

PAPER • OPEN ACCESS

Integrity monitoring for kinematic precise point positioning in open-sky environments with improved computational performance

To cite this article: Ahmed El-Mowafy and Kan Wang 2022 *Meas. Sci. Technol.* **33** 085004

View the [article online](#) for updates and enhancements.

You may also like

- [Inter-frequency code bias handling and estimation for multi-frequency BeiDou-3/Galileo uncombined precise point positioning](#)
Xinyun Cao, Fei Shen, Yulong Ge et al.
- [High-rate multi-GNSS attitude determination: experiments, comparisons with inertial measurement units and applications of GNSS rotational seismology to the 2011 Tohoku Mw9.0 earthquake](#)
Peiliang Xu, Yuanming Shu, Xiaoji Niu et al.
- [Recent advances and perspectives for positioning and applications with smartphone GNSS observations](#)
Jacek Paziewski

Integrity monitoring for kinematic precise point positioning in open-sky environments with improved computational performance

Ahmed El-Mowafy^{1,*}  and Kan Wang^{2,3} 

¹ School of Earth and Planetary Sciences, Curtin University, Perth, Australia

² National Time Service Center, Chinese Academy of Sciences, Xi'an, People's Republic of China

³ University of Chinese Academy of Sciences, Beijing, People's Republic of China

E-mail: a.El-Mowafy@curtin.edu.au

Received 12 January 2022, revised 20 February 2022

Accepted for publication 14 March 2022

Published 5 May 2022



CrossMark

Abstract

Positioning integrity monitoring (IM) is essential for liability- and safety-critical land applications such as road transport. IM methods such as solution separation apply multiple filters, which necessitates the use of computationally efficient algorithms in real-time applications. In this contribution, a new approach that significantly improves the computation time of the measurement update of the Kalman filter is presented, where only one matrix inversion is applied for all filters with measurement subsets. The fault detection and identification method and computation of the protection levels (PLs) are discussed. The computational improvement comes at the expense of a small increase in the PL. Test results for precise point positioning (PPP) with float ambiguities in an open-sky and suburban environment demonstrate the reduced computation time using the proposed approach compared to the traditional method, with 23%–42% improvement. The availability of IM for PPP, i.e. when the PL is less than a selected alert limit of 1.625 m, ranged between 92% and 99%, depending on the allowable integrity risk, tested at 10^{-5} and 10^{-6} , and the observation environment.

Keywords: GNSS, precise point positioning, integrity monitoring

(Some figures may appear in colour only in the online journal)

1. Introduction

In the past decades, accuracy has been the primary key performance indicator (KPI) for positioning using Global Navigation Satellite Systems (GNSS) for land applications. With the

advent of multiple constellations, such as GLONASS, Galileo and BeiDou, positioning availability has come into focus as another KPI. In addition, to ensure positioning reliability, methods such as receiver autonomous integrity monitoring (RAIM) (Powe and Owen 1997, Parkinson and Axelrad 1998) were developed to allow the receiver to autonomously detect and isolate faulty observations without the need to use external integrity information provided by services such as satellite-based augmentation systems (SBAS). In the last decade, there has been a growing interest in using GNSS for precise positioning of unmanned aerial vehicles and land real-time applications, such as autonomous driving. For such applications,

* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

safety is paramount, which brings into focus the integrity of the navigation system as an essential KPI that defines the level of trust of the positioning information.

Integrity monitoring (IM) can be divided into two main tasks. The first task is screening the observations for outliers, a process known as fault detection and identification (FDI). For example, in the solution separation (SS) approach, the difference between the solution based on all observations, and the solution unaffected by the suspected faulty observations, obtained from a subset of the observations by removing the suspected observations, is checked. If this difference is statistically significant, exclusion can be attempted. All observations are screened for the possibility of having faults.

In the second task, and to ensure positioning integrity, positioning errors (PEs) should be confined within a region with a boundary known as the alert limit (AL), selected according to the application and its considered integrity risk. IM aims to monitor that the PEs lie inside this region with a probability equal to at least $(1 - P_{MI})$, where P_{MI} is an application-dependent maximum allowable probability of misleading information (MI). At the same time, integrity aims to satisfy the continuity requirement by ensuring that the maximum probability of raising an alert leading to interruption of operation, without a valid reason, is P_{FA} , the probability of false alarm (FA) (Hassan *et al* 2021). This probability is a sub-allocation of the continuity requirement C_o (presented as a probability), i.e. $P_{FA} < 1 - C_o$, where C_o should also account for the probability of justified alerts in case of the error exceeding the AL (Blanch *et al* 2015). Since in practice PEs, defined as the difference between the estimated position and the true position, are not estimable in the kinematic mode since the true position is unknown, they can be conservatively replaced by a statistical upper bound of the absolute error defined according to the set P_{MI} , known as the protection level (PL). To compute the PL, one should consider several factors, including defining the nominal error model, possible threats identified by investigating GNSS measurement vulnerabilities (Imparato *et al* 2018, Du *et al* 2021) from which the alternative hypotheses (presence of fault modes) are considered, the positioning method used, the work environment, and the probability of occurrence of each fault mode. The PL is monitored not to exceed the threshold AL to declare the availability of IM of the navigation solution.

In the last three decades, IM was developed for aviation, which requires the use of certified signals and methods. Thus, positioning in aviation is typically performed using single-point positioning (SPP) based on a snapshot least-squares adjustment and employing undifferenced observations, supported by SBAS or ground-based augmentation systems (Walter 2017). In the past decade, methods such as advanced RAIM (ARAIM) (Blanch *et al* 2012, Working Group C, EU-US 2016) were developed when SBAS was not available. ARAIM can deal with multi-frequency data and the probability of experiencing multiple faults using GPS only or a combination of constellations (El-Mowafy 2016, 2017). These methods form the basis for IM for land applications where

sub-m or better positioning accuracy is needed for liability- or safety-critical applications. Achieving such accuracy requires the use of GNSS differential approaches, such as real-time kinematic (RTK) or network RTK, to cancel or reduce the spatially correlated measurement errors. Similarly, precise point positioning (PPP) can be considered when all errors, including those that are typically ignored in differential positioning, are dealt with. Therefore, ARAIM approaches need to be adapted to the methodology of these techniques and the land-based work environment, where recursive approaches such as the extended Kalman filter (EKF) are used.

The widely used ARAIM methods, such as the SS method (Jöerger *et al* 2014, Blanch *et al* 2015) require running multiple of these filters using fault-tolerant subsets of observations, i.e. formed by removing suspected faulty observations. Hence, the practical implementation of IM in real time would depend on the use of computationally efficient approaches. One of the most time-consuming processes is matrix inversion. Blanch *et al* (2019) discussed two methods to reduce the computational burden using EKF. The first method computed Kalman gain from all-in-view satellites and does not require a full matrix inversion, and the second applies fault grouping. The computation reduction using these two methods varied between 10% and 70%, but came with the drawback of being suboptimal and introducing large PL, leading to a reduction in the availability of IM. Another approach to improve the computation speed is to apply a rank-one downdating of the state estimates and covariance matrices of the all-in-view set of observations for processing observation subsets formed by removing a single observation from the all-in-view set as shown in Blanch *et al* (2012), Tanil *et al* (2018). The former study applies least squares in a snapshot approach, which is not suitable for PPP. However, using rank-one downdate in both of these studies depends on using the states estimated from the all-in-view satellites, which does not make the solution free from faults in the removed observation in the previous epochs.

In this article, we focus on PPP for road transport to benefit from its use of a single receiver without the need for reference stations, and where sub-m positioning accuracy is sufficient, which can be achieved after reaching PPP initialization and solution convergence. Positioning is considered in open-sky and low-density-structure suburban areas, where the number of GNSS satellites that have a reasonable geometry enabling positioning is assumed available, and frequent initialization of PPP is not experienced. IM for PPP has been recently discussed to some extent in (Gunning *et al* 2018). This article contributes to this effort with a focus on reducing the IM computation in EKF to support its implementation in real-time positioning.

After the introduction, a brief overview of IM and the EKF equations is given, followed by the presentation of a new approach that can reduce the computational load of EKF with ARAIM in PPP. Next, the fault detection method and computation of the PL are discussed. Testing of the proposed methods is then presented, and a summary of the proposed approach and its results conclude the paper.

2. Data processing using EKF

In traditional PPP, dual-frequency observations are considered, which give two ionosphere-free (*IF*) observations per satellite (code pseudorange and carrier-phase observations). Their observation equations and processing methodology have been extensively addressed in the literature (Zumberge *et al* 1997, Kouba and Héroux 2001, El-Mowafy *et al* 2016, Kouba *et al* 2017). In this contribution, the proposed method is presented for PPP using a traditional EKF. Although its equations are well known, it is useful to briefly overview them here as they will be visited frequently when discussing reduction of the computational load of IM in the following sections. The GNSS code and carrier-phase ionosphere-free combination observation equation in the fault-free mode at time (*t*) can be expressed as

$$y_t = A_t x_t + e_t \quad (1)$$

where (y_t) denotes the observation vector, taken as the observed-minus-computed observations based on the approximate user location and satellite position, (A_t) is the geometric Jacobian matrix since the observation equations are nonlinear, (x_t) is the state vector of the unknowns (also known as the error states), defined as the increment from the approximate location, receiver clock offset, the wet troposphere and float ambiguities. (e_t) is the random observation error vector $\sim N(0, R_y)$, where R_y is the covariance matrix of the observations. The undifferenced observations are weighted as a function of the satellite elevation angle (θ) and the carrier-to-noise density power ratio (C/N_o) with the objective of de-weighting the observations that experience significant multipaths and accordingly reduce their contribution to the solution. The model can be expressed as (El-Mowafy 2019)

$$\sigma^2 = \sigma_z^2 \times \frac{m \times 10^{-0.1 \times C/N_o}}{\sin^2(\theta)}, \quad (2)$$

where σ^2 is the slant-observation variance, and σ_z^2 is the variance in the zenith direction in the open sky computed using the method described in El-Mowafy (2015), where the interested reader may refer to this reference for their values. m is a multiplier for model calibration, which depends on the receiver type and environment (El-Mowafy 2019, Wang *et al* 2020). The covariance matrix of the *IF* observation combination is formed by error propagation. After filter initialization, where the initial states are computed from SPP processing using least squares, the time update (prediction) is applied as

$$x_t = \Phi_{t,t-1} x_{t-1} + u_{t-1} \quad \text{such that} \quad \hat{x}_t^- = \Phi_{t,t-1} \hat{x}_{t-1}, \quad (3)$$

$$P_t^- = \Phi_{t,t-1} P_{t-1} \Phi_{t,t-1}^T + \Theta_{u_{t-1}}, \quad (4)$$

where the superscript ($-$) denotes the time update, \hat{x}_t^- is the predicted state vector and P_t^- is its covariance matrix, $\Phi_{t,t-1}$ is the transition matrix, $\Theta_{u_{t-1}}$ is the covariance matrix of the process noise (u_{t-1}), which represents the uncertainty in the state vector prediction with $\sim N(0, \Theta_{u_{t-1}})$.

The measurement update starts by computing the gain matrix (K_t), followed by updating the predicted state vector and its covariance matrix:

$$K_t = P_t^- A_t^T M^{-1}, \quad (5)$$

$$\hat{x}_t = \hat{x}_t^- + K_t (y_t - A_t \hat{x}_t^-), \quad (6)$$

$$P_t = (I - K_t A_t) P_t^-, \quad (7)$$

where \hat{x}_t is the measurement-updated state vector, P_t is its covariance matrix, and $M = (A_t P_t^- A_t^T + R_y)$ is the variance-covariance matrix of the predicted residuals ($y_t - A_t \hat{x}_t^-$), also known as the vector of innovations. The EKF is sub-optimal since only the linear terms of Taylor's series are used in the Jacobian matrix, and thus, an iterative procedure is usually applied, and the assumed observation precisions might not be very accurate.

3. Proposed method for reduction of the computational load

In this study, the fault modes considered include single faulty observations and dual faulty observations, as will be discussed in the following section, and a constellation-wide mode in addition to monitoring a fault-free mode, which are two special cases. The dual faulty observations can be considered from the same satellite or from different satellites, based on studying GNSS observation vulnerabilities (Imparato *et al* 2018, Du *et al* 2021), where one may have a code outlier or phase outlier (e.g. undetected cycle slip), separately or together.

When running multiple parallel filters, where one filter is required for each of the above test modes, the inversion of the matrix M can result in a significant computation time. For n observations and u unknowns, the size of M is n . Firstly, let us consider the case of suspecting one faulty observation and form subsets by removing one observation in turn and computing the estimable unknowns and the difference between each of these solutions and the all-observation solution to check their consistency. In this case, M will be reduced to $M_{i=1 \text{ to } n}$ for the n subsets with n filters, where i refers to the matrix identifier with observation i removed. In this case, n matrix inversions would be needed in equation (5). The following approach can be applied such that only one inversion is computed for all filters. Let the subscript i refer to the case of the i th observation removed, for $i = 1$ to n , and c_i be the i th column of the identity matrix corresponding to the i th observation, such that $c_i = [\dots 0 \ 1_i \ 0 \ \dots]^T$ of size $n \times 1$. A complement matrix C_i of size $n-1 \times n$ is formed such that C_i^T together with c_i forms the identity matrix I_n of size n , such that $C_i c_i = 0$, and $C_i C_i^T = I_{n-1 \times n-1}$. The observation and dynamic models can be expressed as:

$$y_i = C_i (A_t x_t + e_t) \quad \text{with} \quad y_i = C_i y_t; \quad \text{and} \quad \hat{x}_i^- = (\Phi_{t,t-1})_i (\hat{x}_{t-1})_i, \quad (8)$$

where y_t refers to the full set of observations. When the i th observation is removed, M is reduced to $M_i = (C_i A_t P_{t_i}^- A_t^T C_i^T + C_i R_y C_i^T)$, and replacing the initial value of the state covariance matrix $P_{t_i}^-$ by P_t^- , as the only approximation used in this method, i.e.

$$\begin{aligned} M_i &\cong (C_i A_t P_t^- A_t^T C_i^T + C_i R_y C_i^T), \text{ gives} \\ M_i &= (C_i M C_i^T). \end{aligned} \quad (9)$$

It is worth noting that the approximation of using P_t^- instead of $P_{t_i}^-$ is commonly applied in practice when, for instance, one observation is lost in one epoch due to losing track of one satellite (thus one should use $P_{t_i}^-$) whereas the predicted value of the state covariance matrix (P_t^-) from the previous epoch with that observation included is used instead as the only available one. The use of P_t^- instead of $P_{t_i}^-$ is only applied for the computation of M and the correct $P_{t_i}^-$ is applied elsewhere, i.e. in equations (5), (7) and (4). Note here that both P_t^- and $P_{t_i}^-$ are independent of the values of the observations or their possible actual faults. In practice, this approximation has a minor effect on the results after solution convergence, as will be shown in section 7, since the impact of removing one observation on the estimation of P_t^- is limited (but dependent on the number and geometry of satellites) and this initial value will be later adjusted (equation (7)). This effect may also be considered, for example, by adjusting $\Theta_{u_{t-1}}$ for the case considering all observations, leading to a more conservative P_t^- , to compensate for the relatively weaker model when excluding one observation.

Ignoring the time index t in C_i , c_i and M for brevity, the gain matrix is expressed as

$$K_{ti} = P_{t_i}^- (C_i A_t)^T M_i^{-1}, \quad (10)$$

Here we use $P_{t_i}^-$ computed from past epochs, i.e. from the adjusted value $P_{t-1,i}$, by applying equation (4) (or one may apply P_t^- for consistency of the approximation), with

$$\begin{aligned} K_{ti} C_i &= P_{t_i}^- (C_i A_t)^T M_i^{-1} C_i \\ &= P_{t_i}^- A_t^T \{ C_i^T (C_i M C_i^T)^{-1} C_i \}. \end{aligned} \quad (11)$$

From appendix we have

$$M_i^{-1} = C_i \left(M^{-1} - \frac{1}{(c_i^T M^{-1} c_i)} M^{-1} c_i c_i^T M^{-1} \right) C_i^T, \quad (12)$$

which has a size $n-1$, and by pre- and post-multiplying it by C_i^T and C_i , respectively, and from (9) we have

$$C_i^T (C_i M C_i^T)^{-1} C_i = M^{-1} - \frac{1}{(c_i^T M^{-1} c_i)} M^{-1} c_i c_i^T M^{-1}, \quad (13)$$

which gives

$$K_{ti} C_i = P_{t_i}^- A_t^T \left\{ M^{-1} - \frac{1}{(c_i^T M^{-1} c_i)} M^{-1} c_i c_i^T M^{-1} \right\}, \quad (14)$$

and the time-updated state vector is expressed as

$$\hat{x}_{t_i} = \hat{x}_{t_i}^- + K_{ti} C_i (y_t) - K_{ti} C_i A_t (\hat{x}_{t_i}^-), \quad (15)$$

which is equivalent to $\hat{x}_{t_i} = \hat{x}_{t_i}^- + K_{ti} C_i (y_t - A_t \hat{x}_{t_i}^-)$.

Note here the role of C_i in reducing y_t to y_{t_i} and similarly for A_t , as shown in equation (8). Thus, for all parallel filters, when removing one observation in each, we only need to apply equations (13) and (14) with $P_{t_i} = (I - \{K_{ti} C_i\} A_t) P_{t_i}^-$ where only one inversion (M^{-1}) is required for all subsets.

Another expression of equation (14) is

$$\hat{x}_{t_i} = \hat{x}_{t_i}^- + K_{ti} (y_{t_i} - A_{t_i} \hat{x}_{t_i}^-), \quad (16)$$

where K_{ti} is computed from equation (10), and $A_{t_i} = C_i A_t$. In the proposed method, $\hat{x}_{t_i}^-$ is used as the initial value for \hat{x}_{t_i} , making it independent of the values of the observation i or its faults in all epochs, thus outperforming the rank-one downdating method (e.g. in Tanil *et al* 2018) that is similarly used to reduce the processing speed. The rank-one downdating method is typically performed using the states estimated from the all-in-view satellites, i.e. \hat{x}_t , to compute \hat{x}_{t_i} , and thus does not make the solution free from undetected faults in observation i in the previous epochs.

Similarly, when considering the case of suspecting two observations, e.g. i and j , the observation subsets are formed by removing the two observations. For each case, one can make use of C_i defined above and introduce the matrix C_j (size $n - 2 \times n - 1$) by removing the row and column of C_i corresponding to observation j , and taking c_j to be the j th column of the identity matrix of size $n-1$ corresponding to the j th observation, i.e. $c_j = [\dots 0 1_j 0 \dots]^T$ of size $n - 1 \times 1$ and not considering the duplicates. Let $C_{ji} = C_j C_i$; $y_{t_{ji}} = C_{ji} y_t$ such that the observation and dynamic models read

$$\begin{aligned} y_{t_{ji}} &= C_{ji} (A_t x_{t_{ji}} + e_t); \text{ and} \\ \hat{x}_{t_{ji}}^- &= (\Phi_{t,t-1})_{ji} (\hat{x}_{t-1})_{ji}, \end{aligned} \quad (17)$$

where the size of C_{ji} is $n - 2 \times n$, with $K_{t_{ji}} = P_{t_{ji}}^- (C_{ji} A_t)^T (C_{ji} M C_{ji}^T)^{-1}$, which gives

$$\begin{aligned} K_{t_{ji}} C_{ji} &= P_{t_{ji}}^- A_t^T C_{ji}^T (C_{ji} M C_{ji}^T)^{-1} C_{ji} \text{ or} \\ K_{t_{ji}} C_{ji} &= P_{t_{ji}}^- A_t^T C_i^T C_j^T (C_j M_i C_j^T)^{-1} C_j C_i. \end{aligned} \quad (18)$$

One can make use of M_i , computed from equation (12), when considering the exclusion of one observation first, where no inversion is needed other than M^{-1} , for $j= 1$ to $n-2$, and employing the above approximation, we have

$$C_j^T (C_j M_i C_j^T)^{-1} C_j = M_i^{-1} - \frac{1}{(c_j^T M_i^{-1} c_j)} M_i^{-1} c_j c_j^T M_i^{-1}, \quad (19)$$

which gives

$$K_{t_{ji}} C_{ji} = P_{t_{ji}}^- A_t^T C_i^T \left\{ M_i^{-1} - \frac{1}{(c_j^T M_i^{-1} c_j)} M_i^{-1} c_j c_j^T M_i^{-1} \right\} C_i, \quad (20)$$

and

$$\hat{x}_{t_{ji}} = \hat{x}_{t_{ji}}^- + K_{t_{ji}} C_{ji} (y_t) - K_{t_{ji}} C_{ji} A_t (\hat{x}_{t_{ji}}^-), \quad (21)$$

which is equivalent to

$$\hat{x}_{t_{ji}} = \hat{x}_{t_{ji}}^- + K_{t_{ji}} C_{ji} (y_t - A_t \hat{x}_{t_{ji}}^-). \quad (22)$$

Thus, for all cases when removing i and j observations from the full set, the measurement update is carried out using equations (20) and (21) with $P_{t_{ji}} = (I - \{K_{t_{ji}} C_{ji}\} A_t) P_{t_{ji}}^-$. It should be noted that \hat{x}_i and $\hat{x}_{t_{ji}}$ are tolerant to faults in measurements i and ij that could be extended over time (by the inclusion of \hat{x}_i^- and $\hat{x}_{t_{ji}}^-$ in their estimation), as the EKF is applied for the time and observation updates without the inclusion of the suspected faulty observations from the start of the filter. When a satellite appears or drops, the filtering system is updated.

In summary, one can see that for all parallel filters using observation subsets formed by removing one or two observations, and the approximation of replacing the initial value of the covariance matrix $P_{t_i}^-$ by $P_{t_i}^-$ in computing M_i and likewise when computing M_{ji} , only one inversion, i.e. M^{-1} , is performed and the rest is matrix addition or multiplication, which is computationally cheap. The gain can be seen, for example, when observing 16 satellites, assuming the use of a dual-constellation system, e.g. GPS and Galileo, with IF code and phase observations, we have $n = 32$ when testing removing one observation at a time, in addition to the possibility that all observations are unfaulty but have small errors with a combined effect that leads to a position fault. In the case of eliminating two observations, the number of test modes would be 496, computed as $\binom{n!}{2! \times (n-2)!}$. Therefore, for these cases, computing only one matrix inverse will make a significant saving in computation time.

The same concept can be applied when computing the observation weight matrix for the different subsets, estimated as the inverse of the observation covariance matrix when considering a non-diagonal matrix (e.g. for correlated observations). An example of this fully populated observation covariance matrix is when using double-difference observations in the RTK method. Note also that when running parallel filters, the PPP modeled parameters, which are insensitive to positional changes between filters including the phase wind-up, phase center offset and variations, relativistic clocks, dry troposphere estimate, solid earth tide, and ocean loading etc, are computed only once using the all-in-view position.

4. The threat model

In this paper, we test the horizontal probability of MI (P_{hMI}) of 10^{-5} and 10^{-6} (note here the use of the term hMI for *horizontal* MI, such that it is not confused with HMI—the widely used term for *hazardously* MI). For fault rates, the prior probability of single faults $P(H_i)_1$ of 10^{-4} is used for both code and phase observation faults. This probability is taken as the sum of probabilities of different observation error sources (related to satellites, the receiver and its environment), assuming a worst-case scenario where all errors are independent and

taking place together. This probability comprises 10^{-5} due to satellite faults (US. Department of Defense 2020) in addition to other possible faults, such as that owing to multipaths, which accounts for the remaining 10^{-4} . It is worth mentioning that what is relevant here is not the probability of occurrence of a multipath, which is high, but rather the probability of its outliers that can cause a positioning fault. Cycle slips are detected early in the observations before they are introduced in positioning, and undetected cycle slips are considered within the budget of observation faults. The prior probability of two simultaneous faulty observations $P(H_i)_2$, if they belong to the same satellite, is taken to be the same as the $P(H_i)_1$ of 10^{-4} , assuming that they are taking place for the same reasons. If the two simultaneous faults belong to different satellites, it is assumed that they are independent, and thus $P(H_i)_2 = 10^{-4} \times 10^{-4} = 10^{-8}$. The probability of three simultaneous faulty observations is not examined as this will have an overall extremely low probability that lies outside the significant risk allocation budget. The prior probability of a constellation fault $P(H_i)_{const}$ is assumed 10^{-8} (US. Department of Defense 2020), where the same probabilities are assumed for all constellations, and the probability of unmonitored faults $P_{unmonitored}$ is assumed to be 10^{-8} . All probabilities are for faults assumed to occur without the control center of the constellation notifying the users, e.g. through the navigation message, for one hour from the fault commencement. Again, note that these values, other than the referenced ones, are arbitrary and are assumed only to facilitate the demonstration of the presented algorithm, since currently there are no standards set for the above probabilities for land applications. Therefore, further refinements based on extensive testing and analysis of threats is needed, which is outside the scope of this paper.

The total allowable P_{hMI} requirement, for m fault modes, ignoring the unmonitored modes, is thus allocated to the following four cases:

- A fault-free case ($P_{hMI_{fault-free}}$), assumed as 10^{-8} , related to the causes of hMI due to large random errors that can occur with a small probability in the normal operation of the system, such as those caused by receiver noise, multipaths and inaccurate tropospheric delay estimation along with an unfortunate combination of bias errors (El-Mowafy et al 2015).
- The individual satellite fault probability (P_{hMI_1}), of size n , which is the sum $\left(\sum_{k=1}^n\right)$ of the product of the assumed $P(H_i)_1$, the prior probability of occurrence of a single observation fault, and the conditional probability that it is not detected by the FDI process leading to hMI.
- The dual-observation fault probability (P_{hMI_2}), which is the sum $\left(\sum_{k=1}^{m-n-2}\right)$ of the product of the assumed prior probability of occurrence of two misdetected observation faults ($P(H_i)_2$). The number of cases is $m-n-2$ (2 here refers to the next case of a constellation-wide fault in a dual-constellation system).

- The integrity risk of a constellation-wide fault ($P_{hMIconst}$), which is the sum of two cases considered for a dual-constellation system of observations.

From the above, we have

$$P_{hMI_{fault-free}} + P_{hMI1} + P_{hMI2} + P_{hMIconst} = P_{hMI} - P_{unmonitored} - P_{wFDE}, \quad (22a)$$

where P_{wFDE} is a small probability (e.g. 10^{-8}) assumed to account for incorrect observation exclusion in the FDI, which can lead to MI when exclusion is performed.

Concerning the fault duration that needs to be considered, with a focus on aviation, Walter *et al* (2019) considered this duration within the mean time to notify (MTTN), and mentioned that the yearly average MTTN obtained from the constellation service provider can be from ~ 1 h (e.g. for GPS) down to 15 min. In PPP, the MTTN can be much less than 1 h (e.g. 15 min) as the system generating the orbit and clock corrections can also notify users of whether the satellite can be used or not. In safety-critical road transport, when considering the impact of fault duration, this period should ideally be less than the time period of the stopping distance. This time period is a combination of the reaction time and the braking time (which is a function of the vehicle speed); hence, it should be less than 2–3 s. Therefore, the issue of setting representative fault durations and studying their impact on various land applications is still open and needs further research. In our work, once a fault is detected in an observation and the associated satellite continues to be tracked, the observation is excluded for 15 min (arbitrary period until further refinement) before being re-introduced again in the filtering system as a new observation. In this case, a new fault-free subset will be formed with associated estimate, FDI and PL monitoring processes, as will be explained next.

5. FDI

Utilizing the redundancy of the observations, FDI checks the consistency of all possible combinations of the measurements to recognize faulty observations (outliers) that are inconsistent with the rest of the observations. Therefore, increasing the redundancy of measurements enhances the power of FDI. In FDI statistical hypothesis testing, the null hypothesis (H_o) refers to fault-free observations, and the alternative hypotheses, e.g. H_i , refers to all possible fault modes considered, tested against H_o to detect any anomalies in the observations. If anomalies are identified, the corresponding observations can be considered for exclusion. Let us ignore the time index (t) in the sequel of the equations to simplify our presentation. Next, we present the method for the case i only for brevity, which is applicable to suspecting dual faults in ij . Applying FDI in the position domain, the SS test statistic $\Delta\hat{x}_i$ is expressed as

$$\Delta\hat{x}_i = |\hat{x}_i - \hat{x}_0|, \quad (23)$$

computed from the two solutions

$$\hat{x}_o = \hat{x}_o^- + K_o (y - A\hat{x}_o^-), \quad (24)$$

$$\hat{x}_i = \hat{x}_i^- + K_i C_i (y - A\hat{x}_i^-), \quad (25)$$

where \hat{x}_o denotes the solution using all observations, and \hat{x}_i is the solution obtained from a subset of observations, formed by removing the observations suspected to be faulty (the alternative hypothesis i). The SS standard deviation $\sigma_{\Delta\hat{x}_i q}$ for the position component (q) is taken from the SS covariance matrix. This covariance matrix is computed as follows. Starting from

$$\Delta\hat{x} = \hat{x}_i - \hat{x}_o = (\hat{x}_i^- - \hat{x}_o^-) + K_i C_i (y - A\hat{x}_i^-) - K_o (y - A\hat{x}_o^-), \quad (26)$$

$$= (\hat{x}_i^- - \hat{x}_o^-) + (K_i C_i - K_o) y - K_i C_i A (\hat{x}_i^-) + K_o A (\hat{x}_o^-), \quad (27)$$

$$\Delta\hat{x} = (K_i C_i - K_o) y + (I - K_i C_i A) \hat{x}_i^- - (I - K_o A) \hat{x}_o^-, \quad (28)$$

Letting $D_i = (K_i C_i - K_o)$, $B_i = (I - K_i C_i A)$, and $B_o = (I - K_o A)$, we have

$$\Delta\hat{x}_i = \{D_i y + B_i \hat{x}_i^-\} + B_o \hat{x}_o^-, \quad (29)$$

and thus

$$Q_{\Delta\hat{x}_i} = \{D_i R_y D_i^T + B_i P_i^- B_i^T\} + B_o P_o^- B_o^T, \quad (30)$$

where the part on the right-hand side between braces changes for each subset, and $(B_o P_o^- B_o^T)$ is the same for all subsets. Recall that for $K_i = P_i^- (C_i A_i)^T M_i^{-1}$ and $K_o = P_o^- A^T M^{-1}$, no matrix inversion is needed other than M^{-1} , as shown in the previous section.

In the horizontal 2D space, which is of interest for land applications including road transport, one can apply only a single FDI test for the joint Easting and Northing components instead of the two tests traditionally performed for the two directions (El-Mowafy 2019). By extracting $\sigma_{\Delta\hat{x}_{iE}}$ and $\sigma_{\Delta\hat{x}_{iN}}$, which represent the standard deviations of $\Delta\hat{x}_i$ in the Easting and Northing directions and their covariance ($\sigma_{\Delta\hat{x}_{iE}, \Delta\hat{x}_{iN}}$) from the 2D matrix $Q_{\Delta\hat{x}_i}$, a fault is suspected when

$$\{(\Delta\hat{x}_{iE}^T)^2 + (\Delta\hat{x}_{iN}^T)^2\} \geq T_i, \quad (31)$$

where

$$T_i = \chi_{\alpha_i}^2(df, 0) \times \left(\sigma_{\Delta\hat{x}_{iE}}^2 + \sigma_{\Delta\hat{x}_{iN}}^2 + 2\sigma_{\Delta\hat{x}_{iE}, \Delta\hat{x}_{iN}} \right) + \Delta b_i, \quad (32)$$

T_i is the outlier detection test threshold; $\chi_{\alpha_i}^2$ is a central Chi-square value related to the distribution of the test statistic of the joint $(\Delta\hat{x}_{iE}, \Delta\hat{x}_{iN})$ with df degrees of freedom, and α_i is the probability of FA allocated to fault mode i . For example, if fault modes are assumed with the same FA probability, $\alpha_i = \frac{P_{FAhorz}}{m}$, where P_{FAhorz} is the horizontal allocation of P_{FA} (assumed here 10^{-4}) and m is the total number of the considered fault modes, which is taken as the sum of the single- and dual-observation faults and constellation-wide faults. While the same probability of FA can be allocated to single-observation faults and likewise for dual-observation faults, it can be different for the two cases, as will be discussed in the next section. The case of forming subsets by excluding two observations, e.g. i and j , is performed in the same way as above but replacing \hat{x}_i by \hat{x}_{ij} , and $K_{ti} C_i$ by $K_{tji} C_{ji}$.

The term Δb_i is computed from the residual observation nominal biases, denoted as b_y , such that

$$\Delta b_i = |b_o - b_i|, \quad (33)$$

$$b_o = \sqrt{\{|S_o|_E (\zeta_o b_y)\}^2 + \{|S_o|_N (\zeta_o b_y)\}^2}, \quad (34)$$

$$b_i = \sqrt{\{|S_i|_E (\zeta_i b_y)\}^2 + \{|S_i|_N (\zeta_i b_y)\}^2} \quad (35)$$

where b_o and b_i are the projection of the vectors of the maximum nominal bias vector b_y onto the space of the unknowns in \hat{x}_o and \hat{x}_i respectively. It is assumed here that b_y is a step bias, whereas the impact of possible differential temporal nominal observation biases on the solution is considered in the predicted \hat{x}_o^- and \hat{x}_i^- since FDI is applied in the position domain. In our test, empirical values of b_y for each observation type are taken as the mean (m) of the overbounding distribution, which is discussed in the next section for demonstration of the method. These approximate values require future comprehensive study that should take into consideration various possible work conditions. Note also here that Δb_i is small since the PPP solution is driven by phase observations, and the same satellites are shared in the two solutions \hat{x}_o and \hat{x}_i except for one or two satellites according to the test mode i . We only consider here Δb_i for the observations that have satellite elevation angles below 40 degrees where the effect of the multipath could be significant (Zhang and Schwarz 1996). Thus, ζ_o is a reformed identity matrix of size n by replacing its diagonal elements corresponding to the satellites above 40 degrees with zeros. ζ_i is structured in a similar way for n_i , which is the number of observations considered in fault mode i . S_o and S_i are pseudoinverse matrices that project the observations onto the space of the unknowns. Using the least-squares concept as the basis of KF, $S_o = (A^T R_o^{-1} A)^{-1} A^T R_o^{-1}$, where $|S_o|_E$ and $|S_o|_N$ denote its components along the East and North. Similarly, $S_i = (A_i^T R_{y_i}^{-1} A_i)^{-1} A_i^T R_{y_i}^{-1}$, with $|S_i|_E$ and $|S_i|_N$ where $A_i = C_i A$.

It should be noted that since the FDI is applied in the position domain, it is a detector of faults extended over time, since the test statistic $(\Delta\hat{x}_i)$ involves the error state time updates

$(\hat{x}_i^-$ and \hat{x}_o^- , see equations (24) and (25)), which are successively computed from past epochs and using past observations, and \hat{x}_i^- is fault-tolerant over time to the suspected faulty measurement (i).

If a fault is detected, a separability check is applied by computing the correlation between faults to avoid incorrect identification when masking one fault by another correlated fault. For the observations k and j , the correlation coefficient between their corresponding errors denoted as $\xi_{k,j}$ reads (El-Mowafy 2019)

$$\xi_{k,j} = \frac{c_k^T M^{-1} c_j}{\sqrt{c_k^T M^{-1} c_k} \sqrt{c_j^T M^{-1} c_j}}, \quad (36)$$

where c_k and c_j as stated earlier are zero column vectors except for the elements corresponding to the observations k and j , respectively, which equal 1. For a suspected observation error, if a high correlation coefficient is present between it and other observation errors, we consider exclusion of these correlated observations. They are ranked, and the one with the highest test statistic in (31) with a significant difference from the other correlated errors (e.g. empirically twice the test statistic value) will be excluded first. If the test statistic is close for both correlated observation errors and one or both fail, both observations are excluded.

An alternative fault detection test, sometimes called the Chi-square test, uses a test statistic comprising the sum of squared weighted (normalized) observation residuals or innovations. It can be applied over a pre-set time window, by accumulating the test statistic over this period, for the detection of slowly accumulating observation errors (such as ramp errors). To reduce the computer memory used, the test statistic is typically time updated using a recursive formula (Teunissen and Kleusberg 1998, Gunning *et al* 2018). The relationship between this detection method and the SS test is discussed in El-Mowafy *et al* (2019). However, the Chi-square test has the drawback that the detection threshold, which is the inverse Chi-square cumulative distribution function (CDF), increases with time due to the increased degrees of freedom with the accumulation of observations over time and the increase in their number, and accordingly the test statistic. Consequently, the PL that uses this threshold increases with time, unlike the PL using SS, which typically has smaller values (i.e. better IM availability, due to comparing the increased values of PL with the constant pre-set value of AL) and lower computational load (Gunning *et al* 2018). Therefore, the Chi-square method is not used here as the main fault detection method or in the computation of PLs.

To check for ramp observation errors, we apply a relative order testing method using Kendall's Tau correlation coefficient, which is a non-parametric test to check whether the time series of each observation innovation has a trend, whereas they should be temporally uncorrelated random variables. The observation innovations of different subsets are stored for a sliding time window K (e.g. 15 epochs, application-dependent). For each observation innovation time series, we

compute the Kendall's Tau coefficient (T) and the Z-score (Z) as follows:

$$Z = \frac{3 \times T \sqrt{k(k-1)}}{\sqrt{2(2k+5)}}, \quad (37)$$

and compare the p-value of the Z-score against the selected significance level (β), e.g. 0.3% for 3σ , and check whether the p-value $> \beta$. If the test fails, a trend is suspected. This can be confirmed by checking that these innovations exceed a small selected threshold. It can further be verified by applying the Chi-square test (only for confirmation), if the computation power allows. This approach is used for the detection of small ramp errors, whereas outliers can be detected using the SS test.

6. Computation of the PL

Following the fault detection process in the previous step, the PL is computed based on bounding different errors involved in GNSS positioning. The computation of the convolution of all error sources, in case of using empirical distributions for the different error sources, will be complicated to calculate the exact probability of errors. Therefore, the process necessitates replacing the probability distribution of the combined error sources, i.e. their empirical distribution, by one distribution, the so-called overbounding distribution, which bounds errors in both the position and observation domains. For safety reasons, the threat probability computed based on this overbounding distribution should always exceed the threat probability that is computed using the empirical distribution of the combined errors.

In this article, we apply the two-step CDF overbounding method with excess mass described in (Blanch *et al* 2016). The method first determines an intermediate overbounding distribution in the sense of paired overbounding (Rife *et al* 2006) that is symmetrical and unimodal, and next overbound the intermediate distribution by a Gaussian distribution, such that

$$\int_{-\infty}^x f_a(t) dt \leq (1 + \varepsilon) \int_{-\infty}^x f_{su}(t) dt \text{ for any } x$$

$$\int_{-\infty}^x f_{su}(t) dt \leq (1 + \varepsilon) \int_{-\infty}^x f_{ob}(t) dt, \quad (38)$$

$$\text{with } f_{ob}(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}}, \quad (39)$$

where f_a is the PDF distribution of actual data, f_{su} is the distribution of the unimodal and symmetric distribution, and f_{ob} is the final Gaussian overbound distribution, and ε is the excess-mass increment.

In this work, following Wang *et al* (2020), ε is set to 0.01, and to satisfy equations (38) and (39), the overbounding standard deviation and mean $N(\sigma, m)$ are searched for each observation type (code and phase). A nested half-interval search is

performed. For code residuals, the search is conducted from 0.05 m to 1 m with a step of 0.01 m, and for phase residuals from 0.001 m to 0.03 m with a step of 0.001 m. The search stops when equations (38) and (39) are fulfilled. The search is performed for the left-side overbound for $(\sigma_{left}, m_{left})$ and the right-side overbound for $(\sigma_{right}, m_{right})$, and the final (σ, m) are taken as the maximum values from the left and right overbounds. Next, the IF σ_{IFob} is computed, which refers to the zenith direction, and then it is weighted for each satellite along its line of sight using (2).

We assume here that in open-sky and low-density suburban environments the measurement noise is uncorrelated in time and with randomization of the multipath in the kinematic mode due to the constant change of signal geometry between the vehicle and surrounding reflecting sources. In future work, we will study methods that consider temporally correlated measurement noise (Blanch *et al* 2020, Langel *et al* 2020).

Different approaches can be considered for the computation of PLs; for example, the exact, the upper-bound and the approximate upper-bound PL (Blanch *et al* 2015). In practice, results from these approaches are very close to each other as shown in (Gunning *et al* 2018); therefore, and since the computation load is of focus here, we consider the approximate upper-bound of PL expressed as

$$PL = \max \{ PL_o, PL_i |_{i=1}^m \}, \quad (40)$$

for

$$PL_i = T_i + K_{md,i} \times \sigma_i + m_i \text{ and}$$

$$PL_o = K_{md,o} \times \sigma_o + m_o, \quad (41)$$

where $\sigma_i = \sqrt{\sigma_{iE}^2 + \sigma_{iN}^2}$, $\sigma_o = \sqrt{\sigma_{oE}^2 + \sigma_{oN}^2}$, which are computed from the mapped overbound σ_{IFob} weighted using equation (2). PL_o refers to using the all-observations case. The misdetection inflation factors $K_{md,i}$ and $K_{md,o}$ for the above four cases are computed as

$$K_{md,i} |_{i=1}^n = Q^{-1} \left(\frac{P_{hMI_n}}{n \times P(H_i)_1 \times \omega_i} \right), \quad (42)$$

$$K_{md,i} |_{i=n+1}^{m-2} = Q^{-1} \left(\frac{P_{hMI_2}}{(m-n-1) \times P(H_i)_2 \times \omega_i} \right), \quad (43)$$

$$K_{md,i} |_{i=m-1}^{m-1(const)} = Q^{-1} \left(\frac{P_{hMI_{const}}}{P(H_i)_{const} \times \omega_i} \right), \quad (44)$$

and for the fault-free case

$$K_{md,o} = Q^{-1} \left(\frac{P_{hMI_{fault-free}}}{2 \times \omega_o} \right), \quad (45)$$

where $K_{md,i} |_{i=1}^n$, $K_{md,i} |_{i=n+1}^{m-2}$, $K_{md,i} |_{i=m-1}^{m-1(const)}$ are the misdetection inflation factors for single faults, dual-observation faults and constellation faults, respectively, with their cases given. Q^{-1} is the positive value of the inverse of the CDF, with

$\omega_o = (1 + \varepsilon)^{n_o}$ and $\omega_i = (1 + \varepsilon)^{n_i}$, where n_o and n_i are n for the fault-free and fault mode i . Since the PL is computed from the solution separation detection threshold (applied in the position domain) and standard deviations of the covariance matrix of the all-in-view and fault-tolerant estimates that are propagated in time, it takes into consideration the history of the filter and prior observations.

Here, one has to consider how P_{hMI1} and P_{hMI2} are allocated from the linear relationship in equation (22), where $P_{hMI1} + P_{hMI2} = P_{hMI} - P_{unmonitored} - P_{wFDE} - P_{hMI|fault-free} - P_{hMI|const}$. This allocation has to consider the relationship between the significance levels of the FDI test, i.e. the associated probability of FA α_i , and that of the misdetection of any of the four cases above — let us denote it as β_i . It is also well known that α_i and β_i are inversely proportional. The latter probability will affect the inflation factor (K_{md}) for the considered cases. Therefore, the allocation of P_{hMI1} and P_{hMI2} , and that of P_{FAhorz} for the single-observation faults and dual-observation faults, are performed such that the $K_{md,i}$ and T_i values for the two cases do not vary significantly. This was performed iteratively in this study until this condition is approximately reached.

The PLs are statistical bounds mainly computed from the threat and nominal error models, permissible level of integrity risk, the satellite geometry and assumed precision of the observations, but not from the observations themselves and their actual residuals corresponding to the final solution. Therefore, a test is given (El-Mowafy 2019) to quantify the impact of the residuals of the observations passing the FDI and the PL tests on the position solution, also known as the external reliability. In this test, we check whether the resulting positioning discrepancy ($\delta\hat{x}_i$) is less than an allowable value, selected according to the application at hand. The position change can be expressed for the case i as

$$\delta\hat{x}_i = \{K_i C_i (C_i^T M_i^{-1} C_i)^{-1} C_i^T M_i^{-1}\} C_i (y - A\hat{x}_i^-), \quad (46)$$

such that i varies between 0 and m , where $i = 0$ refers to the use of all observations, and the size of C_i is the number of accepted observations passing the FDI and PL tests, where M_i^{-1} is computed previously based on M^{-1} . The horizontal discrepancy $\delta\hat{x}_{H_i}$ is computed as the Euclidean norm of the Easting and Northing horizontal components of the vector $\delta\hat{x}_i$. To consider the validity of the solution, the following condition must be met:

$$\delta\hat{x}_{H_i} < T_H, \quad (47)$$

where T_H is the allowable horizontal error. For example, in transportation applications that require lane identification, the allowable value of T_H can be 1 m, whereas for more precise applications, such as collision alerts, the allowable T_H can be set as 0.5 m or less. If the test fails, this position cannot be validated, and an alternative positioning system should be used. Figure 1 illustrates a flowchart of the IM approach.

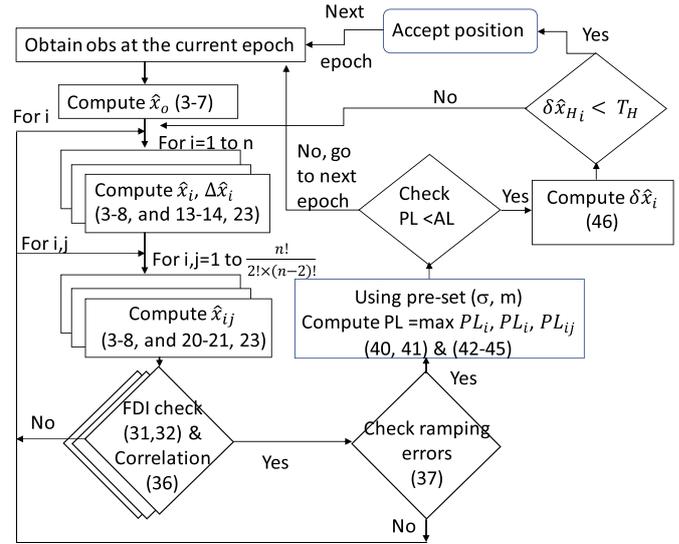


Figure 1. Flow chart of IM approach. Equation numbers are given inside the parenthesis.

7. Testing

The presented method was applied for a kinematic test carried out in Sydney, Australia, in April 2018. The test trajectory includes suburban and open-sky areas, and an area of tall buildings and large trees in the median with signal blockage experienced in low-density urban areas. The data span approximately 2.50 h at a one-second sampling rate. Figure 2 illustrates the test trajectory shown on a Google map, and figure 3 shows the test vehicle, viewing the suburban environment where part of the test was conducted. GPS and Galileo dual-frequency data (L1 and L2 for GPS, and E1 and E5a for Galileo) were processed by PPP with the float-ambiguities solution. The satellite orbit and clock corrections were received in real time, where the test was carried out during a testbed campaign for the new-generation dual-frequency multi-constellation SBAS currently under development in Australia. This new SBAS broadcasts traditional SBAS corrections for GPS L1 signals in addition to dual-frequency GPS and Galileo SBAS data, as well as dual-frequency PPP corrections for the two systems. The data were collected using a Septentrio AsteRx-U receiver with Leica AR10 antenna, and an RF front-end and Linux tablets were used to capture real-time orbit and clock corrections for PPP. These corrections have an accuracy at a sub-decimeter level (Barrios et al 2018, El-Mowafy et al 2020).

For a demonstration of the computational load reduction in the measurement update step of EKF, we consider first the case when suspecting one observation, testing all observable satellites, and forming fault-tolerant subsets without the suspected observations, and next the case when suspecting two observations, and similarly forming subsets by excluding these two observations. A variable number of satellites from 10 to 16 satellites was included in this analysis, noting that at some

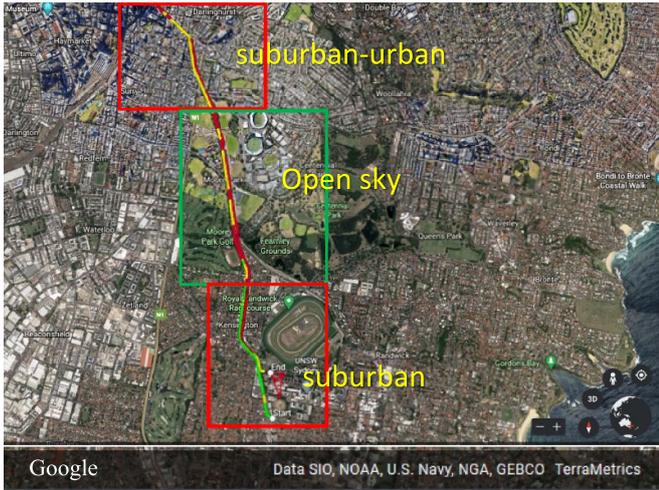


Figure 2. Test trajectory in Sydney, Australia (Google Earth) showing the different test environments.



Figure 3. The test vehicle and surrounding suburban environment.

periods of the test, and due to signal blockage due to buildings, trees, etc, positioning had a poor performance or was not available. As explained earlier, for 16 satellites, we have 32 ionosphere-free code and phase observations, and hence 32 subsets for the first case and 496 subsets for the second case, computed from $\binom{n!}{2! \times (n-2)!}$. For 10 satellites, the number of subsets in the first case is 20 and 190 for the second case. The software code used was developed in Matlab (MathWorks), noting that the authors are not professional programmers. The data were processed using an Intel® Core™ i7 processor with a CPU clock speed of 3.2 GHz with 16 GB RAM. The ratio of the processing time of the measurement update using the proposed method (equations (13) & (14) and (20) & (21)) to the processing time of the measurement update using the traditional method (equations (5) and (6)), i.e. $\frac{\text{time (proposed)}}{\text{time (traditional)}}$, is illustrated in figure 4, showing the average ratio for different numbers of observations over the test period. Since the time update step is the same in both approaches, it was not considered here. The top subplot of figure 3 shows the case of suspecting one observation, and the bottom subplot illustrates the case of suspecting two observations. The figure shows that this

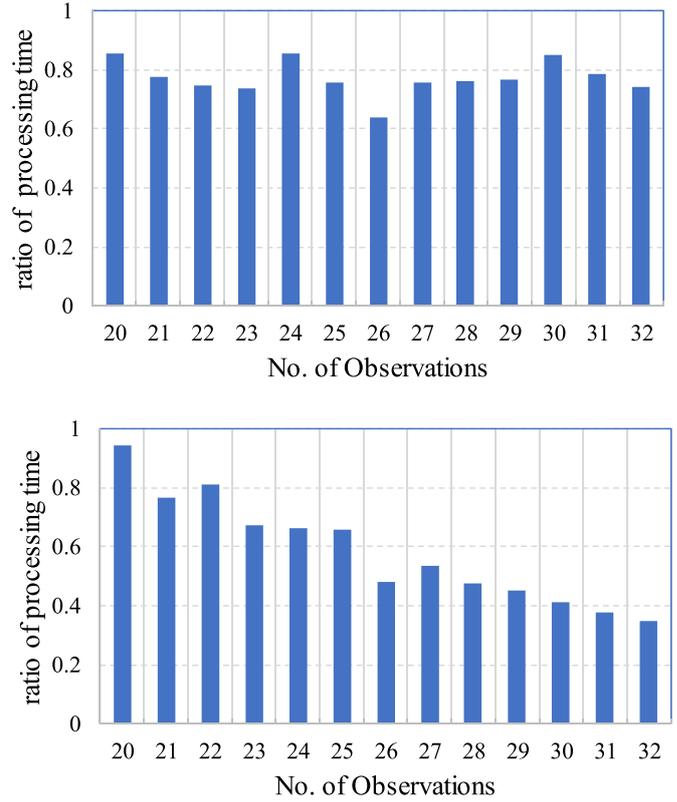


Figure 4. Ratio of processing time of the measurement update step between the proposed method and the traditional KF for (top) subsets formed with one observation removed, and (bottom) subsets with two observations removed.

ratio was on average 0.77 for case 1 (subsets with one observation excluded) and 0.58 for case 2 (subsets with two observations excluded). In the latter case, the reduction in processing time is further evident when the number of observations, and consequently the number of tested subsets, increases.

Compared with the overall results shown in Blanch *et al* (2019), where Kalman gain is derived from the all-in-view observations and does not require a full matrix inversion, the proposed method gives a better reduction in the processing time for case 1 (23% improvement using the proposed method in the computational time vs. 10% improvement in Blanch *et al* 2019), and slightly less improvement for case 2 (42% vs. 50%). The approximation made by replacing the initial value of the state covariance matrix P_t^- by P_t^- in computing M_t has a slight impact on the results, where after solution convergence the PLs increased only within 1–4 cm with a maximum 8 cm from their values when the traditional approach is applied. Similarly, the differences between the computed positions with those when using the traditional KF after convergence were within 5 mm. The proposed method thus has a slight practical impact on the solution results. The proposed method was also compared with the rank-one method that downdates the full-in-view state solution by reducing the effect of removing a single observation without performing a full matrix inversion. For the case of removing two observations, the method needs to be applied sequentially, where stored data from the filters when removing one observation

Table 1. All-satellite average overbounding standard deviations and means of GPS and Galileo signals obtained from data comprising ten kinematic tests and one-month static data.

	GPS			Galileo	
	σ (m)	m (m)		σ (m)	m (m)
Code (L1)	0.593	0.158	Code (E1)	0.508	0.145
Code (L2)	0.570	0.150	Code (E5a)	0.483	0.138
Phase (L1)	0.006	0.008	Phase (E1)	0.005	0.007
Phase (L2)	0.006	0.010	Phase (E5a)	0.005	0.010

(e.g. \hat{x}_i , P_i , etc) are used for the case of removing two observations. Comparing this method with the proposed method in the measurement update step of EKF shows that they have marginal difference in terms of the computational time, where for the case of removing one observation the method of rank-one downdating was faster by $\sim 5\%$ than the proposed method, but for the case of removing two satellites, which is the majority of cases, the proposed method was faster by $\sim 3\%$. Therefore, they are practically comparable in terms of the processing time.

However, while this comparison only considers the processing time. As noted earlier the rank-one downdating method uses states estimated from the all-in-view satellites, i.e. compute \hat{x}_i from \hat{x}_i , and thus does not make the solution free from undetected faults in observation i in the previous epochs. In contrast, the proposed method uses \hat{x}_i and \hat{x}_i^- and thus considers faults in previous epochs.

Next, the presented method was applied for the test data processed in kinematic mode in post-mission, mimicking real-time positioning. The overbounding standard deviation and mean (σ, m) of the distribution of the observation errors were estimated from the data of ten kinematic tests (carried out through 2018) combined with the static data of one month (October 2018), all with a 1 Hz sampling rate. The kinematic tests were performed in the cities of Perth, Wollongong and Sydney, Australia, where the rover antenna was mounted on the roof of a vehicle (see figure 3). The static data were collected at a known reference point on the roof of the Spatial Sciences building at Curtin University, Australia, which has a low multipath level. The observation errors, which were filtered from outliers, were computed as the difference between the collected observations and their ‘assumed-true’ values that were computed from the final precise orbits obtained from the Center for Orbit Determination in Europe (CODE) and the known receiver location. The rover receiver positions in the kinematic tests were determined by RTK-like positioning in post-mission, where the collected data were referenced to a nearby reference station.

The average estimated overbounding parameters (σ, m) for the different code and phase signals from these data are given in table 1. Again, these values should be considered as approximate only, and are utilized here as a representative sample for demonstration of the method. Naturally, practical implementation would require collecting much longer periods of data, in particular from kinematic testing and under various conditions, which is beyond the scope of this study.

The PL results of the example kinematic test described above (shown in figure 2) after PPP initialization and solution convergence are illustrated in figure 5, with a data length of 6614 s. Two values for P_{hMI} were tested (10^{-5} and 10^{-6}), where for road transport, these values are not set yet. The PEs are computed as the difference between the rover-computed PPP-float solution and the rover positions computed in RTK mode. The RTK solution was obtained by processing the rover GNSS data referenced to observations from a station located close to the test area, and the RTK positions were based on ambiguity-fixed solutions with a precision generally better than 3 cm. The detection of observation outliers is performed using the SS test. The relative order testing using Kendall’s Tau correlation coefficient of the observation innovations, and testing suspected observation innovations against a small threshold, was successful in catching small ramp injected artificial errors in a pre-test data set, in which we injected 54 ramping errors in a static 1 Hz GPS data of one hour collected at Curtin University at different satellites and different epochs with an initial value of 0.5 m and a time slope of 2%. All these artificial errors were captured by the proposed method. In the real test of the shown results, we did not inject any artificial errors, and testing did not show the presence of ramp errors. This is not surprising, given that slowly accumulating measurement errors mostly take place due to GPS satellite clock anomalies and incorrect orbit ephemeris parameters broadcast by the satellites, which is not problematic in PPP, since the orbit and clock corrections used are provided by a service that applies a quality control process to prevent such ramp errors.

The test results show that, as illustrated in figure 5, the computed PL bounds the absolute values of PE during all epochs in the test. On average, the PL was 0.98 m for a P_{hMI} of 10^{-5} , increasing to 1.07 m for a P_{hMI} of 10^{-6} . During the period of poor satellite visibility in some suburban/low-density urban areas with taller buildings and trees at the side of the road, between epochs 3860 and 4360, both the PL and PE had large values. By inspecting and comparing the results at each epoch, the PL was larger than the PE (at some epochs this may not be clear in the figure), noting that during that period the RTK also showed a poor performance.

For road transport applications, an AL of 1.625 m is used here, set as half of the average lane width, which typically ranges between 3.0 and 3.5 m. The availability of IM is assessed by checking that $\text{PL} < \text{AL}$, and is expressed as a percentage of the number of epochs when IM is available to the

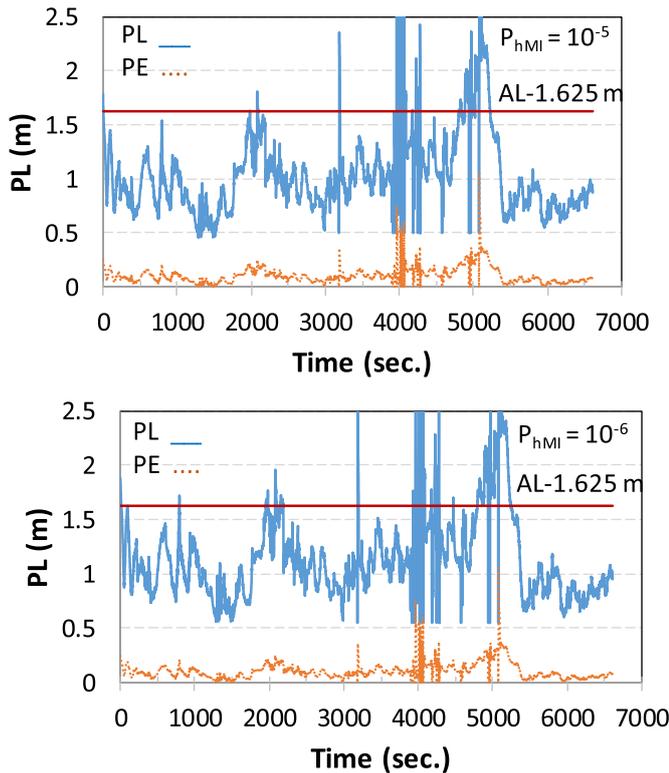


Figure 5. PL, PE and AL (top) using P_{hMI} of 10^{-5} and (bottom) using P_{hMI} of 10^{-6} .

Table 2. Percentage of IM availability.

Environment	AL (m)	$P_{hMI} = 10^{-5}$	$P_{hMI} = 10^{-6}$
Overall (open sky and suburban)	1.625	91.206%	89.007%
Open-sky environment	1.625	99.095%	97.548%

total number of epochs. As shown in figure 5 and summarized in table 2, when P_{hMI} is 10^{-5} , the overall IM availability varies from more than 99% in an open-sky environment to 91% when including the suburban environment. This is mainly due to signal blockage and a reduced number of observations and accordingly reduced PPP positioning availability and accuracy in the latter environment. The IM availability was also reduced by about 2% when using a P_{hMI} of 10^{-6} . This is an expected result and reflects the impact of setting this value for each application, which should be carefully set according to its requirements. To improve the IM availability results, one may apply a positioning method with better performance than the traditional PPP with float ambiguities, such as PPP-AR (PPP with ambiguity resolution), which typically provides better precision and reduced solution convergence time, or RTK.

8. Summary and conclusions

A method for IM using EKF is presented with a reduced computational load to facilitate real-time processing when

applying solution separation ARAIM in PPP, which requires running multiple parallel filters. In the proposed approach, only one matrix inversion is computed for these filters, avoiding the need to invert hundreds of large-size matrices. One approximation is made by replacing the initial value of the state covariance matrix $P_{t_i}^-$ by P_t^- in the covariance matrix of the innovation vector in the sub-filters. Thus, the huge gain in processing speed came at the slight expense of increasing the PL by a few centimeters. The Kalman gain matrix is computed in a simple process presented in equations (13) and (14) computed for testing the single-fault case, and equations (20) and (21) for testing the two-fault case. In the horizontal 2D space, which is of interest for land applications including road transport, the FDI can be applied using only one test for the joint Easting and Northing position components instead of the two tests traditionally performed for the two directions. Furthermore, the upper-bound of PL is applied using the computationally fast formula. The integrity risk is allocated to four modes; namely, a fault-free mode, single-observation faults, dual-observation faults and constellation-wide faults.

The presented method was demonstrated for a kinematic test in suburban and open-sky environments. PPP with float ambiguities was performed, processing GPS and Galileo dual-frequency data with real-time satellite orbit and clock corrections received from the testbed new-generation Australian SBAS service. An improvement in the measurement update of EKF was demonstrated, where the ratio of the processing time using the proposed method to the processing time using the traditional method was on average 0.77 for the case of subsets with one suspected faulty observation, and 0.58 for subsets with two suspected faulty observations. The computational time gain increased with the increase in the number of observations.

The PL results of the test after PPP initialization and solution convergence are presented with an assumed AL of 1.625 m, tested with a P_{hMI} of 10^{-5} and 10^{-6} and assumed prior probabilities (for fault-free, single-fault, dual-fault and constellation fault cases). For a P_{hMI} of 10^{-5} , the percentage of IM availability (when $PL < AL$) varies from approximately 99% in the open-sky environment to 91% when including the suburban environment, mainly due to signal blockage and a reduced number of observations at some epochs. The IM availability was reduced by about 2% when the P_{hMI} was reduced to 10^{-6} . To improve the IM availability, the use of PPP-AR is recommended. Future work will investigate refinement of the assumed probabilities and biases and include methods that bound temporally correlated non-Gaussian measurement noise that may be present, particularly in urban environments. It will also investigate the impact of the method approximation on various satellite geometries.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This study benefited from discussions with Dr Amir Khodabandeh. Fronter SI, GeoScience Australia, GMV (Spain) and Transport for New South Wales are acknowledged for their support and assistance in obtaining the data used, which were collected during the new-generation Australia/New Zealand SBAS testbed campaign. The precise orbits used for computation of the observation errors for estimation of the overbounding parameters were obtained from the Center for Orbit Determination in Europe (CODE).

Funding

This research was funded by the Australian Research Council, discovery Project No. DP 190102444, the National Time Service Center, Chinese Academy of Sciences (CAS) (No. E167SC14) and the CAS ‘Light of West China’ Program (No. XAB2018YDYLO1).

Conflict of interest

The authors declare no conflict of interest.

Appendix

Proof of equation (12), i.e.

$$M_i^{-1} = C_i \left(M^{-1} - \frac{1}{(c_i^T M^{-1} c_i)} M^{-1} c_i c_i^T M^{-1} \right) C_i^T. \quad (12)$$

For brevity let us replace the symmetric matrix M^{-1} by B , and its elements will be defined as b , such that (12) reads

$$M_i^{-1} = C_i \left(B - \frac{1}{(c_i^T B c_i)} B c_i c_i^T B \right) C_i^T. \quad (48)$$

Recall that $c_i = [\dots 0 1_i 0 \dots]^T$ is of size $n \times 1$ and C_i is of size $n-1 \times n$, which is formed such that C_i^T together with c_i forms the identity matrix I_n of size n . When eliminating the observation i , the denominator term $(c_i^T B c_i)$ in (48) becomes the element (b_{ii}) , where the subscripts here refer to the row and column numbers, respectively.

In (48), $B c_i$ gives the column i of B , and $c_i^T B$ gives the row i of B . When they are multiplied, i.e. the numerator $(B c_i c_i^T B)$ in (48), and for illustration let us arrange its elements as i, j, k, n in this order, we have the symmetric matrix

$$\frac{1}{(c_i^T B c_i)} B c_i c_i^T B = \frac{1}{b_{ii}} \begin{bmatrix} b_{ii}^2 & b_{ii} b_{ji} & b_{ii} b_{ki} & b_{ii} b_{ni} \\ \cdot & b_{ji}^2 & b_{ji} b_{ki} & b_{ji} b_{ni} \\ \cdot & \cdot & b_{ki}^2 & b_{ki} b_{ni} \\ \cdot & \cdot & \cdot & b_{ni}^2 \end{bmatrix}. \quad (49)$$

Again, note here that the subscripts of the matrix elements on the right-hand side refer to the row and column numbers. By substituting into the part between parentheses in (48), we get

$$B - \frac{1}{(c_i^T B c_i)} B c_i c_i^T B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & b_{jj} - \frac{b_{ji}^2}{b_{ii}} & b_{jk} - \frac{b_{ji} b_{ki}}{b_{ii}} & \dots & b_{jn} - \frac{b_{ji} b_{ni}}{b_{ii}} \\ 0 & \cdot & b_{kk} - \frac{b_{ki}^2}{b_{ii}} & & b_{kn} - \frac{b_{ki} b_{ni}}{b_{ii}} \\ & & \vdots & \ddots & \vdots \\ 0 & \cdot & \cdot & \dots & b_{nn} - \frac{b_{ni}^2}{b_{ii}} \end{bmatrix} \quad (50)$$

Note the symmetry of the matrix, i.e. $b_{ij} = b_{ji}$, etc, and that the zero column and row refer to the location of observation i , which were the first row and column in this example. Pre-multiplying (50) by C_i and post-multiplying by C_i^T removes these zero row and column of i , and reduces the size of the matrix to $n-1$ (the size of M_i^{-1}), i.e.

$$C_i \left(B - \frac{1}{(c_i^T B c_i)} B c_i c_i^T B \right) C_i^T = \begin{bmatrix} b_{jj} - \frac{b_{ji}^2}{b_{ii}} & b_{jk} - \frac{b_{ji} b_{ki}}{b_{ii}} & \dots & b_{jn} - \frac{b_{ji} b_{ni}}{b_{ii}} \\ \cdot & b_{kk} - \frac{b_{ki}^2}{b_{ii}} & & b_{kn} - \frac{b_{ki} b_{ni}}{b_{ii}} \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \dots & b_{nn} - \frac{b_{ni}^2}{b_{ii}} \end{bmatrix} \quad (51)$$

which is simply M_i^{-1} , where from linear algebra

$$\begin{bmatrix} b_{jj} - \frac{b_{ji}^2}{b_{ii}} & b_{jk} - \frac{b_{ji} b_{ki}}{b_{ii}} & \dots & b_{jn} - \frac{b_{ji} b_{ni}}{b_{ii}} \\ \cdot & b_{kk} - \frac{b_{ki}^2}{b_{ii}} & & b_{kn} - \frac{b_{ki} b_{ni}}{b_{ii}} \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \dots & b_{nn} - \frac{b_{ni}^2}{b_{ii}} \end{bmatrix} = M_i^{-1} \quad (52)$$

which proves (12).

ORCID iDs

Ahmed El-Mowafy  <https://orcid.org/0000-0001-7060-4123>

Kan Wang  <https://orcid.org/0000-0001-5688-6937>

References

Barrios J *et al* 2018 Update on Australia and New Zealand DFMC SBAS and PPP system results *Proc. ION GNSS+ 2018 (Miami, 24–28 September 2018)* pp 1038–67
 Blanch J, Gunning K, Walter T, De G L and Norman L 2019 Reducing computational load in solution separation for Kalman filters and an application to PPP integrity *Proc. 2019 Int. Technical Meeting of the Institute of Navigation (Reston, VA, January 2019)* pp 720–9

- Blanch J, Phelts E, Gunning K, Walter T, de Groot L and Norman L 2020 Bounding temporally correlated measurement noise with an application to GNSS integrity *Proc. 33rd ITM of the Institute of Navigation ION GNSS+ 2020 Virtual Meeting (22–25 September 2020)* pp 27–38
- Blanch J, Walter T and Enge P 2016 Gaussian bounds of sample distributions for integrity analysis *IEEE Trans. Aerosp. Electron. Syst.* **55** 1806–15
- Blanch J, Walter T, Enge P and Lee Y 2012 Advanced RAIM user algorithm description: integrity support message processing, fault detection, exclusion, and protection level calculation *Proc. ION GNSS (Nashville, TN, 18–21 September 2012)* pp 2828–49
- Blanch J, Walter T, Enge P and Lee Y 2015 Baseline advanced RAIM user algorithm and possible improvements *IEEE Trans. Aerosp. Electron. Syst.* **51** 713–32
- Du Y, Wang J, Rizos C and El-Mowafy A 2021 Vulnerabilities and integrity of precise point positioning for intelligent transport systems: overview and analysis *Satell. Navig.* **2** 1–22
- El-Mowafy A 2015 Estimation of multi-constellation GNSS observation stochastic properties using a single-receiver single-satellite data validation method *Surv. Rev.* **47** 99–108
- El-Mowafy A 2016 Pilot evaluation of integrating GLONASS, Galileo and BeiDou with GPS in ARAIM *Artif. Satell.* **51** 31–44
- El-Mowafy A 2017 Advanced receiver autonomous integrity monitoring using triple frequency data with a focus on treatment of biases *Adv. Space Res.* **59** 2148–57
- El-Mowafy A 2019 On detection of observation faults in the observation and position domains for positioning of intelligent transport systems *J. Geod.* **93** 2109–22
- El-Mowafy A, Cheung N and Rubinov E 2020 First results of the second generation SBAS in Australian Urban and Suburban Road Environments *J. Spatial Sci.* **65** 99–121
- El-Mowafy A, Deo M and Rizos C 2016 On biases in precise point positioning with multi-constellation and multi-frequency GNSS data *Meas. Sci. Technol.* **27** 035102
- El-Mowafy A, Imperato D, Rizos C, Wang J and Wang K 2019 On hypothesis testing in RAIM algorithms: generalized likelihood ratio test, solution separation test and a possible alternative *Meas. Sci. Technol.* **30** 075001
- Gunning K, Blanch J, Walter T, de Groot L and Norman L 2018 Design and evaluation of integrity algorithms for PPP in kinematic applications *Proc. 31st Int. Technical Meeting of the Satellite Division of the Institute of Navigation ION GNSS+ 2018 (Miami, 24–28 September 2018)* pp 1910–39
- Hassan T, El-Mowafy A and Wang K 2021 A review of system integration and current integrity monitoring methods for positioning in intelligent transport systems *IET Intell. Transp. Syst.* **15** 43–60
- Imparato D, El-Mowafy A, Rizos C and Wang J 2018 Vulnerabilities in SBAS and RTK positioning in intelligent transport systems: an overview *Proc. IGNSS Symp. 2018 (Sydney, 7–9 February 2018)* pp 1–18
- Jöerger M, Chan F C and Pervan B 2014 Solution separation versus residual-based RAIM *Navigation* **61** 273–91
- Kouba J and Héroux P 2001 Precise point positioning using IGS orbit and clock products *GPS Solut.* **5** 12–28
- Kouba J, Lahaye F and Tétreault P 2017 Precise point positioning *Handbook of Global Navigation Satellite Systems* ed P J G Teunissen and O Montenbruck (Berlin: Springer) pp 339–60
- Langel S, Crespillo O G and Jöerger M 2020 Overbounding sequential estimation errors due to non-gaussian correlated noise *Proc. 33rd ITM of the Institute of Navigation ION GNSS+ 2020 Virtual Meeting (22–25 September 2020)* pp 1054–67
- Parkinson B W and Axelrad P 1998 A basis for the development of operational algorithms for simplified GPS integrity checking *Proc. First Technical Meeting of the Satellite Division of the Institute of Navigation Colorado Springs Colorado* pp 269–76
- Powe M and Owen J 1997 A flexible RAIM algorithm *Proc. ION GPS 1997 (Kansas City, MS, 16–19 September 1997)* pp 439–49
- Rife J, Pullen S, Enge P and Pervan B 2006 Paired overbounding for nonideal LAAS and WAAS error distributions *IEEE Trans. Aerosp. Electron. Syst.* **42** 1386–95
- Tanil C, Khanafseh S, Joerger M and Pervan B 2018 Sequential integrity monitoring for kalman filter innovations-based detectors *Proc. 31st Int. Technical Meeting of the Satellite Division of the Institute of Navigation ION GNSS+ 2018, (Miami, FL, 24–28 September)* pp 2440–55
- Teunissen P J G and Kleusberg A 1998 *GPS for Geodesy* 2nd edn (New York: Springer)
- U.S. Department of Defense 2020 *Global Positioning System Standard Positioning Service Performance Standard* 5th edn (Washington DC: Office of the Department of Defense) pp 1–196 (available at: www.gps.gov/systems/gps/performance/)
- Walter T, Blanch J, Gunning K, Joerger J and Bervan B 2019 Determination of fault probabilities in ARAIM *IEEE Trans. Aerosp. Electron. Syst.* **55** 3505–16
- Walter T 2017 Satellite based augmentation systems *Handbook of Global Navigation Satellite Systems* ed P J G Teunissen and O Montenbruck (Berlin: Springer) pp 723–51
- Wang K, El-Mowafy A, Rizos C and Wang J 2020 Integrity monitoring for RTK positioning in road transport: new weighting model and overbounding CDF in open-sky and suburban scenarios *Remote Sens.* **12** 1–23
- Working Group C, EU-US 2016 *Cooperation on Satellite Navigation, ARAIM Technical Subgroup, Milestone 3 Report* (available at: www.gps.gov/policy/cooperation/europe/2016/working-group-c/)
- Zhang Q J and Schwarz K P 1996 Estimating double difference GPS multipath under kinematic conditions *Proc. Position, Location and Navigation Symp. PLANS'96* pp 285–91
- Zumberge J F, Heflin M B, Jefferson D, Watkins M M and Webb F H 1997 Precise point positioning for the efficient and robust analysis of GPS data from large networks *J. Geophys. Res.* **102** 5005–17