all possible failures. One possible interval is to take the same value as the minimum sampling rate of a current GPS receiver, say, 50ms, 100ms, or 1s.

3.0 DYNAMIC SAMPLING SCHEME FOR INTEGRITY ASSESSMENT

To perform integrity monitoring, the factors discussed in Section 2 should be considered. In real time applications, the sampling interval should be as short as possible since the catastrophic and transient failures could occur at any time. However, integrity assessment is normally carried out by simulation to assess not only the effect of the deterministic error (failure) but also the random error.

In order to avoid over-sampling and reduce the sample size to a feasible level, the correlation based philosophy is used to determine a practical sample size. If a failure can be detected and excluded under a certain condition including random errors, then the expectation is that the same should apply in other 'similar' conditions. The similarity is determined by the correlation of both measurement errors and geometry. The change in the number of visible satellites, which causes geometry de-correlation, is also considered in the approach proposed here. The frequency of the change in the number

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

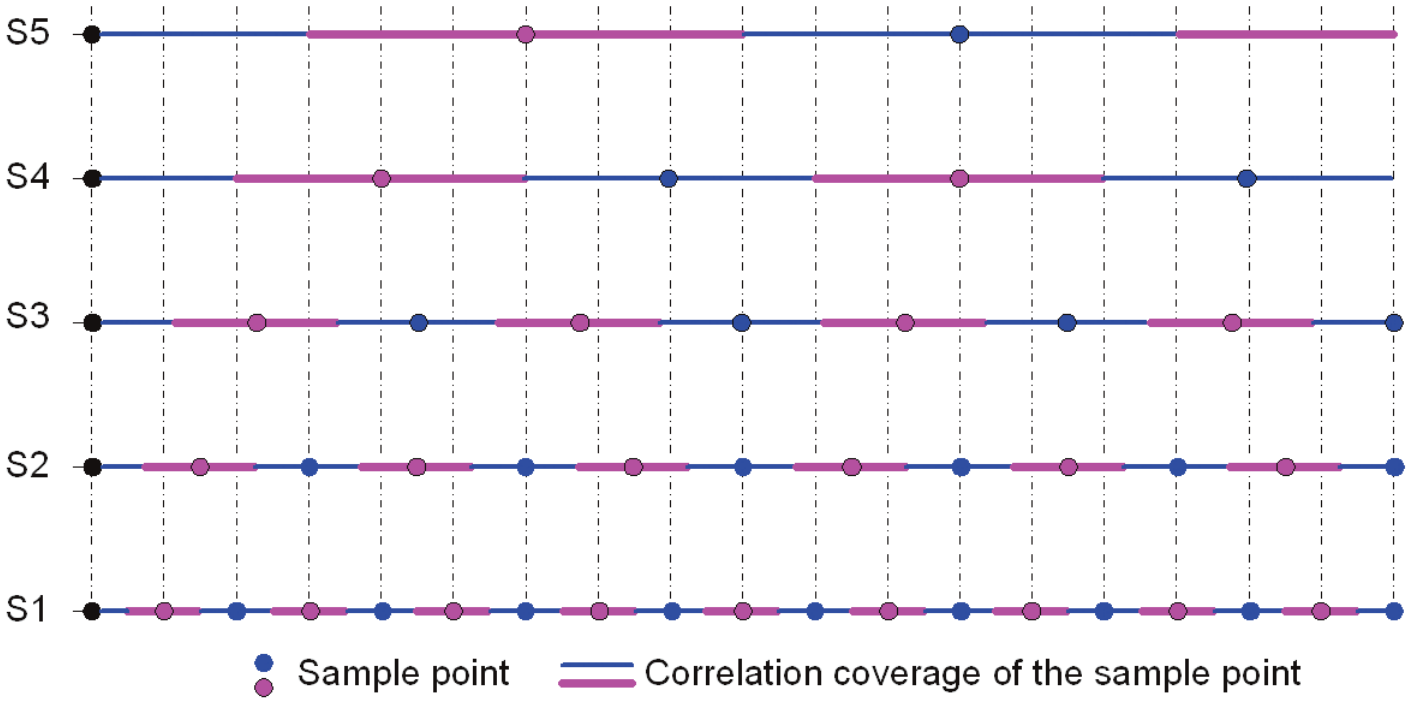

Figure 4. Correlation coverage of sample points with different sample intervals.

of visible satellites is due to satellite motion, receiver platform motion and attitude, and the relative position of the satellites and receiver platform. Clearly, the factors that affect this frequency will depend on the type of operation e.g. over a single point, an area (or global), or a flight path.

3.1 Dynamic sampling method

For each error source, a series of sample points are determined by correlation, which can be called a set of sample points. Suppose five sample sets, *S1*, *S2*, *S3*, *S4*, and *S5* are determined for the error sources based on correlation of measurement and geometry. The whole sample set of correlation is the union of these sets, and can be expressed as

$$S_C = S1 \cup S2 \cup S3 \cup S4 \cup S5 \quad \ldots (8)$$

However, the determination of the set does not follow a pure mathematical method. S_c contains all error sources. Among these sample sets, the contribution of each item to integrity monitoring is different. One way to determine the whole set is to take the set determined by the dominant factor. Suppose *S3* is the set of the dominant factor, then

$$S_C = S3 \quad \ldots (9)$$

The other approach is to consider the set that has the smallest sampling interval while covering the entire sampling space (*S1* in Fig. 4). In this case:

$$S_C = S1 \quad \ldots (10)$$

In the nominal GPS operational constellation, the satellite orbits repeat almost the same ground track once each sidereal day. The orbit altitude is such that the satellites repeat the same track and configuration over any point approximately each 24 hours (four minutes earlier each day). Hence a 24-hour geometry is taken to represent a full geometry in this paper (i.e. captures all significant geometric variations with respect to a sampling point). However, it should be noted that there may be some variations to this because of medium to long term changes in geometry due to constellation management.

Unlike accuracy assessment, good geometry is not always equally useful for integrity assessment. The geometry factor in RAIM algorithms is used to calculate the protection radius, which is the deterministic position error corresponding to the minimum detectable bias (no noise is considered)[10]. The protection radius parameter (taken directly as protection level sometimes) is not as reliable for good geometries as it is for bad ones[10]. Therefore, selecting a number of typical geometries may miss some critical points and cannot yield credible integrity assessment results. Under the full geometry consideration, how often the number of visible satellites changes depends on the relative position of the user and the satellites. In spatial and temporal domains, there exist some points near the intersection points of the coverage boundaries of two or more satellites, where the number of satellites could change quite often and within short time periods (or distances). These points are the critical points for geometry de-correlation and should be sampled. Suppose these sample points form a set S_G, then the whole sample set S can now be expressed as:

$$S = S_C \cup S_G \quad \ldots (11)$$

The determination of this whole set follows pure mathematical rules since S_C and S_G are independent.

The sample set of correlation can be based on constant intervals. However, the set of geometry de-correlation is dynamic, and therefore, the whole sample set is dynamic, hence the 'dynamic sampling method' proposed in this paper.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

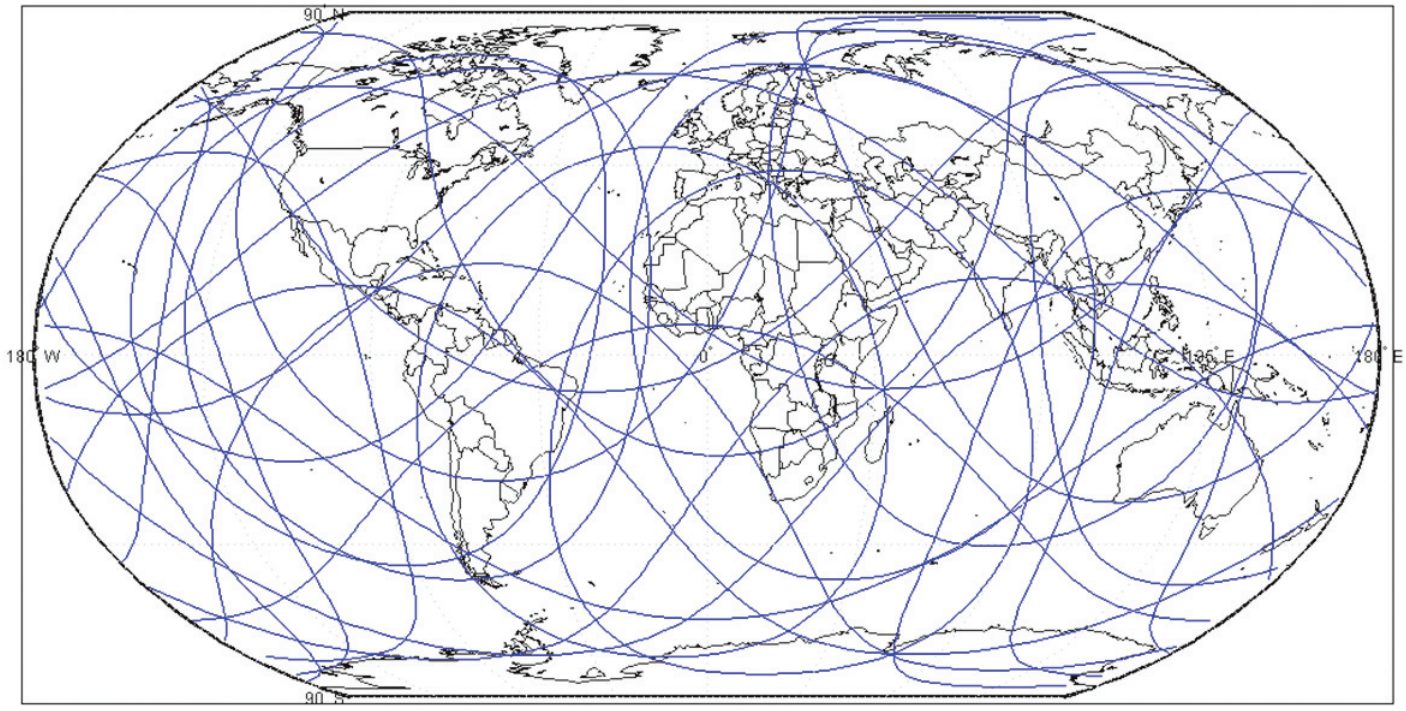

Figure 5. Patches formed by the boundaries of visible satellites.

3.2 Dynamic sampling scheme for a single point

For single point integrity assessment, there is no spatial correlation and hence only temporal correlation is considered. Post selective availability (SA), it is conservatively assumed in the literature that there are ten independent measurements in one hour (i.e. a 360-second sampling interval). Under normal conditions, the expectation is that these measurements should be highly correlated. However, the conservative assumption above was made in order to allow for non-nominal variations due to multipath or ionosphere[11]. The other error sources have longer correlation time than 360 seconds. The geometry correlation time is also longer than 360 seconds except when the number of visible satellites changes very often. Therefore, in the method proposed here the sample points are determined based on a 360-second constant time interval plus the time points (critical points) when the number of visible satellites changes.

The assessment scheme scans the geometry every second. When either a point based on the constant sampling interval or a critical point is encountered, a random simulation (with an appropriate measurement error model) is carried out at a fixed interval. The process is repeated until all the points are sampled.

3.3 Dynamic sampling scheme for an area

In the spatial domain, the coverage boundaries of the satellites divide the Earth surface into hundreds of areas (patches) at anytime as illustrated in Fig. 5. The constant grid interval assessment method may miss some small patches and further miss possible integrity holes.

The philosophy of the dynamic sampling scheme is to add the critical points to a defined grid to ensure that any small patch will not be missed. The intersection points of the boundaries of any two satellites are considered as critical points because the number of visible satellites around each point is different as illustrated in Fig. 3(a) and 3(b). The coverage boundaries are based on the elevation angle of possible users to the satellite[12].

Additional points are required which should be distributed uniformly around the critical point to ensure that each patch around it has one sample point. If small positive and negative shifts are made to the elevation angle, four additional intersection points around the critical point can be obtained as shown in Fig. 6(a), where θ_0 is the elevation angle for determining the boundaries and $\Delta\theta$ is the increment. These points are the intersection points determined by the elevation angle pairs of two satellites ($\theta_0 + \Delta\theta$, $\theta_0 + \Delta\theta$), ($\theta_0 - \Delta\theta$, $\theta_0 + \Delta\theta$), ($\theta_0 + \Delta\theta$, $\theta_0 - \Delta\theta$) and ($\theta_0 - \Delta\theta$, $\theta_0 - \Delta\theta$). The new points are also considered as critical points. In this case, each patch has a number of critical points. This number is exactly the number of satellite coverage boundaries that form the patch as demonstrated in Fig. 6(b), where the number of both boundaries and additional points is four for the shadowed patch.

In this case the sample points are determined by the combination of grid points and the critical points. Performance assessment using the grid points is suitable for large patches based on the measurement error and geometry correlation, while the critical points ensure that there are at least a few points to be assessed even for small patches. As a result, using a larger grid (e.g. three degrees) plus critical points is more efficient than using a small grid (e.g. 15 minutes of arc), with a low risk of missing small integrity holes.

3.4 Dynamic sampling for a flight path

For flight path integrity assessment, the aircraft moves both in temporal and spatial domains and forms space-time points. A combination of a spatial and temporal point will determine a unique point in the flight path.

The attitude variation of the aircraft may cause a sudden change in the number of visible satellites due to the changes in the relative elevation angle (in the body frame). If the attitude variation has not caused changes in the number of visible satellites, the geometry is not expected to change significantly. However, the measurements undergo small changes depending on the gain distribution of the antenna.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

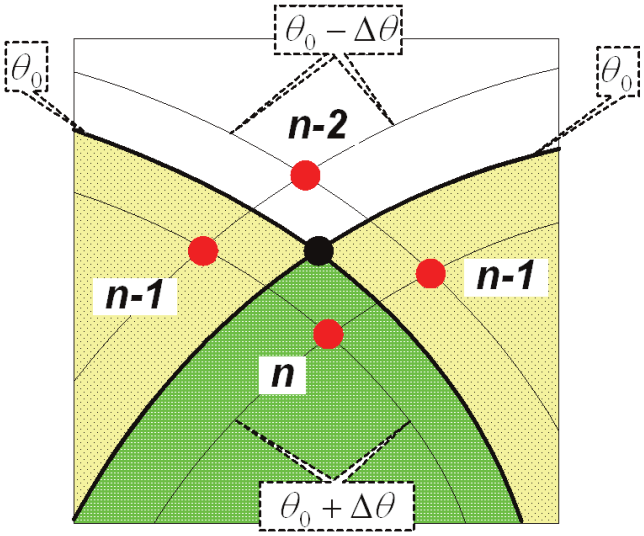

Figure 6(a). Critical points around an intersection point.

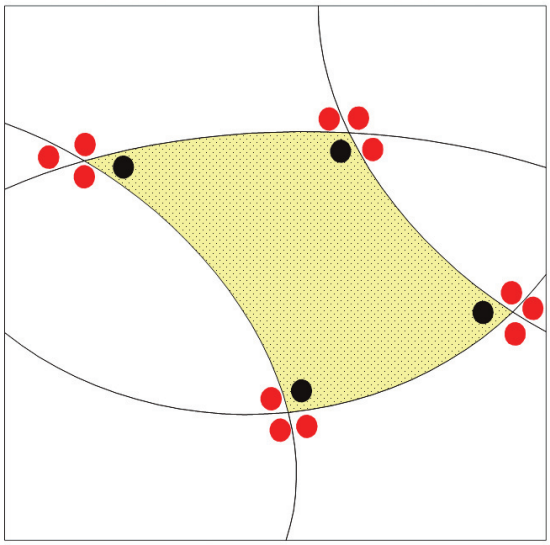

Figure 6(b). Critical points on each patch.

There are two ways to determine samples for flight path integrity assessment. One is to transform the temporal domain points to the spatial domain points by multiplying the aircraft velocity by the temporal sampling interval. Assume that an aircraft flies from the start point at a certain velocity and attitude, calculate the aircraft position and attitude and the number of visible satellites based on the elevation angle in the aircraft body frame at each following second. In addition to the sampling points based on the constant sampling interval, if the number of visible satellites changes, this space-time point is a critical point and must also be assessed by random simulation. The other way is to assess the waypoints on the flight path considering both the attitude of the aircraft and satellite number changes. In addition, the points, where the number of visible satellites changes in between adjacent waypoints, are also considered to be critical points for random simulation. At each waypoint and identified additional points, a single point assessment is carried out as described in Section 3.2.

4.0 DYNAMIC SAMPLING RESULTS FOR RAIM

4.1 The simulation scheme

The simulation scheme is based on the following basic rules

- Consideration of both random error and bias.
- Application of critical bias to the most difficult satellite to detect.
- Running of a credible number of random simulations at each sample point.
- Counting of the number of missed detections.
- Comparison with the expected number of missed detections to determine level of performance against requirements.

In integrity assessment, both potential system failures and random errors are considered. However, not all failures are difficult to detect using integrity monitoring algorithms. Even for insidious failures, only when the value of the error caused by a failure is within a certain range of the threshold (critical area in Fig. 7) can it pose a risk to integrity. It is therefore important to emphasise that in the simulation-based integrity assessment, it is only necessary to take into account the error magnitude representative of the critical area, i.e. the critical bias. The critical bias is a function of the minimum detectable bias and the diagonal element of the **T** matrix (Section 2.3) corresponding to the satellite under consideration[13].

The contribution of the error in each satellite measurement to the position error is different depending on the relative positions of all satellites. As stated earlier a failure on the satellite with $SLOPE_{MAX}$ is the most difficult to detect. Therefore, to avoid valueless simulations and reduce the number of simulations, only the satellite with the maximum slope value is considered with the critical bias and random errors applied.

The number of random simulations depends on the probability of missed detection and the confidence level of the results. In general, the larger the number of simulations, the more credible the results. The probability of missed detection is 0·001 corresponding to an integrity risk of 1×10^{-7} per hour (assuming a failure probability of 1×10^{-4}). With a confidence level of 95%, 3,300 simulations are required per sample point.

In the process of simulation, the number of missed detections is recorded. The missed detection is counted only when the test statistic is less than the threshold while the horizontal position error (HPE) is larger than the horizontal protection level (HPL). The maximum allowable number of missed detections should 4·7 for 3,300 based on statistical theory and the probabilities above.

4.2 Single point simulation

To verify the dynamic sampling scheme, four sampling intervals were simulated. The GPS almanac for Week 238 was used in the simulation. In order to verify the correlation characteristics, a modified critical bias (to give more missed detections) was applied. The sampling intervals were selected to be 2, 10, 60 and 360 seconds. The critical points where the number of visible satellites changed were also taken as sample points. At each of the sample points, the modified critical bias was applied to the satellite with maximum slope and 3,300 simulations were performed with random errors added. The modified critical bias was used in order to generate more missed detections for correlation analysis. Figures 8(a) to 8(d) show the results in terms of the number of missed detections, maximum slope, HDOP and sample points. The maximum slope and HDOP experience a sudden jump at points where the number of visible satellites changed. The left vertical

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

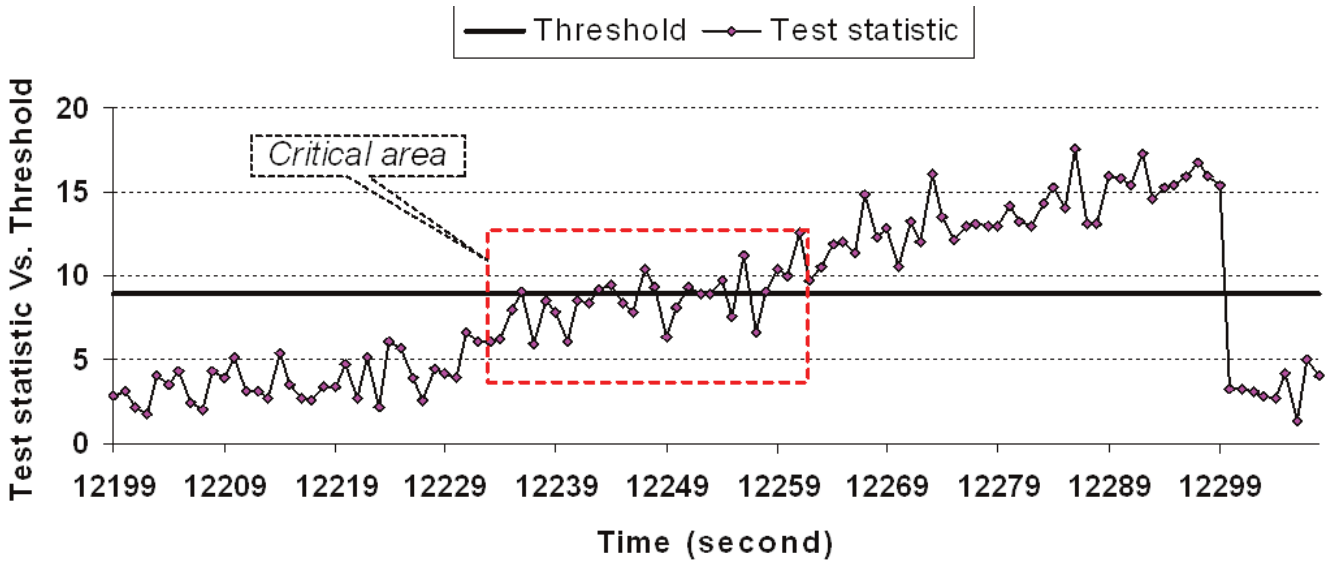

Figure 7. Critical areas for failure detection.

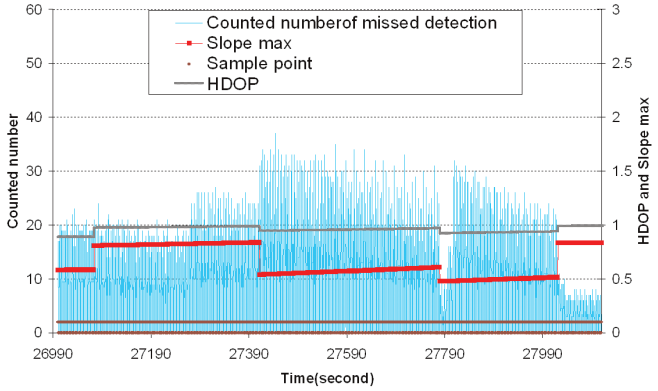

Figure 8(a). Single point dynamic sampling at a two second interval.

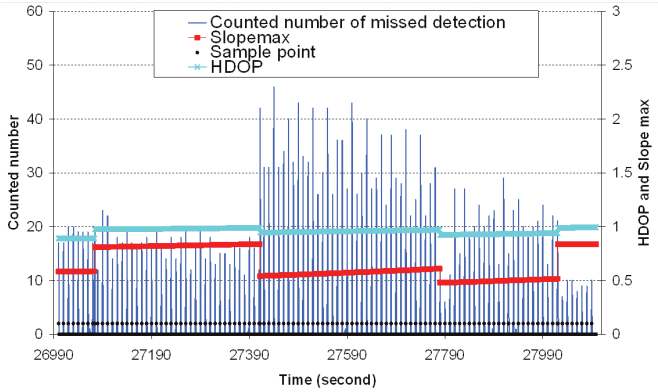

Figure 8(b). Single point dynamic sampling at a ten second interval.

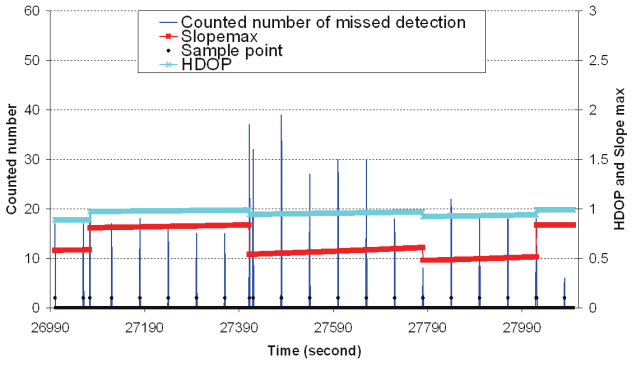

Figure 8(c). Single point dynamic sampling at a 60 second interval.

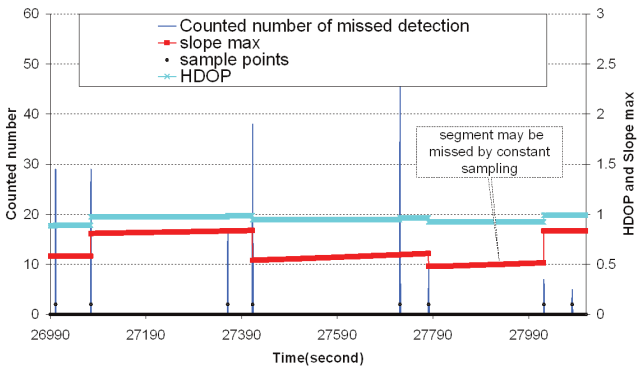

Figure 8(d). Single point dynamic sampling at a 360 second interval.

axis in each Figure is the number of missed detection counts and the right axis is the value of the maximum slope and HDOP.

The statistical results for the different sampling intervals are shown in Table 1.

The statistical results are consistent at different sampling intervals due to the geometry and measurement error correlations. It should be noted here that the use of a constant sampling interval may not be able to make full geometry assessments as demonstrated in Fig. 8(d).

4.3 Area patch simulation

Figures 9(a) to 9(e) show a patch changing over time which is formed by the coverage boundaries of satellites. The nominal satellite constellation was used[9]. The users are assumed to be on the surface of the Earth. A three degree grid was laid over a map. As shown in Fig. 9(a), at the 430,000th second, the patch on the map did not contain any grid point. A constant spatial sampling interval method would not be able to

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

Table 1
Statistical results of single point dynamic sampling

Sample interval	2s	10s	60s	360s
Average counted number of missed detections	20·3	22·1	18·7	22·6

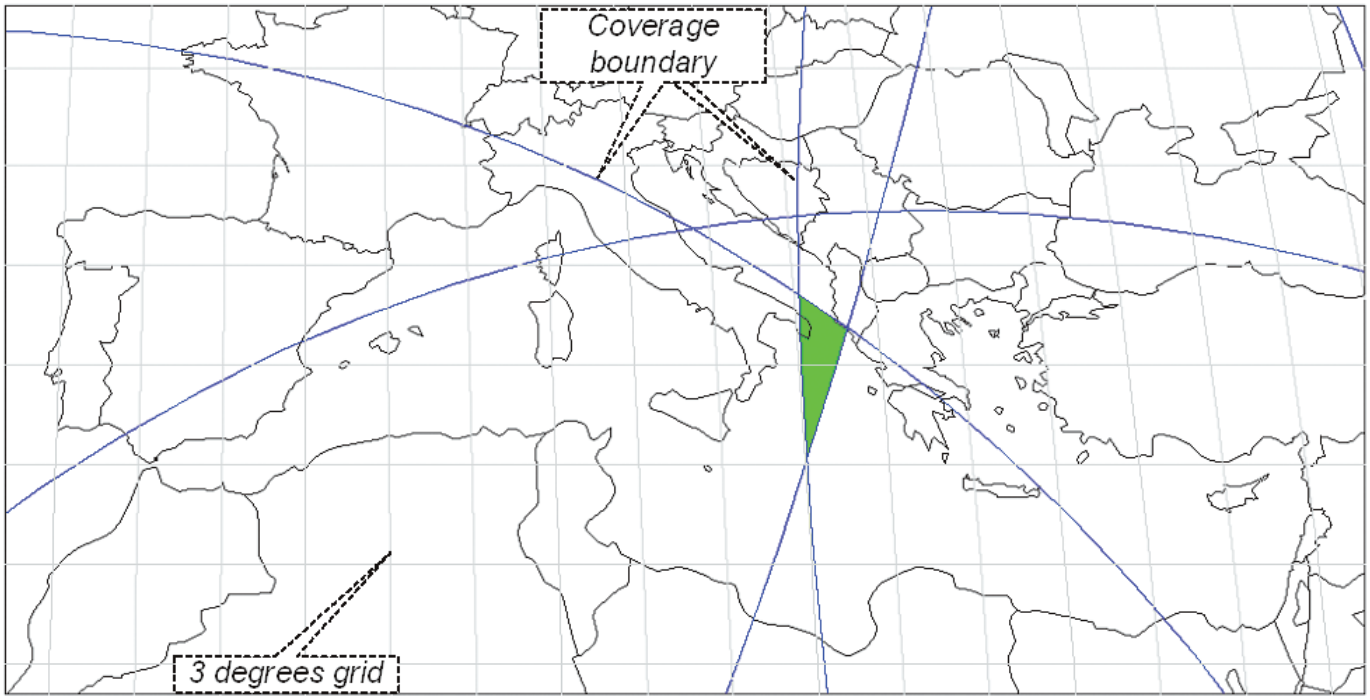

Figure 9(a). One patch at 430,000 seconds.

include this area. As captured in Figures 9(a) to 9(c) the patch shrinks over time due to user-satellite geometry changes after approximately 600 seconds. A new small patch is formed thereafter and expands with time without any grid point until approximately 1,200 seconds later. As illustrated it is clear that using a constant sampling interval in the spatial domain for integrity assessment will result in missing some critical areas where integrity holes may be present. However, with the dynamic sampling scheme, three samples will be taken from the patch at each temporal point, and will not result in missing any integrity hole.

4.4 Flight path simulation

Non precision approach (NPA) to Gatwick Airport in the UK was selected as the simulation case for integrity assessment. Fig. 10 shows the vertical profile of the flight path. The second method described in Section 3.4 was selected for the assessment.

At each waypoint on the flight path, the pitch and yaw angles of the aircraft are selected to be the same values as the angles of adjacent flight path segments, and the roll angle is assumed to be zero. There are two sets of attitude values at each waypoint in between the start and end points. The selection of pitch angles is conservative because the actual pitch angle is smaller than the flight path angle during segments with slope.

If satellite coverage boundaries divide any flight path segment into three or more segments, additional points in between adjacent boundaries on the flight path segment will be taken as critical points and assessed.

The attitude of the aircraft does not affect the relative position of satellites and the aircraft. However, the elevation angles of satellites are different between the aircraft body frame and the navigation frame (normally, East-North-Up), which may cause differences in the number of visible satellites. The mask angle in the body frame can be selected to be smaller than the mask angle in the navigation frame, which is normally taken to be five degrees. In order to assess the impact of the mask angle in the body frame, three scenarios are assessed: a zero degree mask angle both in the navigation and body frames; five degrees in the navigation frame and zero degree in the body frame; and five degrees both in the navigation and body frames.

The same satellite constellation as described in Section 4.2 was used for the simulation. A critical bias was applied to the measurement of the satellite with maximum slope, and random (nominal) errors were applied to the measurements from all satellites. To capture the full geometry, the simulation was run over 24 hours for each waypoint. At each determined critical point, 3,300 simulations were run. Table 2 shows the statistical

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

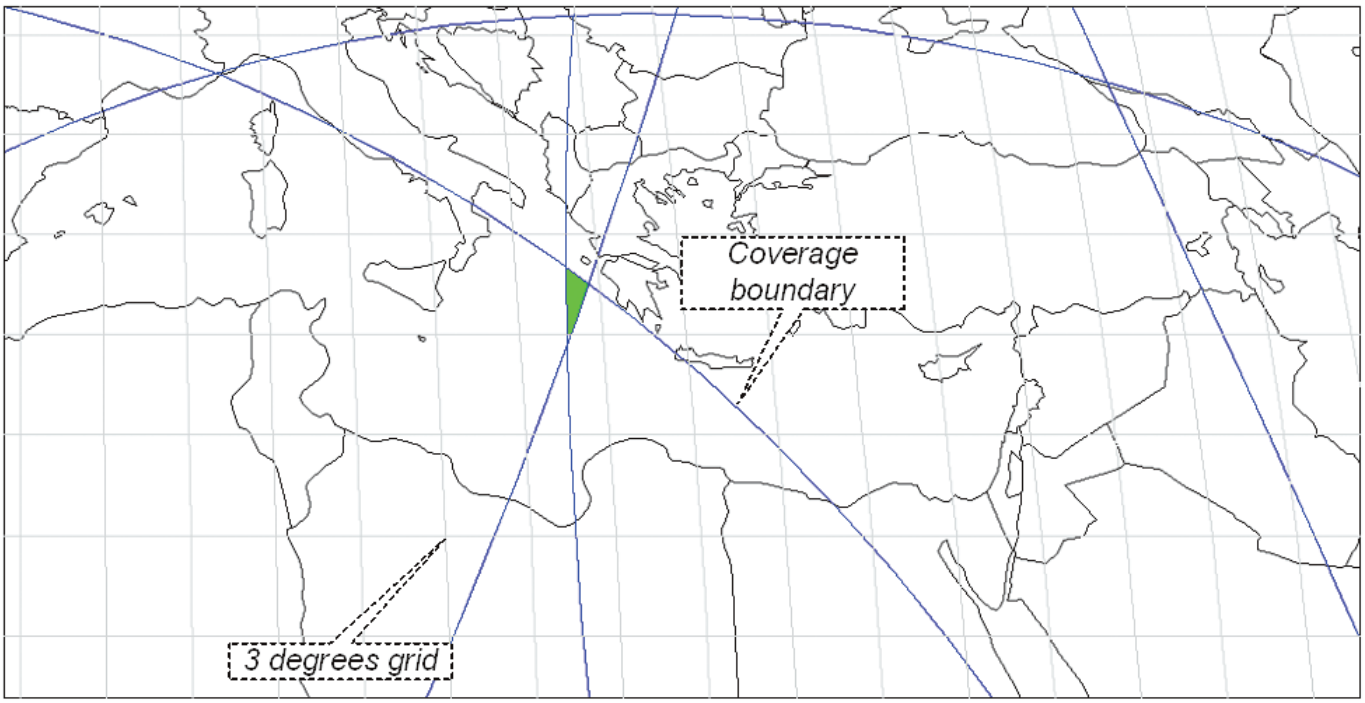

Figure 9(b). The patch at 430,300 seconds.

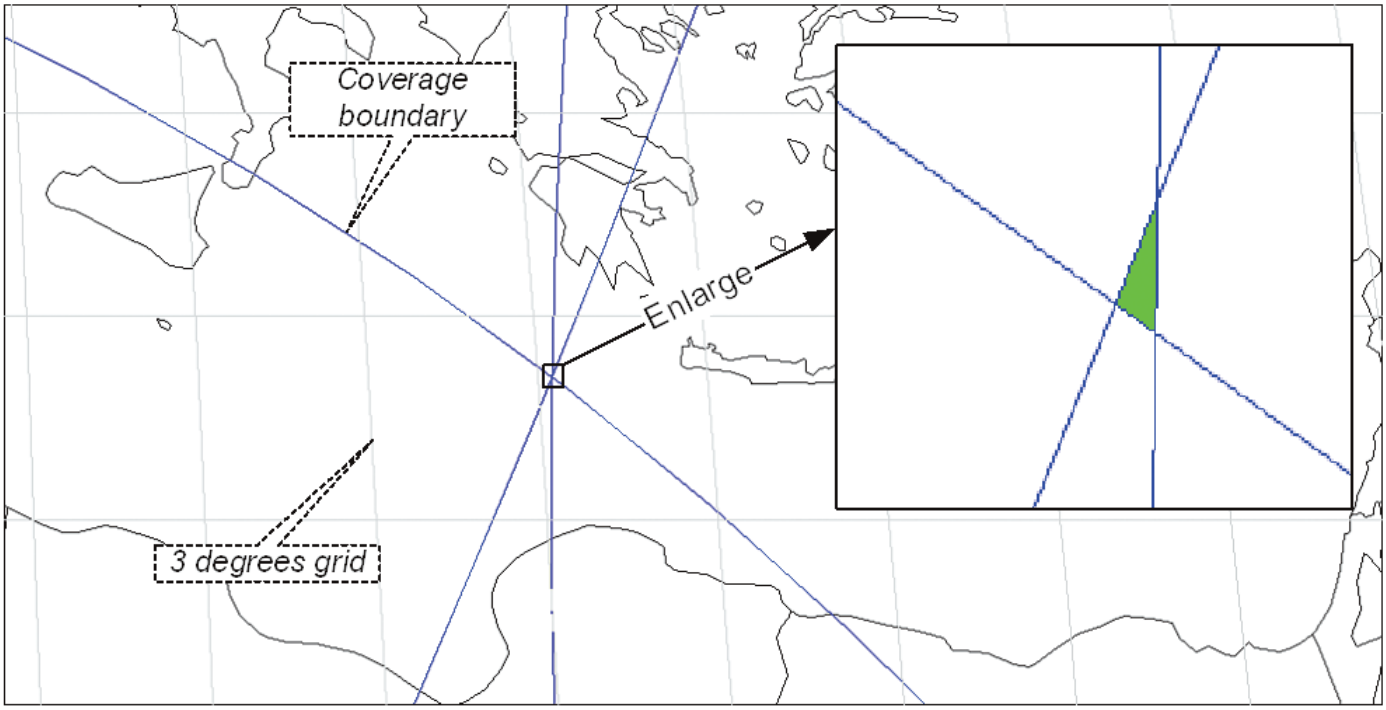

Figure 9(c). The patch at 430,600 seconds.

results of the simulation considering both attitude and mask angles.

From Table 2, the overall probability of missed detection is within the ICAO requirements for NPA (10^{-3})[15]. The missed detection (average) probability is determined by dividing the number of missed detections ('missed number') by the product of the number of the sample (critical) points and 3,300 (Section 4.1).

5.0 DISCUSSION AND CONCLUSION

The dynamic sampling scheme proposed in this paper helps to reduce the sample size for integrity assessment. It is based on a combination of a constant sample interval determined by correlation and critical points determined by uncorrelated factors both in temporal and spatial domains. The potential integrity risks of

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

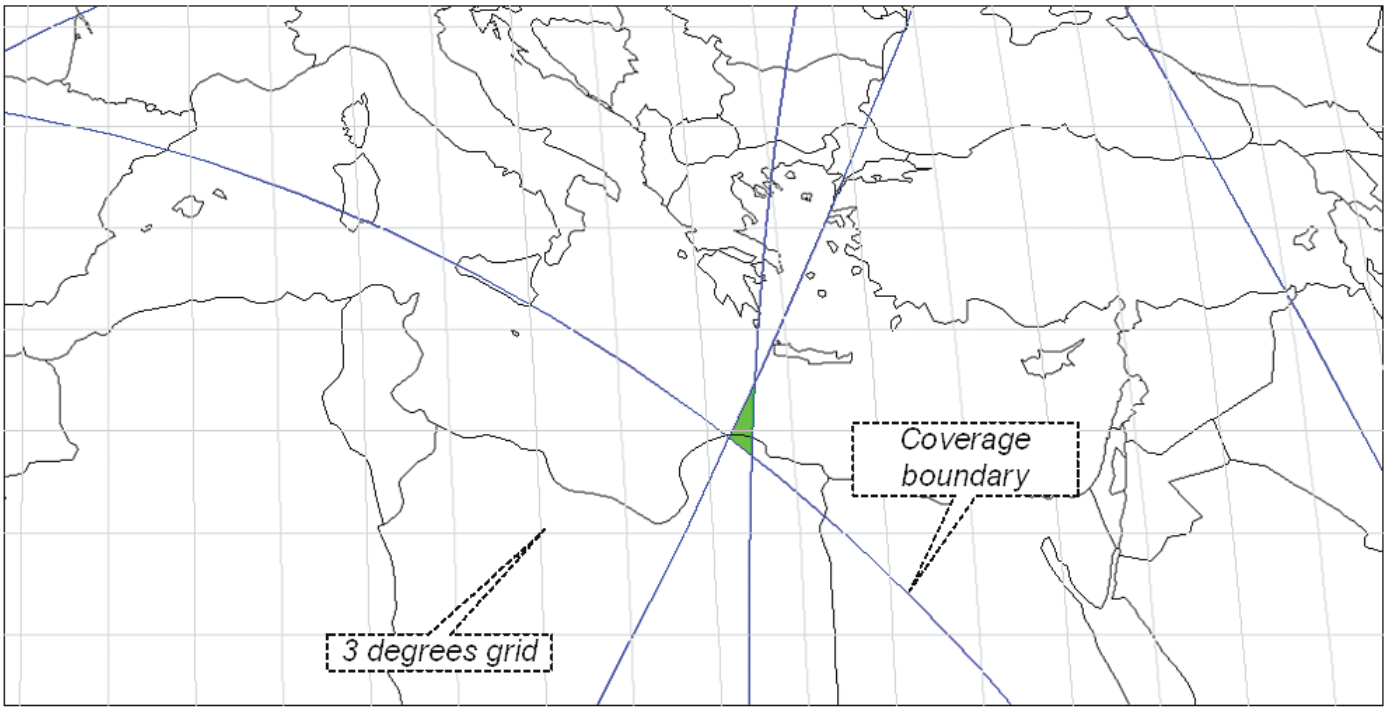

Figure 9(d). The patch at 430,900 seconds.

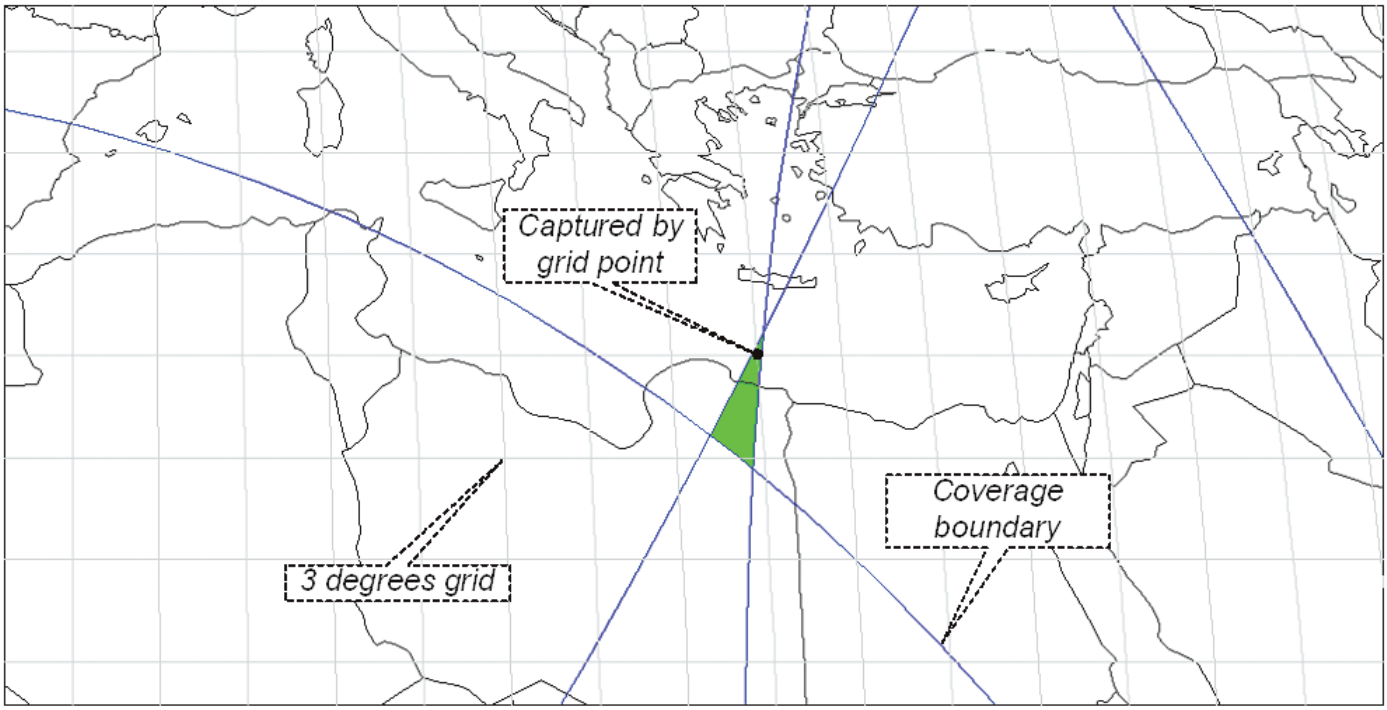

Figure 9(e). The patch at 431,200 seconds.

constant interval samples are thus removed by the dynamic sampling scheme proposed.

Although the main purpose of this paper is not to comprehensively assess the integrity at a point, over an area, or along a flight path, the results demonstrate the ability of the scheme to capture areas of potential risk to integrity which are otherwise missed using existing schemes.

The combination of GPS with other systems (e.g. GLONASS or Galileo) will increase the computation burden of performance assessment using this method. This is not considered a problem given the power of computers these days. Assuming that there are no interoperability issues (e.g. time and spatial reference systems), the overall impact of adding other systems should result in higher integrity.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

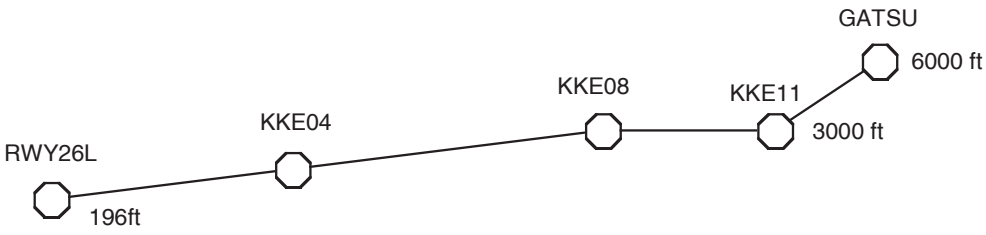

Figure 10. The flight path for NPA to Gatwick airport.

Clearly, the approach developed requires a clear definition of operation type. Although, the area assessment carried out in this paper was on the assumption that the users are on the surface of the ellipsoid. It is very easy to integrate a digital terrain model to account for topographic features such as mountains. Alternatively, it is quite possible to define spatial height layers so that the simulation is carried out in four dimensions.

All unexpected sudden changes in measurement (such as caused by ionosphere scintillations) are considered as failures. The RAIM algorithm has the ability to detect such failures.

The dynamic sampling method proposed should be incorporated into methods for the assessment of the capability of navigation systems and integrity monitoring methods to protect against significant failures. It can be used by regulators to assess the level of safety afforded by GNSS for multi-modal applications. The method can also be used by engineers to assess performance to support system design and development.

ACKNOWLEDGEMENT

This study was supported by the UK Civil Aviation Authority Safety Regulation Group (SRG). The authors wish to acknowledge the contributions made by Steve Griffin and Mark Denney of the CAA SRG and Air Safety Support International (ASSI), respectively.

Table 2
Statistical results of flight path integrity assessment

Sample points	First half day		Second half day		Missed Detection (average)	Remark
	Missed number	sample points	Missed number	sample points		
GATSU 0,0	284	166	334	164	5.6749×10^{-4}	**Column 1**
GATSU 5,0	271	159	296	163	5.336×10^{-4}	α, β (example 5,0)
GATSU 5,5	256	151	248	160	4.9108×10^{-4}	α is the values of mask angle in degree for navigation frame,
KKE11 (Horizontal)	272	161	256	163	4.9383×10^{-4}	β is the values of mask angle in degree for body frame,
KKE11 0,0	281	166	328	164	5.5923×10^{-4}	
KKE11 5,0	251	159	297	163	5.1571×10^{-4}	
KKE11 5,5	253	151	259	161	4.9728×10^{-4}	
KKE08 (Horizontal)	271	160	282	163	5.1881×10^{-4}	
KKE08 0,0	317	167	294	169	5.5105×10^{-4}	
KKE08 5,0	298	163	305	163	5.6051×10^{-4}	
KKE08 5,5	247	161	296	163	5.0786×10^{-4}	
KKE04 0,0	293	167	289	169	5.2489×10^{-4}	
KKE04 5,0	259	161	275	163	4.9944×10^{-4}	
KKE04 5,5	246	159	292	163	5.0631×10^{-4}	
RWY26L 0,0	284	167	302	169	5.2850×10^{-4}	
RWY26L 5,0	259	161	272	163	4.9663×10^{-4}	
RWY26L 5,5	253	159	283	163	5.0442×10^{-4}	
Other critical points	23			14	4.9784×10^{-4}	

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press

REFERENCES

1. OCHIENG, W.Y., SHERIDAN, K.F., SAUER, K. and HAN, X. An assessment of the RAIM performance of a combined Galileo/GPS navigation system using the marginally detectable errors (MDE) algorithm, *GPS Solution*, 2002, 5, 3, pp 42-51.
2. RTCA/DO-229C. Minimum operational performance standards for global positioning system/wide area augmentation system airborne equipment, November 2001.
3. Department of Transportation Federal Aviation Administration (FAA) Aircraft Certification Service. Airborne supplemental navigation equipment using the global positioning system (GPS)-TSO-C129a February 1996, Washington, DC, USA.
4. Eurocontrol, AUGUR, http://augur.ecacnav.com/augurHelp.html.
5. PARKINSON, B.W. and ENGE, P. K. Differential GPS, *Global Positioning System: Theory and Applications*, 1996, Volume 2, Chapter 1, pp 3-79, AIAA.
6. MONTEIRO, L.S., MOORE, T. and HILL, C. What is the accuracy of DGPS?, 2004, European Navigation Conference GNSS 2004, 16-19 May 2004, Rotterdam, The Netherlands.
7. VAN DYKE, K.L. The World after SA: Benefits to GPS integrity, Position Location and Navigation Symposium, IEEE 2000, 13-16 March 2000, pp 387– 394.
8. OCHIENG, W.Y., SAUER, K., WALSH, D., BRODIN, G., GRIFFIN, S and DENNEY, M. GPS integrity and potential impact on aviation safety, *J Navigation*, **56**, (1), pp 51-65.
9. US Department of Defense. Global Positioning System Standard Positioning Service Performance Standard, October 2001, Assistant secretary of Defense for command, control, communication and intelligence.
10. BROWN, R.G. and CHIN, Y.G. GPS RAIM: calculation of thresholds and protection radius using Chi-Square methods-A geometric approach Global Positioning System: *Navigation V*, 1998, pp 155-178.
11. CORRIGAN, T.M., HARTRANFT, J.F., LEVY, L.J., PARKER, K.E., PRITCHETT, J.E., PUE, J.E., PULLEN, S. and THOMSON, T. *GPS Risk Assessment Study*, Final Report, 1990, http://gps.faa.gov/Library/Data/gps_risk.pdf.
12. FENG, S., OCHIENG, W., WALSH, D. and IOANNIDES, R. A highly accurate and computationally efficient method for predicting RAIM holes, *J Navigation*, to be published.
13. KELLY, R.J. The Linear Model, RNP, and the near-optimum fault detection and exclusion algorithm, Global Positioning System, *Navigation V*, 1998, pp 227-259.
14. GREWAL, M., RAYTHEON, H.H. and SCHEMPP, T.R. Overview of the WAAS Integrity Design, 2003, ION GPS/GNSS 2003, Portland, OR, USA, 9-12 September 2003, pp 2750-2759.
15. ICAO/SARPS. Annex 10 Volume I *Radio and Navigation Aids*, 25 November 2004.

https://doi.org/10.1017/S0001924000001111 Published online by Cambridge University Press