Vol. 18, No. 1, 2022

Journal of Global Positioning Systems

ISSN 1446-3156 (Print Version) ISSN 1446-3164 (CD-ROM Version)

International Association of Chinese Professionals in Global Positioning Systems (CPGPS)

Journal of Global Positioning Systems

Aims and Scope

The Journal of Global Positioning Systems is a peer-reviewed international journal for the publication of original research papers, review articles, invited contributions, also including selected papers presented at non-referred conferences and symposiums. Articles should present discussions of technologies and applications of any positioning systems, including Global Navigation Satellite Systems (GNSS) and their various augmentations and local components, and other positioning and navigation technologies, such as indoor positioning, and inertial, visual navigation systems and their integrations. Articles presenting advances of other related areas, such as wireless communications, intelligent vehicle systems, sensor networks, spatial information and geosciences are also welcome. Short research and technical notes, book reviews, lecture series and commercial advertisements can be included. Specific questions about the suitability of prospective manuscripts may be directed to the Editor-in-Chief.

Publication and Copyright

The Journal of Global Positioning Systems is an official publication of the International Association of Chinese Professionals in Global Positioning Systems (CPGPS). It is published twice a year, in June and December. The Journal is available in both print version (ISSN 1446-3156) and CD-ROM version (ISSN 1446-3164), which can be accessed through the CPGPS website at http://www.cpgps.org/journal.php. Whilst CPGPS owns all the copyright of all text material published in the Journal, the authors are responsible for the views and statements expressed in their articles. Neither the authors, the editors nor CPGPS can accept any legal responsibility for the contents published in the journal.

ISSN 1446-3156 (Print Version)/ISSN 1446-3164 (CD-ROM Version)

Journal of Global Positioning Systems
2022
© CPGPS

Journal of Global Positioning Systems

Published by
the International Association of Chinese Professionals
in Global Positioning Systems

Vol. 18, No. 1, 2022

Table of Contents

Articles

PhD Dissertation Abstracts

Journal of Global Positioning Systems (*2022*)
Vol. 18, No. 1: 1-14
DOI: 10.5081/jgps.18.1.1

Journal of Global Positioning Systems
2022
© CPGPS

An indoor magnetic field matching positioning solution based on consumer-grade IMU for smartphone

Jian Kuang[1], Taiyu Li[1] and Xiaoji Niu(✉)[1,2]
1. GNSS Research Center, Wuhan University, Wuhan, China
2. Artificial Intelligence Institute, Wuhan University, Wuhan, China
✉: **corresponding author, xjniu@whu.edu.cn**

Abstract: Magnetic field matching positioning (MFMP) has become one of the mainstream indoor positioning methods for mass application. However, the problem of the large workload of magnetic field mapping and the instability of the magnetometer bias remains to be solved. This paper designs an indoor MFMP scheme based on consumer-grade Inertial Measurement Units (IMUs). In the magnetic field mapping stage, the high-precision poses of the smartphone obtained by combining a foot-mounted IMU, a smartphone built-in IMU, and a few control points in the building are employed to generate a magnetic field grid map with high efficiency. In the real-time positioning stage, the relative trajectory generated by pedestrian dead reckoning (PDR) is used to improve the position discrimination of the magnetic field feature by adding spatial profile; and the differential magnetic field strength in the sensor frame (instead of in the reference frame) is used to achieve matching positioning that is immune to the magnetometer bias. The consistency of the magnetic field maps built using different smartphones show that the proposed magnetic mapping scheme achieves an average efficiency of 37 m^2/min and is not affected by the mapping trajectory and walking speed. The real-time positioning results using multiple smartphones show that the proposed MFMP algorithm is barely affected by the magnetometer bias, and can achieve an average RMS value of ±0.83 meters in a typical office scenario.

Keywords: Magnetic Matching; Pedestrian Dead Reckoning (PDR); Foot-mounted Inertial Measurement Unit (IMU); Indoor Positioning; Pedestrian Navigation

1 Introduction

The indoor geomagnetic field has the ubiquitous distortion feature due to the interference of steel materials in the building structure, which can be used for indoor positioning. Compared with the common indoor radio positioning signals (including Wi-Fi [1], Bluetooth [2], UWB [3], etc.), indoor magnetic field signals have the advantages of ubiquity, stability, and immunity from human body influence. Therefore, magnetic field matching positioning (MFMP) has become one of the mainstream indoor positioning methods for mass application [4, 5].

MFMP includes magnetic field map generation and real-time positioning parts [5, 6]. In the magnetic field map generation stage, the correlation between magnetic field features and geographic coordinates is established. Compared with the methods of point-by-point and crowdsourcing, the walking survey is the most widely used method for magnetic field map data collection by achieving a balance between accuracy and cost [6, 7]. The basic idea is the surveyors must walk along a straight line between two control points at a uniform speed, and the coordinates of the control points are determined by using professional measurement methods (e.g., total

Editor in charge: Dr. Jian Wang

station). However, the requirement of a straight trajectory will cause the number of control points to increase dramatically in complex indoor areas; at the same time, it is difficult for surveyors to ensure uniform straight-line walking. In general, the traditional walking survey still has the problems of heavy workload and unsure accuracy.

In the real-time positioning stage, the current position of the user is determined by calculating the similarity between the observed magnetic field feature coming from the magnetometer and the reference magnetic field feature in the magnetic field map. Dynamic time wrap (DTW) [8, 9] and particle filter (PF) [10, 11] are the two most frequently used methods in the published literature. The DTW method has the characteristics of a simple algorithm and stable positioning performance by transforming the MFMP problem into the similarity calculation of two magnetic field feature sequences. However, the real user's motion trajectories will be unpredictable in an open indoor area, then the DTW method will not work properly because there is no reference magnetic field feature sequence corresponding to the real-time user motion trajectory in the database [8]. The PF method sets a large number of particles to retain all possible positions of the user, uses the difference between the observed and the reference magnetic feature to filter out wrong particles, and obtains the current position of the user through continuous iterative calculation. The PF method has the advantages of high positioning accuracy and stable performance. However, the computational load of the PF method still needs to be further reduced for smartphones [11].

The positioning performance of the above-mentioned magnetic field matching algorithm depends on the pre-calibrated magnetometer bias. Smartphones are easily affected by nearby magnetic materials or strong currents, thereby the magnetometer bias needs to be calibrated and corrected frequently. Many researchers try to use the differential magnetic field feature in navigation coordinates (n-frame) to eliminate the effect of magnetometer bias [6, 10]. However, the assumption that the magnetometer bias is a fixed value in n-frame does not hold in the scenario of attitude angle fluctuation. In particular, since the user often shakes when using a smartphone and the walking route is irregular, the differential magnetic field feature in n-frame cannot achieve the purpose of eliminating the magnetometer bias.

Aiming at the above-mentioned typical problems of MFMP, this paper designs an indoor MFMP solution based on consumer-grade inertial measurement units (IMUs) as follow.

a) In the magnetic field map generation stage, a foot-mounted IMU, a handhold IMU (i.e., the smartphone built-in IMU) and combined with a few control points (e.g., as sparse as 50 meters between two adjacent control points) are used to provide the poses of the smartphone [12]. The collection of magnetic field map data can be achieved with high efficiency and sufficient precision by reducing the number of control points and releasing all user motion requirements.
b) In the real-time positioning stage, the relative trajectory generated by pedestrian dead reckoning (PDR) is used to improve the position discrimination of the magnetic field feature, and the *differential magnetic field feature in the sensor frame* is used to achieve the magnetometer bias irrelevant matching positioning.
c) Finally, multiple tests using various models of smartphones in real-world scenarios are conducted to verify the feasibility and localization performance of the proposed scheme.

The following content of this article is arranged as follows: Part 2 provides an overview of the proposed MFMP method, Part 3 describes the magnetic field map generation method based on the pedestrian positioning and orientation system in detail, and Part 4 describes the magnetic field matching and positioning algorithm in detail, Part 5 verifies the feasibility and effectiveness of the proposed magnetic field positioning scheme in the real environment, and Part 6 summarizes and concludes.

2 Architecture of the magnetic field positioning solution

The position error of the pure inertial navigation system (INS) based on consumer-grade IMUs will reach several meters within a few seconds, which is far from the needs of practical applications. Then, it is necessary to extract the constraint information formed by the pedestrian motion law for improving the relative positioning ability of the INS. The foot-mounted IMU utilizes the fact that the feet of pedestrians periodically contact the ground to obtain a very powerful relative positioning capability. This is the fundamental reason why the combination of foot- mounted IMU and hand-held IMU can be used to efficiently collect magnetic field vector maps [12]. At the same time, the pedestrian walking pattern is employed for controlling the velocity error of the smartphone built-in IMU-based INS, which provides reliable relative position and attitude for improving the stability and positioning accuracy of MFMP [13]. In general, the consumer-grade IMUs play an indispensable role as the auxiliary means for the smartphone-based indoor MFMP solution.

Fig. 1 shows the flow of the MFMP scheme based on consumer-grade IMUs. The scheme can be divided into two parts:

1) Magnetic field map generation stage. A pedestrian positioning and orientation system (P-POS) composed of a foot-mounted IMU, a smartphone built-in IMU, and a few control points is used to provide high-precision pose. On this basis, a linear interpolation method is used to generate a high-precision grid map of the magnetic field vector.

2) Online positioning stage. The relative position and attitude coming from the pedestrian dead reckoning (PDR) algorithm are used for correlating the observed magnetic field strength to form a magnetic field profile [12]. Additionally, a constructed differential magnetic field profile in the sensor frame (b-frame) is used to eliminate the influence of the magnetometer bias and provide stable positioning results.

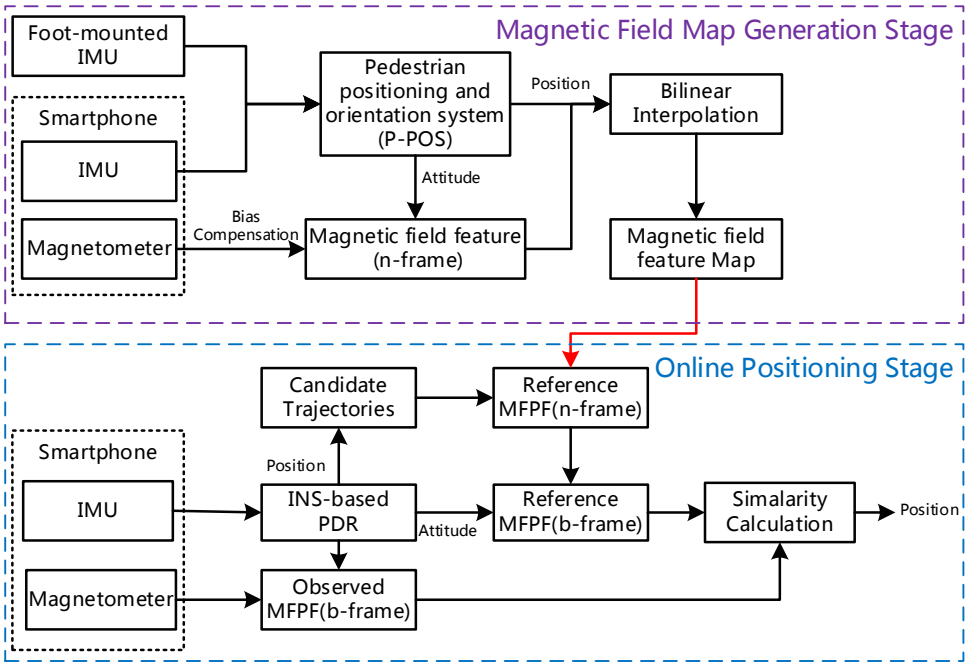


Fig. 1. The flow of the MFMP scheme based on consumer-grade IMUs.

3 Magnetic Field Map Generation Stage

The magnetic field map is the basis of the matching positioning scheme. The magnetic field map generation method is different depending on the different application requirements, which is essentially a compromise between measurement accuracy and measurement efficiency/cost. This paper uses the P-POS-based walking survey method

to collect the magnetic field map data with high-efficiency. The proposed method reduces the number of control points and releases the movement demand of the data collectors by utilizing the superior relative positioning capability of the foot-mounted inertial navigation. After data collection, a linear interpolation method is used to generate a uniformly distributed grid map of the magnetic field vector.

Pedestrian Positioning and Orientation System (P-POS)

Fig. 2 shows the hardware setup of P-POS, including a foot-mounted IMU and a smartphone built-in IMU. P-POS can be divided into Foot-INS (Foot-mounted IMU-based Inertial Navigation System) and Foot-INS/IMU integrated. Foot-INS is a typical pedestrian dead reckoning algorithm, and the positioning error will continue to accumulate with the walking distance, the specific algorithm can be found in [14]. Foot-INS assumes that the feet of a pedestrian will periodically come into contact with the ground, that is, there is a short static state within a cycle of the footstep, as shown in Fig. 3. In the stationary state, the speed of the foot can be considered zero, which can greatly reduce the position drift error of the INS. Test results in many documents show that the positioning performance of Foot-INS based on consumer-grade IMU fluctuates greatly, and the typical relative positioning accuracy is 0.3~3% of the total walking distance [12, 15]. To control the position drift error of Foot-INS, this scheme introduces control points (i.e., the coordinates are known) to periodically correct the position drift error of Foot-INS, and uses a reverse smoothing algorithm to further improve the positioning accuracy.

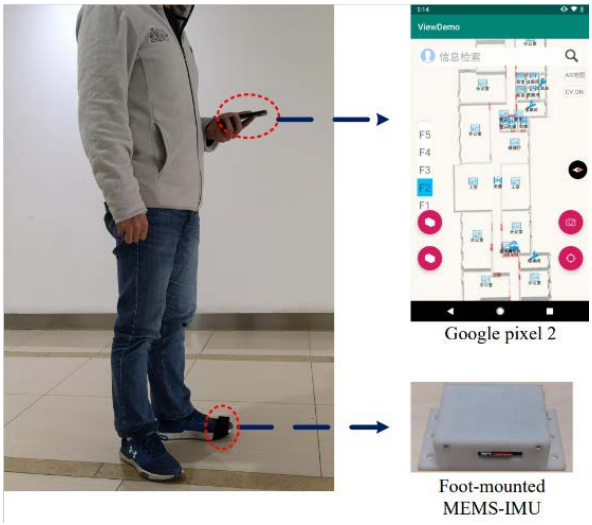


Fig. 2. The hardware setup of P-POS, including a foot-mounted IMU and a handhold smartphone built-in IMU (near the waist).

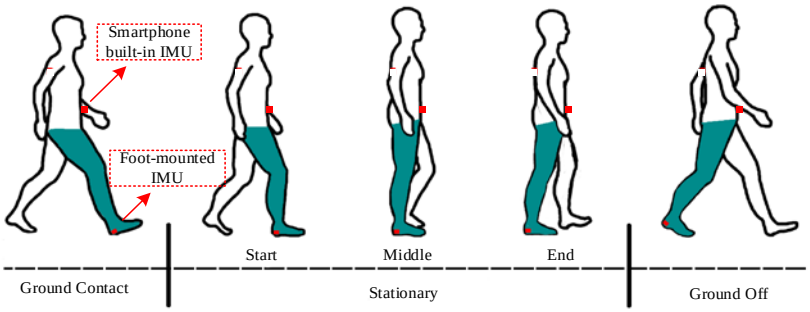


Fig. 3. Foot cycle in normal walking. The relative positional relationship between the foot and the waist in the middle of the stationary period in a step cycle that can be regarded as fixed.

Due to the large difference in magnetic field features between the foot and the waist, the trajectory coordinates generated by Foot-INS cannot be directly used to generate a magnetic field map. To obtain the precise pose of the smartphone, the position estimated by Foot-INS is used to correct the position error of the smartphone built-in IMU. This is the core idea of the Foot-INS/IMU combination, which is similar to the GNSS/INS integrated positioning. However, the relative positional relationship between the foot and the waist is constantly changing (as shown in Fig. 3), and the position of the foot cannot

be considered equal to the position of the waist. After a thorough analysis of pedestrian walking, we found that the relative positional relationship between the foot and the waist in the middle of the stationary period in a single step cycle can be regarded as fixed. Then, the estimated position of Foot-INS can be accurately projected to the waist. [12] gives a detailed algorithm description, and its results show that the estimation accuracy of the position and attitude of the smartphone under the condition that the distance between adjacent control points is 50 meters can reach the decimeter level and the degree level, respectively.

Magnetic Field Grid Map Generation

As data collectors cannot guarantee uniform and high-density collection of magnetic field features, this paper uses the bilinear interpolation method to generate a magnetic field grid map. Many research works have shown that rigorous theoretical interpolation methods can achieve higher accuracy than bilinear interpolation[16]. However, the accuracy improvement of magnetic field maps is very limited for pedestrian localization scenarios. The magnetic field map generation includes two steps: rasterization and linear interpolation.

Rasterization: 1) Divide the test area into a uniform grid in the east-west direction and north-south direction. 2) Mark the grid where each magnetic field feature is located according to the estimated smartphone coordinates. 3) Average all the magnetic field features in a single grid.

Linear interpolation: 1) Determine the position coordinates of the grid to be interpolated (e.g., grid No. 0 in Fig. 4), and set the search radius (e.g., 1m) of the effective grid. 2) Traverse the eight directions (i.e., east, south, west, north, northwest, northeast, southeast, and southwest) of grid No. 0. The grid with a valid magnetic field feature will be marked available (e.g., the grids numbered 1~7 in Fig. 4). 3) The grid No. 3 will be eliminated based on the fact that the magnetic field features can only be interpolated; 4) The other valid grids are used for obtaining the magnetic field features of grid No. 0 by linear interpolation. The linear interpolation formula can be expressed as

$$\mathbf{M}_0^n = \frac{\mathbf{M}_{1,5}^n + \mathbf{M}_{2,6}^n + \mathbf{M}_{4,7}^n}{3} \tag{1}$$

where

$$\mathbf{M}_{1,5}^n = \frac{d_{5-0}\mathbf{M}_1^n + d_{1-0}\mathbf{M}_5^n}{d_{1-0} + d_{5-0}}, \quad \mathbf{M}_{2,6}^n = \frac{d_{6-0}\mathbf{M}_2^n + d_{2-0}\mathbf{M}_6^n}{d_{2-0} + d_{6-0}},$$

$$\mathbf{M}_{4,7}^n = \frac{d_{7-0}\mathbf{M}_4^n + d_{4-0}\mathbf{M}_7^n}{d_{4-0} + d_{7-0}}, \ \mathbf{M}_i^n$$ is the magnetic field feature in n-frame of the i-th grid, $d_{i-0} = \sqrt{(x_i - x_0) + (y_i - y_0)}$ is the distance between the i-th grid and the 0-th grid. In addition, the bias compensation is necessary for the observed magnetic field signal accurately reflects the real environmental magnetic field. And the simple ellipsoid fitting method is employed for calibrating the magnetometer bias [17].

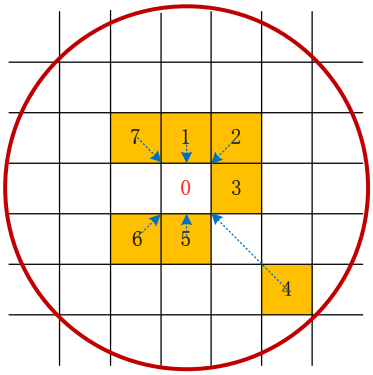


Fig. 4. Obtain a magnetic field feature by linear interpolation. The presence of two valid magnetic field features in the same direction will only be used for the interpolation method.

Compared with the traditional walking survey method, the proposed P-POS-based method has the advantages of high efficiency and high precision. This proposed solution allows the surveyor to walk on a trajectory of any shape at any walking speed, which greatly simplifies the data collection process. At the same time, the demand for the number of control points is greatly reduced, so the workload of surveyors to mark control points and the risk of incorrect control point correction are reduced. In addition, the proposed solution can provide the accurate smartphone attitude for expanding the

magnetic field map feature from 1~2 dimensions to 3 dimensions at a low cost, which supports the projection of the reference magnetic field feature from n-frame to b-frame, thus forming a matching algorithm independent of the magnetometer bias, as will mentioned in the next section.

4 Real-time Positioning Stage

The relative trajectory and attitude generated by PDR play a very important role in our MFMP algorithm. And the relative trajectory is used to correlate the magnetic field features to form a magnetic field profile, which helps to improve the position discrimination of the magnetic field features; the attitude is used to project the reference magnetic field profile from n-frame to b-frame, thereby constructing a matching algorithm independent of magnetometer bias.

Magnetic Field Profile

A magnetic field feature at a single location has low dimensionality (e.g., at most 3 dimensions) and is far from sufficient to provide accurate MFMP. Therefore, magnetic field feature time sequence (MFFTS) is widely used as an improvement scheme. As shown in Fig. 5, the user has passed through four positions A, B, C, D, etc., and combined the magnetic field features of the four points to form a magnetic field feature time sequence. However, the MFFTS still does not fully exploit the strong correlation between magnetic field features and spatial position. Therefore, this paper uses the relative trajectory generated by the PDR to correlate the MFFTS to give the relative spatial topological properties between the four positions (such as the azimuth and distance between the next point and the previous point), thereby further improving the position discrimination of the magnetic field features, called the magnetic field profile (MFP).

The INS-based PDR (INS-PDR) algorithm is used to provide the relative trajectory, which the detailed algorithm can be found in [13]. The reason to use INS-PDR is that it can achieve more robust positioning performance as the regular step-model-based PDR, while INS-PDR can also provide 3D position, velocity, and attitude at higher frequencies (e.g., 20 Hz). Based on the relative position and attitude output by the PDR algorithm, the observed magnetic field profile can be expressed as

$$\mathbf{MFP}_{obs} = \begin{Bmatrix} \mathbf{r}_1^n & \left(\mathbf{C}_b^n\right)_1 & \tilde{\mathbf{M}}_1^b \\ & \cdots & \\ \mathbf{r}_k^n & \left(\mathbf{C}_b^n\right)_k & \tilde{\mathbf{M}}_k^b \end{Bmatrix} \tag{2}$$

where $\mathbf{r}^n$ is the plane position in n-frame, $\mathbf{C}_b^n$ is the cosine matrix of the direction from b-frame to n-frame, provided by the INS-PDR algorithm, $\tilde{\mathbf{M}}^b$ is the raw observation of the magnetometer, and k is the length of a magnetic field profile. Since high sampling rate data will bring a lot of useless calculations, we use the magnetic field profile with a sampling rate of 2Hz to reduce the computational load of the matching algorithm while ensuring positioning accuracy.

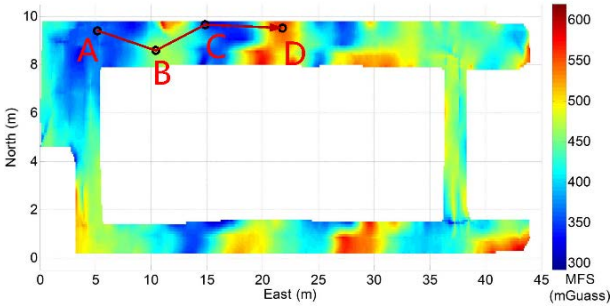


Fig. 5. Magnetic profile on magnetic field map. A magnetic field profile consists of magnetic field time series and corresponding relative trajectories.

Magnetic Field Profile Based Matching Algorithm

The basic condition for MFMP to be feasible is that the magnetic field features observed by different users on the same path are consistent. Thus, the problem of MFMP can be simplified to find the conversion relation between relative trajectory and absolute trajectory. And the conversion relationship mainly includes translation and rotation. Since the conversion relationship and absolute trajectory coordinates are parameters to be solved, they cannot be directly obtained through mathematical analysis. Therefore, we generate all possible reference

trajectories using a parametric search method and compare the similarity of the observed MFP with the reference MFP to determine the conversion relationship. Fig. 6 shows the specific process of generating candidate reference trajectory. 1) Obtain the relative trajectory S based on the relative position sequence calculated by PDR, and rotate the trajectory S around its initial point $\Delta\theta$ to obtain the trajectory S'. 2) Translate the trajectory S' in the north-south direction Δn to obtain the trajectory S'', 3) Translate the trajectory S'' along the east-west direction translate Δe to obtain trajectory S'''.

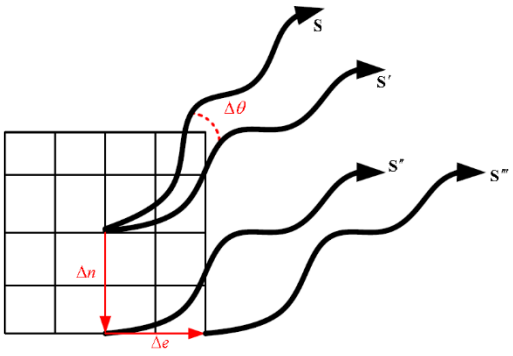


Fig. 6. The generation process of multiple candidate trajectories. All possible candidate trajectories are generated by traversing the translation parameter $(\Delta n, \Delta e)$ and rotation parameter $\Delta\theta$.

The candidate trajectory S''' can be expressed as:

$$\mathbf{r}_j^{n'} = \mathbf{C}(\Delta\theta)\left(\mathbf{r}_j^n - \mathbf{r}_1^n\right) + \mathbf{r}_1^n + \Delta\mathbf{r}^n \tag{3}$$

where $\Delta\mathbf{r}^n = \begin{bmatrix}\Delta n & \Delta e\end{bmatrix}^T$,

$$\mathbf{C}(\Delta\theta) = \begin{bmatrix}\cos(\Delta\theta) & -\sin(\Delta\theta)\\ \sin(\Delta\theta) & \cos(\Delta\theta)\end{bmatrix},$$

$\mathbf{r}_j^{n'}$ is the coordinates of the j-th point of the candidate reference trajectory. The corresponding direction cosine matrix $\mathbf{C}_b^n$ also needs to be adjusted accordingly

$$\left(\mathbf{C}_b^{n'}\right)_j = \mathbf{C}_n^{n'}\left(\mathbf{C}_b^n\right)_j \tag{4}$$

where $\left(\mathbf{C}_b^{n'}\right)_j$ is the directional cosine matrix from b-frame to n-frame corresponding to the j-th point of the candidate reference trajectory and

$$\mathbf{C}_n^{n'} = \begin{bmatrix}\cos(\Delta\theta) & -\sin(\Delta\theta) & 0\\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0\\ 0 & 0 & 1\end{bmatrix}.$$

Since the magnetic field map is composed of uniformly distributed reference points, the sampling points of the candidate reference trajectory cannot be exactly coincident with the reference points. Therefore, the bilinear interpolation method is used to obtain the reference magnetic field feature with higher resolution [5], as shown in Fig. 7.

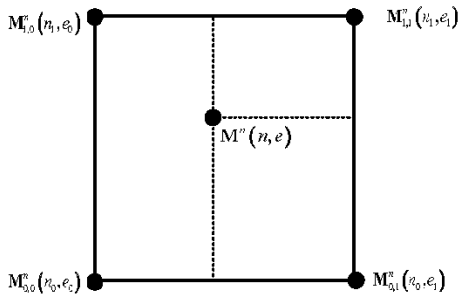


Fig. 7. Reference magnetic feature at (n, e) from the bilinear interpolation method.

A given point coordinate $(n\ e)$, its corresponding reference magnetic field feature is

$$\mathbf{M}^n \approx \alpha_1\mathbf{M}^n_{0,1} + \alpha_2\mathbf{M}^n_{0,0} + \alpha_3\mathbf{M}^n_{1,1} + \alpha_4\mathbf{M}^n_{1,0} \tag{5}$$

where

$$\alpha_1 = \frac{(n_1-n)(e-e_0)}{(n_1-n_0)(e_1-e_0)}, \alpha_2 = \frac{(n_1-n)(e_1-e)}{(n_1-n_0)(e_1-e_0)}$$
$$\alpha_3 = \frac{(n-n_0)(e-e_0)}{(n_1-n_0)(e_1-e_0)}, \alpha_4 = \frac{(n-n_0)(e_1-e)}{(n_1-n_0)(e_1-e_0)}$$

The reference magnetic field profile can be expressed as

$$\mathbf{MFP}_{ref} = \begin{Bmatrix}\mathbf{r}_1^{n'} & \left(\mathbf{C}_b^{n'}\right)_1 & \mathbf{M}_1^b\\ & \cdots & \\ \mathbf{r}_k^{n'} & \left(\mathbf{C}_b^{n'}\right)_k & \mathbf{M}_k^b\end{Bmatrix} \tag{6}$$

where $\mathbf{M}_1^b = \left(\mathbf{C}_b^{n'}\right)_1^T \mathbf{M}_1^n$ is the reference magnetic field profile in b-frame.

Due to the magnetometer bias being a constant value in a short time (e.g., 15 seconds), the

differential MFP in b-frame is used for eliminating the influence of the magnetometer bias. However, the large noise characteristic of the smartphone built-in magnetometer affects the selection of the reference point for the calculation of the differential MFP. The proposed solution performs de-averaging processing on the observed MFP and the reference MFP respectively. And the DTW algorithm is used to calculate the similarity between the observed MFP and the reference MFP, the latest position of the reference trajectory with the highest similarity is the estimated user position. DTW compresses or stretches the reference axis of the two sequences to be matched so that two sequences with different lengths have better matching results. This will help solve the issue that the PDR algorithm cannot accurately estimate the pedestrian step length.

5 Test Results and Analysis

Test Description

The test scene is a typical office building scene with size of 94m×22m. Fig. 8 shows the specific indoor structure, and the red dash line marks the test area. To evaluate the performance of the real-time magnetic field matching positioning algorithm, this paper uses an offline method to analyze the positioning accuracy, and the reference truth position is provided by P-POS. Table 1shows the error parameters of all the IMUs involved in this paper.

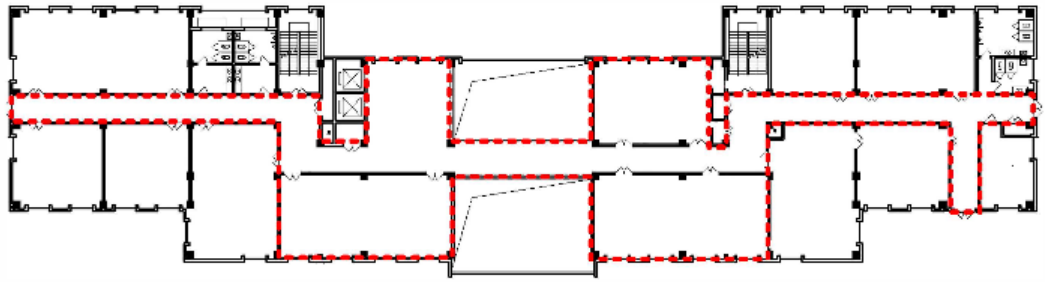


Fig. 8. Indoor structure. The red box is the test area.

Table 1. Error Parameters of the IMUs

Parameter	Foot-mounted IMU	Honor V10	Google pixel2 & Mi8	Google pixel3
Gyroscope noise (ARW)	0.004 °/s /√Hz	0.005 °/s /√Hz	0.007 °/s /√Hz	0.0038 °/s /√Hz
Gyroscope bias change vs. temperature	0.01°/s/°C	0.05°/s/°C	0.05°/s/°C	0.01°/s/°C
Accelerometer noise (VRW)	0.1 mg/√Hz	0.13 mg/√Hz	0.3 mg/√Hz	0.13 mg/√Hz
Accelerometer bias change vs. temperature	1 mg/°C	0.5 mg/°C	1 mg/°C	0.1 mg/°C

5.2 Magnetic Field Map

Three smartphones (such as Google pixel2, pixel3, and Mi8) are employed to generate the magnetic field map of the test area, which take 13.3, 13.7, and 12.8 minutes, respectively. And the effective area is about 500 m^2, so the average data collection efficiency of the magnetic field map is about 37m^2/min. If an indoor open area with an area of 100 m^2 is tested, based on the "S" path method and calculated at a pedestrian speed of 1.2 m/s, the data collection efficiency will reach 55 m^2/min. Compared with the data collection efficiency (i.e., 30 m^2/min) of the state-of-the-art conventional method in [18], the proposed method almost doubled the efficiency. Moreover, the measurement efficiency evaluation of traditional methods usually ignores the workload of coordinate measurement of control points and re-measurement caused by the error of the surveyors.

Therefore, the proposed method can achieve much higher measurement efficiency than traditional methods in practical applications due to fewer control points and simpler operation procedures.

Due to the lack of a high-precision magnetic field map reference, we use the within-accuracy coincidence of the magnetic field map to evaluate the performance of the proposed method. Fig. 9 shows the magnetic field map generated using Google pixel2. The vertical axis and the horizontal axes are the north and east positions respectively, and the colors represent the values of the magnetic field feature. Subgraphs (a) ~ (c) are the north, east and down components. The magnetic field features transition smoothly with the change of the region, which is in line with the real case. Fig. 10 shows the difference between the magnetic field map using Mi-8 and Google pixel2. The differences in the magnetic field map in most areas are distributed around 0 milligauss (mGauss), indicating that the P-POS provides position and attitude with good repeatability. In addition, the difference reached about 100 mGauss in some areas. This is because the accuracy of P-POS is limited (such as decimeter level) and there is a position projection error caused by the relative position change between the smartphone and the Foot-INS. In addition, the magnetic field feature decay with the 3rd power of the spatial distance [19], small position and attitude errors will cause obvious magnetic field feature deviation when approaching a magnetic field interference source.

To present the consistency of the magnetic maps in a quantitative way, Fig. 11 shows the cumulative density function of the difference in the magnetic field maps using two different smartphones. Table 2 shows the root mean square (RMS), 68%, and 95% of the difference in the magnetic field maps using two different smartphones. The difference in the three directions of the magnetic field maps using any two smartphones is less than 20 mGauss (RMS), and 95% of the difference is less than 40 mGauss. Compared with the noise level of magnetometers built-in most smartphones are about ±10~±20 mGauss, and the errors caused by P-POS and map generation algorithms are very small. Therefore, the proposed magnetic field map generation method has the characteristics of high efficiency and high precision

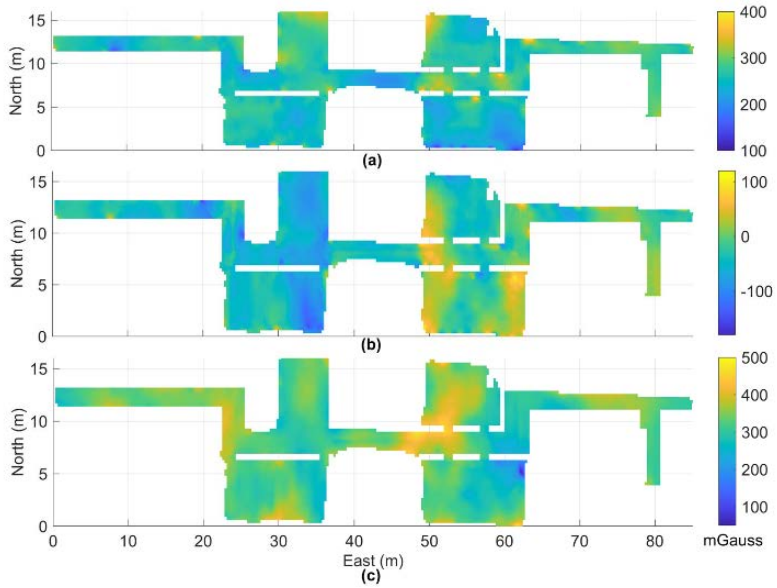


Fig. 9. Magnetic field map using Google pixel2. (a) North, (b) East, (c) Down

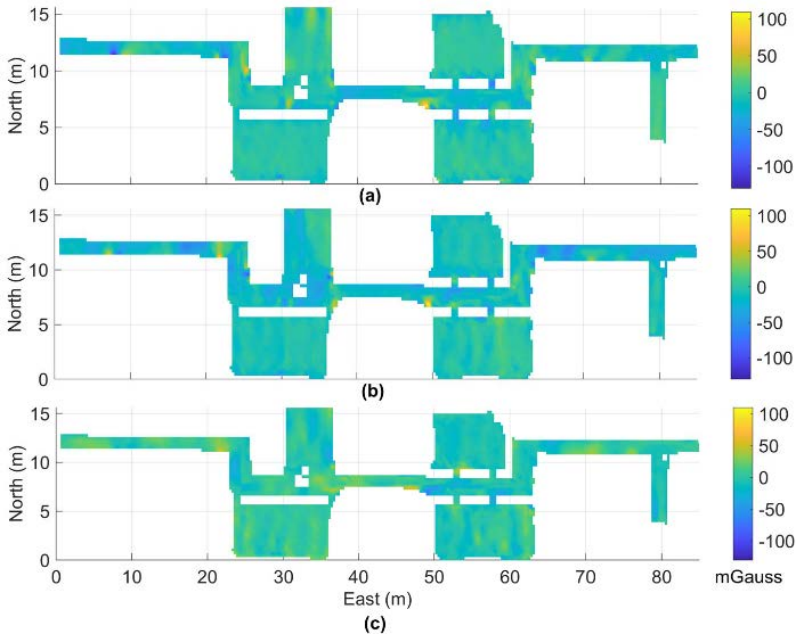


Fig. 10. The difference of two magnetic field maps using Mi 8 and Pixel 2. (a) North, (b) East, (c) Down.

Table 2. Root mean square, 68% and 95% of the difference of magnetic field maps using different smartphones (unit: mGauss)

	Mi8- Pixel3		Mi8-Pixel2		Pixel2- Pixel3	
	RMS	68%/95%	RMS	68%/95%	RMS	68%/95%
North	±16.0	±12.1/±32.3	±13.0	±10.1/±25.2	±18.4	±16.2/±35.6
East	±19.4	±16.7/±39.1	±14.7	±13.1/±28.7	±19.1	±18.3/±36.3
Down	±13.6	±11.6/±27.7	±17.0	±15.8/±32.6	±18.4	±17.6/±36.2

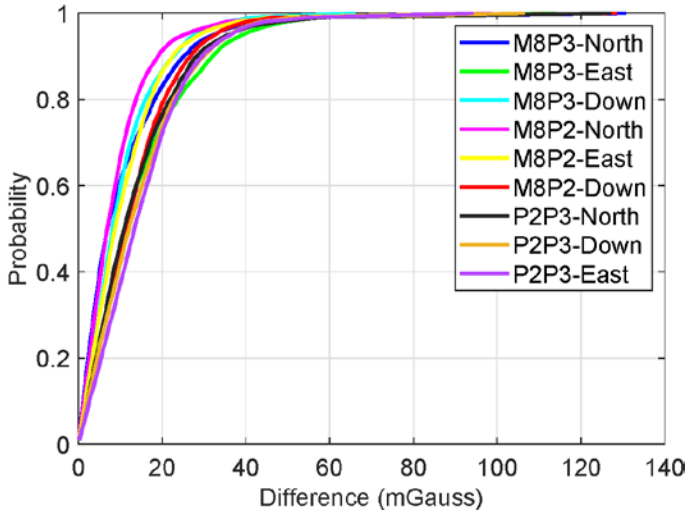


Fig. 11. The cumulative density function of the difference of three generated magnetic field maps using two different smartphones

5.3 Real-time Positioning Performance Analysis

Since the magnetic field features do not have the global positioning capability, MFMP is usually used as an auxiliary positioning method. Here, the initial position is manually given, and WiFi/Bluetooth can be used to give a rough position (such as the position error less than 10m) for the real-time positioning

application. As the proposed algorithm is insensitive to the magnetometer bias, so the real-time positioning algorithm evaluation stage will no longer need to perform bias compensation on the magnetometer observations.

Based on the magnetic field map generated with Pixel2, we conducted 8 tests using 4 different models of smartphones (including Honor V10, Google Pixel2, Pixel3, and Mi8). To be consistent with the user movements in the real scene, the testers walked on a straight trajectory in corridor area and walk on an irregular curved trajectory in open area. Fig. 12 shows the trajectories of these 8 tests. The red line is the reference truth trajectory (given by P-POS); the blue line is the PDR trajectory, and the green line is the trajectory of MFMP. The sub-pictures (a) and (b) correspond to Honor V10, (c) and (d) correspond to Google Pixel3, (e) and (f) correspond to Mi 8, (g) and (h) correspond to Google Pixel2. The trajectories generated by the PDR in all tests have different scale and deformation error, and the trajectories of MFMP has a good degree of coincidence with the reference trajectories. We can learn that the relative trajectory and attitude with errors provided by PDR can be used to improve the performance of magnetic field matching.

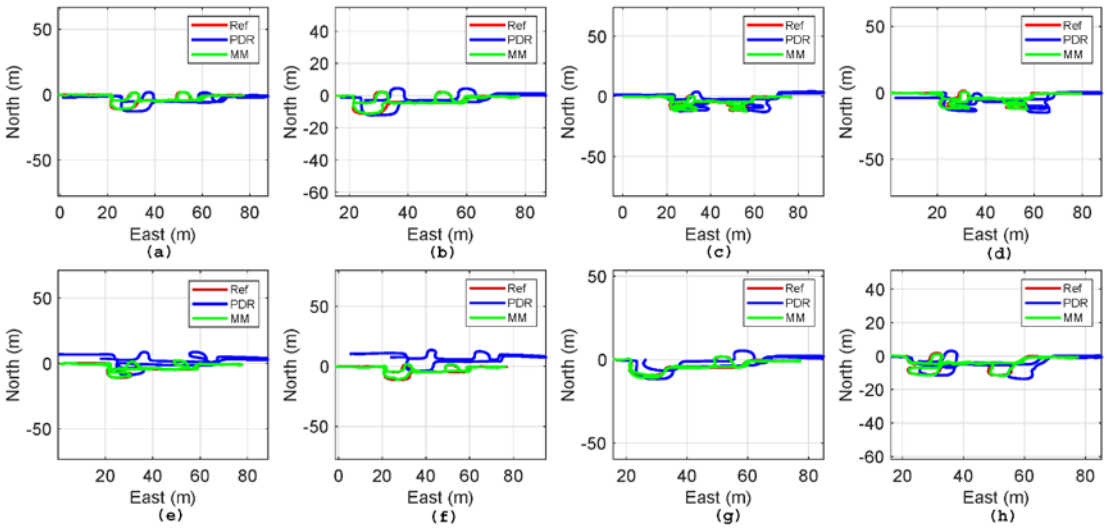


Fig. 12. The trajectories of 8 tests using 4 smartphones. (a) and (b) correspond to Honor V10, (c) and (d) correspond to Google Pixel3, (e) and (f) correspond to Mi 8, (g) and (h) correspond to Google Pixel2

Fig. 13 shows the cumulative density function of position error for 8 tests, and "V10-1" is the first test of Honor V10. The positioning errors are relatively concentrated and most of which are within 1.5 meters. Table 3 summarizes the RMS, 68% and 95% of the position error for 8 tests. The RMS of the position errors of the 8 tests distribute in ±0.67~±1.01m, and the fluctuation range (i.e., 0.34m) of the positioning error is smaller than the length of a pedestrian step (about 0.6 meters). We can learn that the MFMP algorithm designed in this paper is not sensitive to the magnetometer bias, and the positioning performance difference between multiple smartphones is also small. The mean value of the RMS, 68% and 95% are ±0.83m, ±0.79m, and ±1.60m, respectively, which shows that the method proposed in this paper can reach the meter-level/sub-meter-level positioning accuracy.

6 Conclusion and Outlook

This study proposed an indoor magnetic field matching positioning solution of smartphone based on consumer-grade IMU. The solution has greatly improved the relative positioning and attitude estimation ability of the consumer-grade IMUs by using the constraint information formed by the pedestrian movement characteristics. The magnetic field map construction efficiency and real-time positioning stability have consequently been enhanced significantly.

Table 3. RMS: 68% and 95% of positioning error of 8 tests (unit: m)

Test	V10-1	V10-2	Pix3-1	Pix3-2	Mi8-1	Mi8-2	Pix2-1	Pix2-1	Mean
RMS	±0.67	±0.79	±1.01	±0.99	±0.78	±0.87	±0.74	±0.84	±0.83
68%	±0.66	±0.74	±1.00	±0.88	±0.78	±0.86	±0.70	±0.72	±0.79
95%	±1.27	±1.53	±2.10	±1.89	±1.48	±1.43	±1.32	±1.77	±1.60

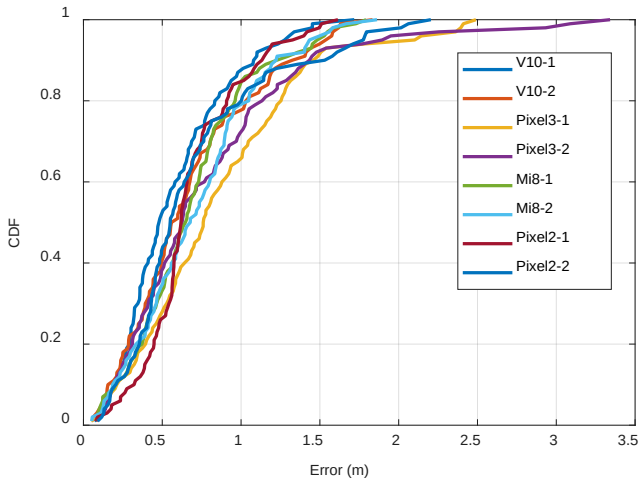


Fig. 13. Cumulative density function of position error of 8 tests

In the magnetic field map generation stage, a P-POS is used to provide decimeter-level positioning and degree-level attitude (including roll, pitch, and heading) of the smartphone. The test results of using three smartphones to generate magnetic field maps show that the data collection efficiency of the proposed method has reached 37m^2/min and 55m^2/min maximum, and the inconsistency of the magnetic field maps using different smartphones is less than 20 mGauss (RMS).

In the real-time positioning stage, the position and attitude provided by PDR are used to improve the position discrimination of the magnetic field features and obtain the transformation relationship from the navigation frame to the sensor frame, so that the differential MFP in b-frame can be used for eliminating the impact of the magnetometer bias. The results from 8 field tests using 4 models of smartphones showed the positioning errors between ±0.67 and ±1.01 meters, reaching an average RMS value of ±0.83 meters. The experimental results have completely verified that the MFMP method proposed in this study is immune to the magnetometer bias, and there is no significant difference in positioning performance between different models of smartphones.

Because the smartphone-based indoor MFMP scheme proposed in this study is highly dependent on the stability of the PDR, we will focus on automatically monitoring the integrity of the PDR and adapting it to a variety of typical smartphone usage modes in the future, such as texting, calling, and swinging. More importantly, we will explore the method of generating magnetic field maps based on crowdsourced data to further reduce the cost of the whole solution.

Remark: This paper has been originally published in the proceeding of China Satellite Navigation Conference (CSNC 2021).

References:

[1]. Zhuang, Y. and N. El-Sheimy, Tightly-Coupled Integration of WiFi and MEMS Sensors on Handheld Devices for Indoor Pedestrian Navigation. IEEE Sensors Journal, 2015. 16(1): p. 224-234.

[2]. Zhuang, Y., et al., Smartphone-based indoor

localization with bluetooth low energy beacons. Sensors, 2016. 16(5): p. 596.
[3]. Van Herbruggen, B., et al., Wi-PoS: A Low-Cost, Open Source Ultra-Wideband (UWB) Hardware Platform with Long Range Sub-GHz Backbone. Sensors, 2019. 19(7): p. 1548.
[4]. Lin, Y., et al., Autonomous aerial navigation using monocular visual-inertial fusion. Journal of Field Robotics, 2018. 35(1): p. 23-51.
[5]. Kuang, J., et al., Indoor Positioning Based on Pedestrian Dead Reckoning and Magnetic Field Matching for Smartphones. Sensors, 2018. 18(12): p. 4142.
[6]. Li, Y., Integration of MEMS Sensors, WiFi, and Magnetic Features for Indoor Pedestrian Navigation with Consumer Portable Devices. 2016, University of Calgary.
[7]. Jung, S.H., B.C. Moon and D. Han, Performance Evaluation of Radio Map Construction Methods for Wi-Fi Positioning Systems. IEEE Transactions on Intelligent Transportation Systems, 2017. PP(99): p. 1-10.
[8]. Shu, Y., et al. Last-Mile Navigation Using Smartphones. in International Conference on Mobile Computing and NETWORKING. 2015.
[9]. Subbu, K.P., B. Gozick and R. Dantu, LocateMe- Magnetic Fields Based Indoor Localization Using Smartphones. ACM Transactions on Intelligent Systems and Technology, 2013. 4(4): p. 1-27.
[10]. Kim, B. and S. Kong, Indoor Positioning based on Bayesian Filter using Magnetometer Measurement Difference, in IEEE Vehicular Technology Conference Proceedings. 2015.
[11]. Shi, L., M. Yu and W. Yin, PDR/Geomagnetic Fusion Localization Method Based on AOFA Improved Particle Filter. IEEE Transactions on Instrumentation and Measurement, 2021.
[12]. Niu, X., et al., A Novel Position and Orientation System for Pedestrian Indoor Mobile Mapping System. IEEE Sensors Journal, 2021. 21(2): p. 2104-2114.
[13]. Kuang, J., X. Niu and X. Chen, Robust Pedestrian Dead Reckoning Based on MEMS-IMU for Smartphones. Sensors, 2018. 18(5): p. 1391.
[14]. Foxlin, E., Pedestrian tracking with shoe-mounted inertial sensors. IEEE COMPUTER GRAPHICS AND APPLICATIONS, 2005. 25(6): p. 38-46.
[15]. Liu, T., et al., A Simple Positioning System for Large-Scale Indoor Patrol Inspection Using Foot-Mounted INS, QR Code Control Points, and Smartphone. IEEE Sensors Journal, 2021. 21(4): p. 4938-4948.
[16]. Hensel, S., et al., Application of Gaussian Process Estimation for Magnetic Field Mapping. 2021, Springer International Publishing: Cham. p. 284-298.
[17]. Tabatabaei, S., A. Gluhak and R. Tafazolli, A Fast Calibration Method for Triaxial Magnetometers. IEEE Transactions on Instrumentation & Measure..., 2013.
[18]. Liu, G., et al., Human-interactive Mapping Method for Indoor Magnetic Based on Low-cost MARG Sensors. IEEE Transactions on Instrumentation and Measurement, 2021: p. 1-1.

Authors

Jian Kuang is an Associate Researcher Fellow of GNSS Research Center at Wuhan University in China. He got his Ph.D. and bachelor degrees from Wuhan University in 2019 and 2013 respectively. Her research interest includes pedestrian navigation, vehicle positioning, magnetic field matching, and multi-system fusion positioning.

Taiyu Li received the B.Eng. degree in Geodesy and Survey Engineering from China University of Mining and Technology, Xuzhou, China, in 2019. He is currently a Master student in the GNSS Research Center, Wuhan University, Wuhan, China. His research interests focus on pedestrian navigation and indoor positioning.

Dr. Xiaoji Niu is a Professor of GNSS Research Center at Wuhan University in China. He got his Ph.D. and bachelor degrees from Tsinghua University in 2002 and 1997 respectively. He did post-doctoral research at the University of Calgary and worked as a senior scientist in SiRF Technology Inc. Dr. Niu leads a multi-sensor navigation group focuses on GNSS/INS integrations, low-cost navigation sensor fusion, and the relevant new applications. Dr. Niu has published 100+ academic papers and own 30+ patents.

Journal of Global Positioning Systems (*2022*)
Vol. 18, No. 1: 15-27
DOI: 10.5081/jgps.18.1.15

Journal of Global Positioning Systems
2022
© CPGPS

Estimation and analysis of BDS-3 multi-frequency differential code bias using MGEX observations

Haijun Yuan, Zhuoming Hu, Xiufeng He(✉) and Zhetao Zhang
School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
✉ **Corresponding author:** Xiufeng He, xfhe@hhu.edu.cn

Abstract: Differential code bias (DCB) significantly affects the ionosphere modeling, precise positioning, and navigation applications when using code observations. With the fully completed BeiDou navigation satellite system (BDS-3), there exist various DCBs of new frequencies and types which should be handled. However, limited types of DCB products for BDS-3 are provided by the analysis institutions (e.g., Chinese Academy of Science (CAS) and German Aerospace Center (DLR)). Hence, for some DCB corrections of new frequencies, they are generally generated by complex linear combinations, which are not friendly to users and may degrade the accuracy. In this study, the estimation method of DCB for BDS-3 is introduced first. Then, the BDS-3 observations from 40 globally distributed stations are selected to estimate the DCBs, including 19 types of DCBs of new frequencies for BDS-3. Moreover, the estimated DCBs are carefully analyzed in terms of inner consistency, external consistency, and stability. For the results of inner consistency, most closure error series are within 0.2 ns, and the closure error series of each satellite fluctuate near zero and have no obvious systematic deviations. For the results of external consistency, the mean deviations of estimated DCBs of each satellite are mainly within 0.3 ns and 0.2 ns for the common types of DCB products of CAS and DLR, respectively. For the results of stability, the mean values of monthly STDs for the estimated DCBs are all smaller than 0.12 ns, which exhibits good stability. The STDs of the directly estimated DCBs are generally smaller than that of the DCB combinations of DLR and CAS. In this sense, the directly estimated DCBs for BDS-3 exhibits good performance in terms of accuracy and stability in this study, which can further provide the DCB corrections for precise positioning and navigation applications.

Keywords: GNSS; BDS-3; multi-frequency; differential code bias; MGEX

1. Introduction

Differential code bias (DCB) physically is defined as the difference in time delays between two different types of code observations. It includes intra-frequency DCB (e.g., C1W-C1C DCB) and inter-frequency DCB (e.g., C1W-C2W DCB) at the satellite or receiver side [1–3]. The DCB not only significantly affects the ionosphere modeling [4,5], but also is the error source in precise positioning, navigation, and timing applications of Global Navigation Satellite System (GNSS) when using code observations [6–8]. Hence, it is of great importance to estimate the DCBs and analyze their property, which can further provide the DCB corrections for GNSS applications.

There exist two major DCB estimation methods depending on the way of ionosphere modeling. The first one is to estimate the DCBs with a global or local ionosphere model simultaneously [9,10]. Another one is to estimate the DCBs after eliminating the ionosphere delays with a prior ionosphere model (e.g., German Aerospace Center (DLR)) [11,12]. The prior ionosphere models generally include the global ionospheric map (GIM) and some broadcast ionospheric models [13]. In addition, the DCBs can also be estimated by modeling the vertical total electron content (VTEC) above each station in the GNSS networks (e.g., Chinese Academy of Science (CAS)) [3,5]. Apart from the estimated DCBs by GNSS observations from ground stations, the estimation of DCBs can rely on the onboard observations from low-earth orbit (LEO) satellites [14–16]. The LEO solutions can achieve comparable and even better stability compared to the ground-based solutions from the global ground GNSS networks [17,18]. The

Editor in charge: Dr. Huizhong Zhu

aforementioned methods and observation sources for DCB estimation are used to obtain the DCB products of Global Positioning System (GPS) and BeiDou Satellite Navigation System (BDS) in recent years [19–21].

Currently, the construction of BDS-3 was fully completed with providing global services to GNSS users. However, there exist various DCBs of new frequencies and types which should be handled for BDS-3 [22,23]. At this time, it leads to a new situation with the following problems and limitations. First, some analysis institutions (e.g., DLR and CAS) only provide specific and limited types of DCB for BDS-3, which is not friendly to GNSS users in the situation of many new frequencies and code types. Second, the DCB corrections are usually added to the code observations by linear combinations of different types of DCB products currently. This will degrade the accuracy of DCB corrections, because the value derived from the linear combination cannot fully represent the true value due to various reasons, which may affect the GNSS applications.

To cope with these problems and limitations, there exists an urgent need to estimate the DCBs of new frequencies and code types for BDS-3. Therefore, in this study, the estimation method of DCB for BDS-3 is introduced first. Then, the BDS-3 observations from 40 globally distributed multi-GNSS experiment (MGEX) stations are selected to estimate the DCBs, including 19 types of DCBs of new frequencies for BDS-3. Moreover, the estimated DCBs are carefully analyzed in terms of inner consistency, external consistency, and stability compared to the DCB products provided by CAS and DLR.

2. Estimation method of BDS-3 DCBs

The dual-frequency code observations of arbitrary frequencies can be expressed as follows:

$$\begin{cases} P_{r,i}^s = \rho_r^s + c \cdot (dt_r - dt^s) + T_r^s + I_{r,i}^s + b_{r,i} + b_i^s + \varepsilon_{r,i}^s \\ P_{r,j}^s = \rho_r^s + c \cdot (dt_r - dt^s) + T_r^s + I_{r,j}^s + b_{r,j} + b_j^s + \varepsilon_{r,j}^s \end{cases} \quad (1)$$

where $P_{r,i}^s$ and $P_{r,j}^s$ denote the code observations on frequency i and j, respectively, the superscript s and subscript r denote the satellite and receiver, respectively, ρ_r^s denotes the geometric distance between the satellite and receiver, c is the light speed in vacuum, dt_r and dt^s denote the clock errors of the receiver and satellite, respectively, T_r^s denotes the tropospheric delays, $I_{r,i}^s$ and $I_{r,j}^s$ denote the ionospheric delays on frequency i and j, respectively, $b_{r,i}$ and $b_{r,j}$ denote the receiver time delays on frequency i and j, respectively, b_i^s and b_j^s denote the satellite time delays on frequency i and j, respectively, and $\varepsilon_{r,i}^s$ and $\varepsilon_{r,j}^s$ denote the code noises on frequency i and j, respectively. Based on (1), the code geometry-free (GF) combination on frequency i and j is deduced as follows:

$$P_{r,ij}^s = P_{r,i}^s - P_{r,j}^s = \left(I_{r,i}^s - I_{r,j}^s\right) + \left(b_{r,i} - b_{r,j}\right) + \left(b_i^s - b_j^s\right) + \Delta\varepsilon_{r,ij}^s \quad (2)$$

where $P_{r,ij}^s$ denotes the code GF combination. As the ionospheric delays are frequency dependent, the combined delay can be eliminated when the selected frequencies of GF combination are same. Otherwise, the ionospheric delays can be corrected by the high-precision GIM products and the corresponding mapping function.

The effect of observation noises can be degraded by averaging the results of GF combinations. At this time, the DCB combination of satellite and receiver can be expressed as follows:

$$DCB_{r,ij} + DCB_{ij}^s = \frac{1}{N}\sum_{t=1}^{N}\left(P_{r,i}^s(t) - P_{r,j}^s(t) - 40.28 \cdot \left(\frac{1}{f_i^2} - \frac{1}{f_j^2}\right) \cdot STEC_r^s\right) \quad (3)$$

where $DCB_{r,ij} = b_{r,i} - b_{r,j}$ is the receiver's DCB on frequency i and j, $DCB_{ij}^s = b_i^s - b_j^s$ is the satellite's DCB on frequency i and j, N is the number of epochs over a continuous period, t is the current observation time, f is the frequency of code observations, and $STEC_r^s$ is the slant total electron content (STEC) in TECU. Therefore, the DCB estimation model can be deduced as follows:

$$\boldsymbol{L}_{\text{sum}} = \boldsymbol{F} \cdot \widehat{\boldsymbol{X}}_{\text{DCB}} \quad (4)$$

wherein $\boldsymbol{L}_{\text{sum}}$ is the vector of DCB combination of satellite and receiver, $\boldsymbol{F}$ is the design matrix, and $\widehat{\boldsymbol{X}}_{\text{DCB}}$ is the estimated DCB vector. Specifically, $\boldsymbol{F}$ and $\widehat{\boldsymbol{X}}_{\text{DCB}}$ are given as follows:

$$\begin{cases} \underset{(u1+u2)\times 1}{\widehat{\boldsymbol{X}}_{\text{DCB}}} = \left[\underset{u1\times 1}{\widehat{\boldsymbol{X}}_{s,\text{DCB}}}, \underset{u2\times 1}{\widehat{\boldsymbol{X}}_{r,\text{DCB}}}\right]^{\text{T}} \\ \underset{n\times(u1+u2)}{\boldsymbol{F}} = \left[\underset{n\times u1}{\boldsymbol{A}}, \underset{n\times u2}{\boldsymbol{B}}\right] \end{cases} \quad (5)$$

where $\widehat{\boldsymbol{X}}_{s,\text{DCB}}$ and $\widehat{\boldsymbol{X}}_{r,\text{DCB}}$ are the DCB vector of satellite and receiver, respectively, $\boldsymbol{A}$ and $\boldsymbol{B}$ are the design matrix for satellite and receiver, respectively, and $u1$ and $u2$ are the numbers of satellites and receivers, respectively. The zero-mean constraint for the DCB of all satellites is introduced to separate the DCBs of satellites and receivers, which is as follows:

$$\begin{cases} \boldsymbol{S} \cdot \widehat{\boldsymbol{X}}_{\text{DCB}} = \boldsymbol{0} \\ \underset{1\times(u1+u2)}{\boldsymbol{S}} = \left[\underset{1\times u1}{\boldsymbol{e}}, \underset{1\times u2}{\boldsymbol{0}}\right], \underset{1\times u1}{\boldsymbol{e}} = [1, \cdots, 1] \end{cases} \quad (6)$$

where $\boldsymbol{S}$ is the design matrix of the zero-mean constraint. Then, the DCBs of the satellites and receivers can be estimated simultaneously by the least-square adjustment, which is as follows:

$$\begin{cases} \hat{\boldsymbol{X}}_{\mathrm{DCB}} = (\boldsymbol{N} + \boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{S})^{-1}\boldsymbol{W} \\ \boldsymbol{D}_{\hat{X}\hat{X}} = \sigma_0^2 \cdot (\boldsymbol{N} + \boldsymbol{S}^{\mathrm{T}}\boldsymbol{S})^{-1} \\ \boldsymbol{N} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{F}, \boldsymbol{W} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{L}_{\mathrm{sum}} \end{cases} \quad (7)$$

where $\boldsymbol{D}_{\hat{X}\hat{X}}$ is the variance matrix of the estimated DCBs, σ_0^2 is the variance of unit weight, and $\boldsymbol{P}$ is the weight matrix.

3. Experiments and analysis

In this section, the data description and processing strategy are given first. Then, the internal consistency, the external consistency, and the stability of the estimated BDS-3 DCBs are carefully analyzed compared to the products of DLR and CAS.

3.1 Data description and processing strategy

The BDS-3 observations from 40 globally distributed MGEX ground stations are selected to estimate the DCBs in this study. The processing period is continuous 30 days from day of the year (DOY) 60 to 89 in 2021. The distribution of the 40 GNSS stations is displayed in Figure 1. In addition, the types of DCBs of new frequencies for BDS-3 are shown in Table 1, as well as the used stations for estimating each type of DCB. It can be found that there exist 19 types of DCBs that need to be estimated, and the numbers of the used stations for each type of DCB are all larger than 14. There exists a realignment problem before comparing the estimated DCBs in this study, and the specific procedure of solving it can refer to [24]. For the processing strategy, the ionospheric delays of the code GF combinations are corrected by the GIM products provided by CODE with an interval of one hour. The cut-off satellite elevation is set to 20°. The minimum continuous observation arc is set as one hour. The least-square estimator is selected, and the weight matrix is determined by the STD of the GF combination series. The flowchart of the DCB estimation is also shown in Figure 2.

Table 1 Type of the DCB of new frequencies for BDS-3 and the corresponding used stations

Type	Number of the used stations
C1P-C2I	22
C1P-C5P	22
C1P-C6I	22
C1X-C5X	17
C1X-C6I	17
C1X-C7Z	14
C1X-C8X	14
C2I-C1X	17
C2I-C5X	18
C2I-C7Z	15
C2I-C8X	15
C5P-C2I	22
C5P-C6I	22
C5X-C6I	15
C5X-C7Z	15
C5X-C8X	15
C6I-C7Z	15
C6I-C8X	15
C7Z-C8X	15

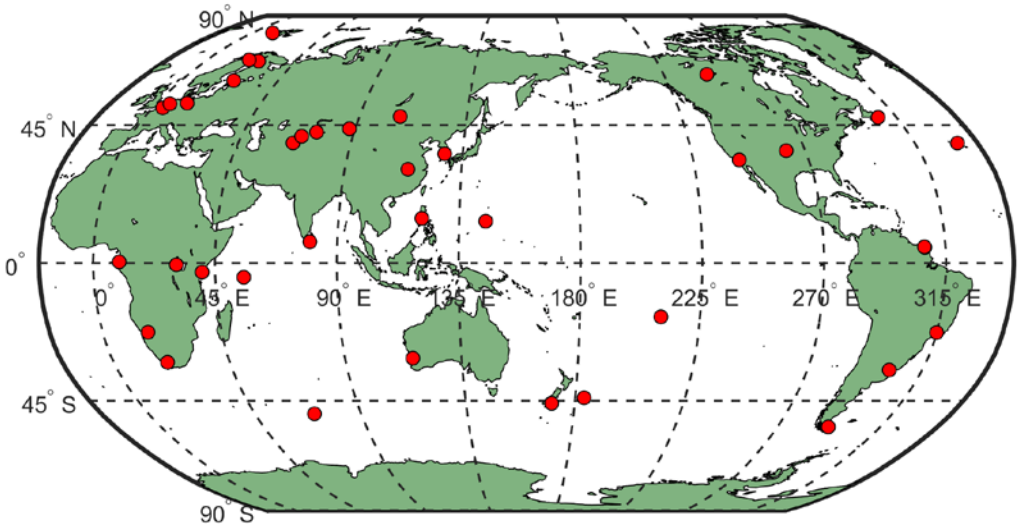


Figure 1 Distribution of the 40 globally distributed MGEX ground stations

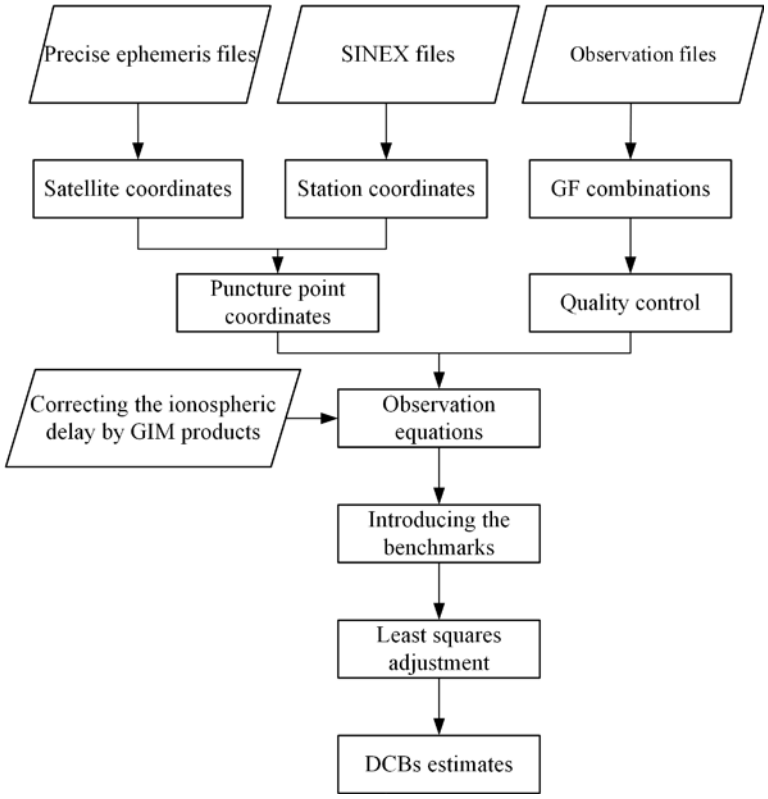


Figure 2 Flowchart of the DCB estimation for BDS-3, where the benchmarks denote the zero-mean constraint condition

3.2 Inner consistency of the estimated BDS-3 DCBs

Theoretically, the closure error for each satellite of these types of DCBs is equal to zero (e.g., $\mathrm{DCB}_{C2I-C5P} - \mathrm{DCB}_{C1P-C2I} - \mathrm{DCB}_{C1P-C5P} = 0$). However, the closure error usually exhibits non-zero in practical estimation due to the effects of observation noises and unmodeled errors. Hence, the property of the closure error can be used to evaluate the inner consistency of these estimated BDS-3 DCBs. For example, the daily closure error series of the estimated DCBs for each satellite are shown in Figures 3 to 6, including the types of C1P-C2I-C5P, C1P-C5P-C6I, C1X-C2I-C5X, and C1X-C5X-C6I. It can be found that most of these closure error series for each satellite are within 0.2 ns, and the closure error series of each satellite fluctuate near zero and have no obvious systematic deviations. In addition, the closure error series of C1P-C2I-C5P and C1P-C5P-C6I are within 0.1 ns, and smaller than that of other types. However, the closure error for some satellites exhibits large on some days (e.g., the closure error of C45 reaches 0.45 ns on DOY 67 and 88 for the type of C1X-C2I-C5X). The reason for this can be attributed to the insufficient number of processing epochs for the satellite on that day.

The mean values and STDs of closure error for the types of C1X-C2I-C5X and C1X-C5X-C6I are further shown in Figures 7 to 8, respectively. As shown in the figures, the mean values of closure error for each satellite are within 0.2 ns, and most of them are within 0.1 ns. In addition, the STDs of closure error for each satellite are within 0.3 ns. Hence, the inner consistency of estimated BDS-3 DCBs in this study exhibits good performance.

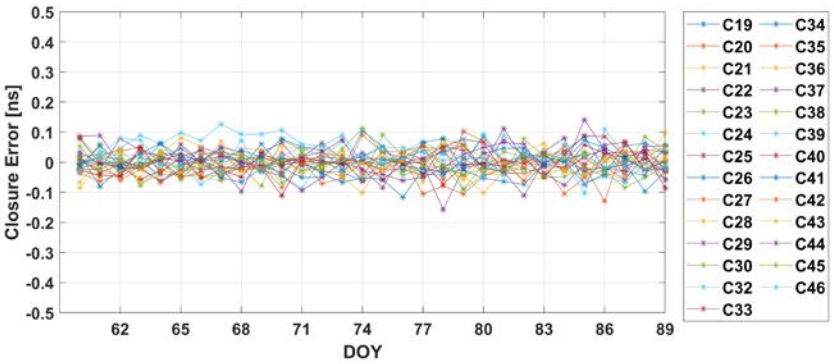


Figure 3 Daily closure error series for the type of C1P-C2I-C5P

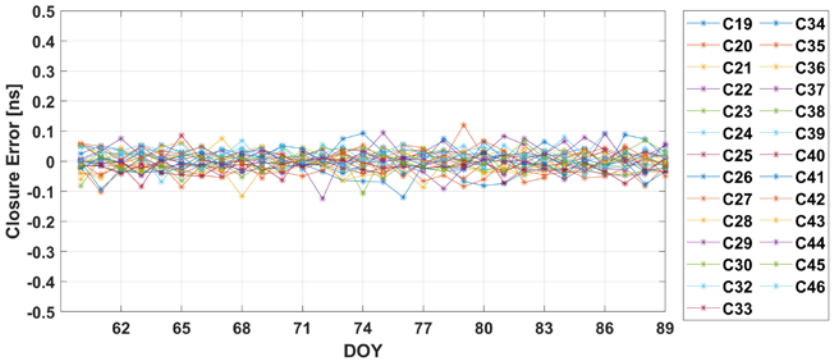


Figure 4 Daily closure error series for the type of C1P-C5P-C6I

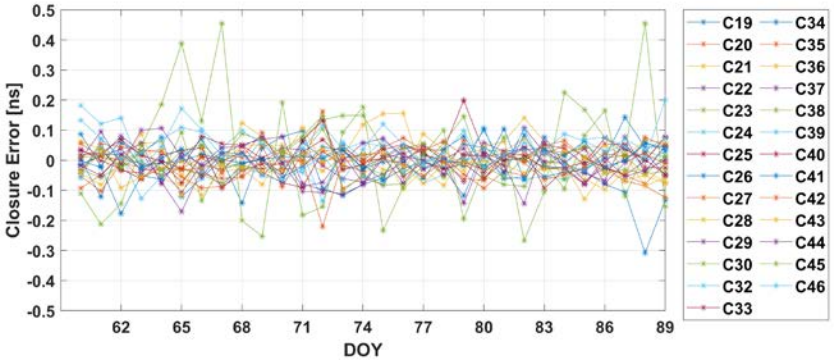


Figure 5 Daily closure error series for the type of C1X-C2I-C5X

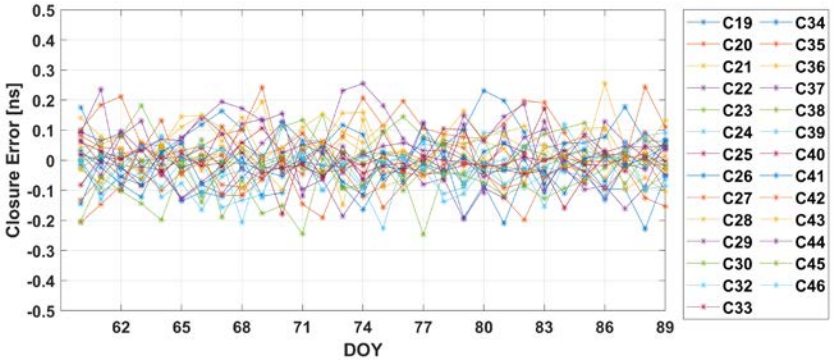


Figure 6 Daily closure error series for the type of C1X-C5X-C6I

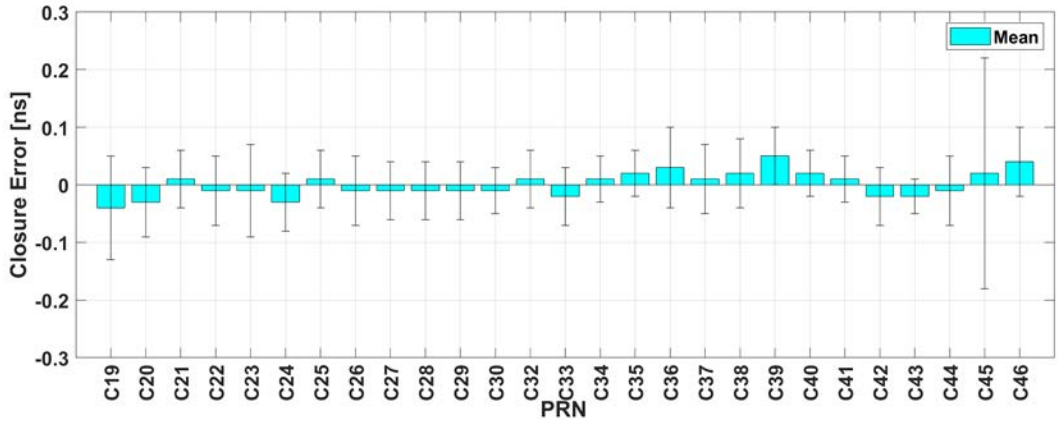


Figure 7 Mean values and STDs of closure error for the type of C1X-C2I-C5X

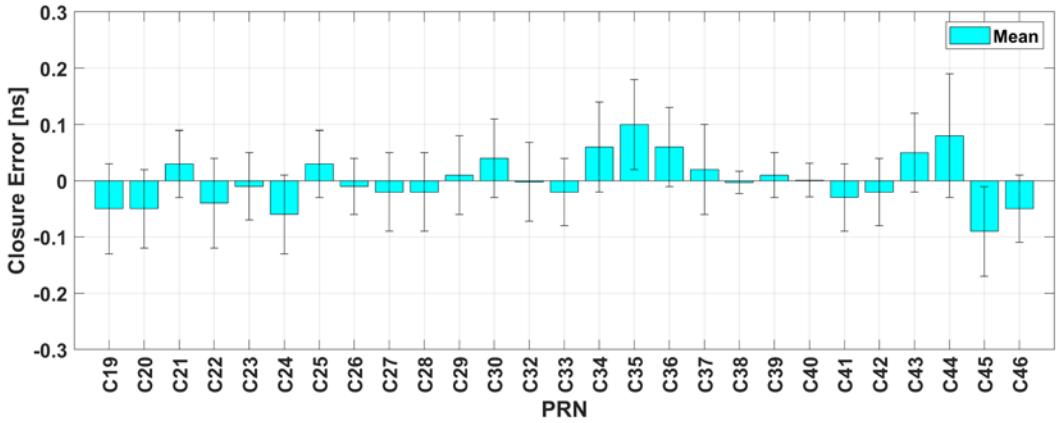


Figure 8 Mean values and STDs of closure error for the type of C1X-C5X-C6I

3.3 External consistency of the estimated BDS-3 DCBs

To further evaluate the external consistency of the estimated BDS-3 DCBs in this study (Hohai University, HHU), the DCB products of both DLR and CAS are used as references. Specifically, the DCB products of DLR and CAS are available at ftp.aiub.unibe.ch/CODE/ and ftp.gipp.org.cn/product/dcb/mgex/, respectively. For the common types of DCB products in HHU and the other two institutions, the deviation results can be obtained directly, which are defined as the directly estimated DCBs in this study. However, due to the limited DCB products provided by DLR and CAS, some estimated DCBs cannot be compared with the DCB products of DLR and CAS directly. Hence, the deviation results can be obtained by using the linear combinations of DCB products, which are defined as the DCB combinations in this study. Also, there may exist a datum transformation between the compared satellite sets [3]. The daily deviation series of the estimated DCBs for the types of C1P-C6I and C2I-C1X compared to CAS and DLR are depicted in Figures 9 to 10, respectively. It can be found that the accuracy of the estimated DCBs exhibits comparable performance compared to CAS and DLR, and the daily deviations for most satellites are within 0.2 ns. In addition, the monthly mean deviations of each satellite between HHU and the other two institutions are shown in Figures 11 to 14. It can be found that the mean deviations for each satellite between HHU and CAS are mainly within 0.3 ns, and the mean deviations for each satellite between HHU and DLR are mainly within 0.2 ns for the directly estimated DCB products. This verifies a good external consistency for the estimated BDS-3 DCBs in this study. Moreover, most mean deviations for each satellite between HHU and DCB combinations products of CAS and DLR are both within 0.6 ns. The accuracy of the directly estimated DCBs behave better than that of DCB combinations products, which further verifies the necessity to estimate the multi-frequency DCBs directly for BDS-3 satellites.

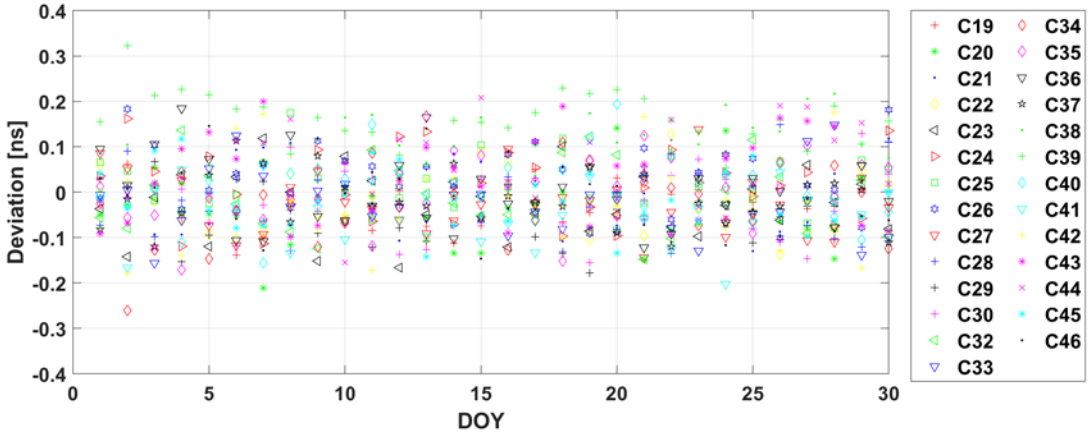


Figure 9 Daily deviation series of the estimated DCBs for the type of C1P-C6I

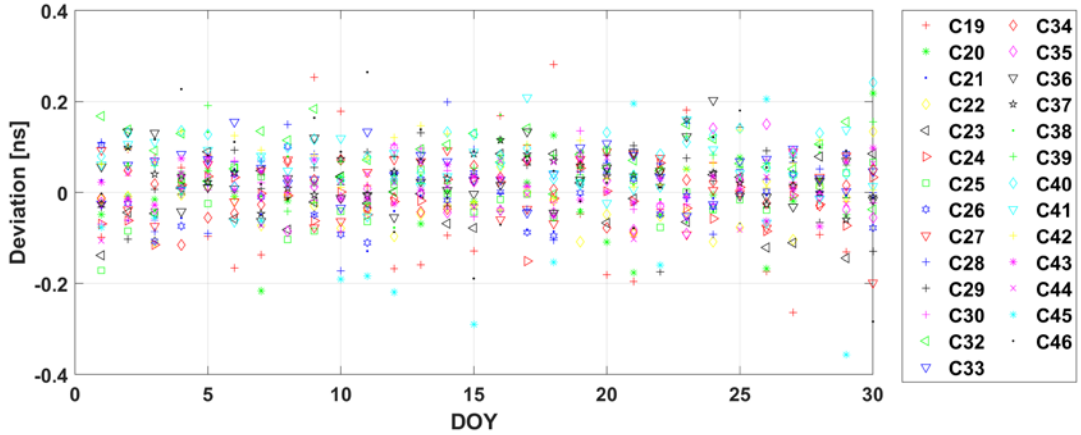


Figure 10 Daily deviation series of the estimated DCBs for the type of C2I-C1X

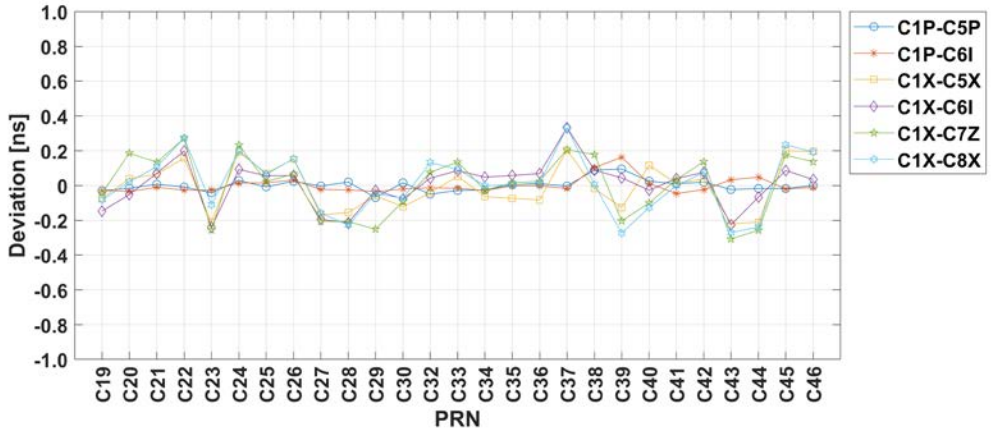


Figure 11 Mean deviations of each satellite for the directly estimated DCBs between HHU and CAS

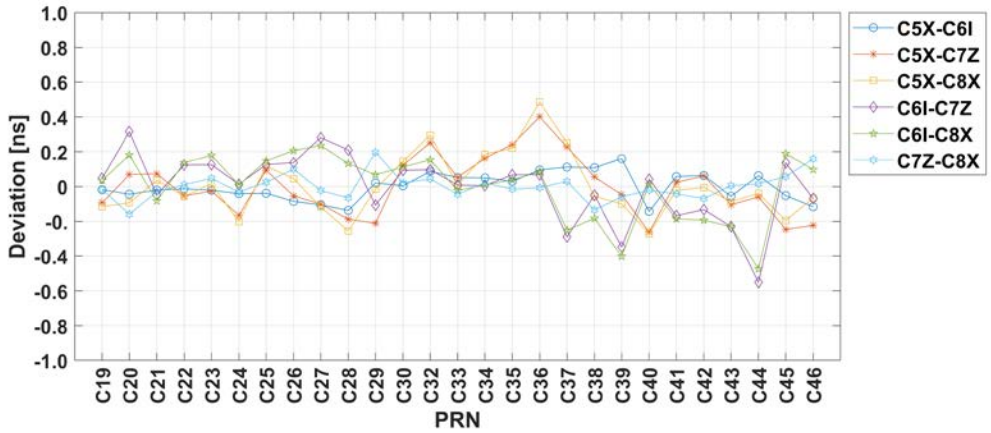


Figure 12 Mean deviations of each satellite for the DCB combinations between HHU and CAS

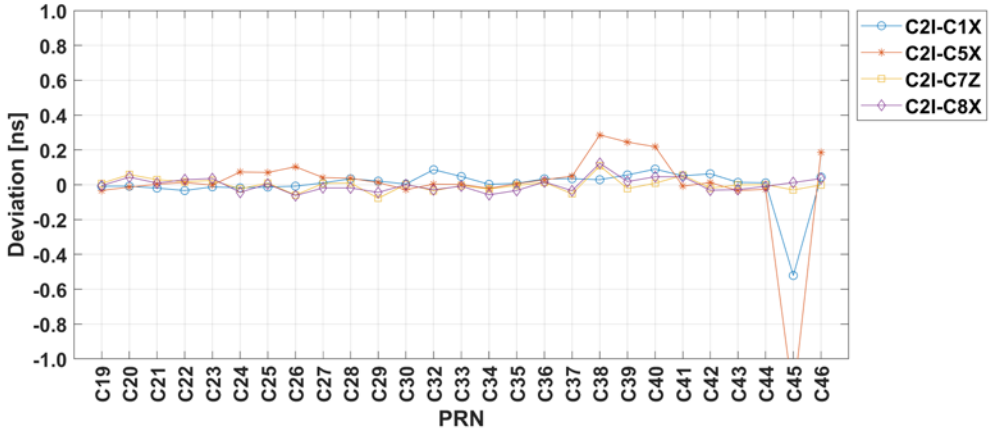


Figure 13 Mean deviations of each satellite for the directly estimated DCBs between HHU and DLR

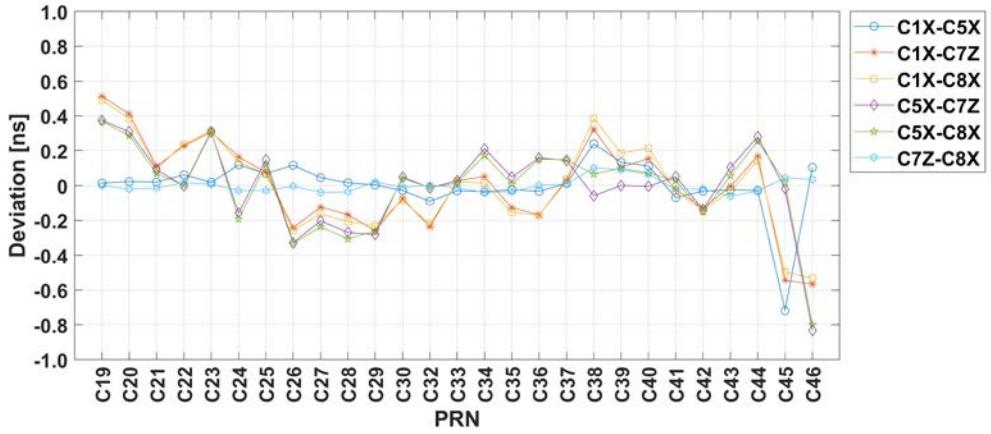


Figure 14 Mean deviations of each satellite for the DCB combinations between HHU and DLR

The monthly mean deviations of each satellite between HHU and the other two institutions are further listed in Tables 2 and 3 in detail. The HHU-CAS-1 denotes the mean deviations between the directly estimated DCBs of HHU and the DCB combinations products of CAS. The HHU-CAS-2 denotes the mean deviations between directly estimated DCBs of HHU and CAS. Similar definitions are used for HHU-DLR-1 and HHU-DLR-2, respectively. The bolded values are the mean deviations which are relatively large in the Tables. It can be found that the numbers of bolded values of HHU-CAS-1 and HHU-DLR-1 are significantly more than HHU-CAS-2 and HHU-DLR-2, respectively. Similar conclusions can be found in Figures 11 to 14, and this verifies the necessity to estimate the multi-frequency DCBs directly for BDS-3 satellites again.

Table 2 Monthly mean deviations of DCBs between HHU and CAS for each satellite (unit: ns)

PRN	HHU-CAS-1						HHU-CAS-2					
	C5XC6I	C5XC7Z	C5XC8X	C6IC7Z	C6IC8X	C7ZC8X	C1PC5P	C1PC6I	C1XC5X	C1XC6I	C1XC7Z	C1XC8X
C19	-0.02	-0.09	-0.12	0.05	0.03	-0.02	-0.03	-0.03	-0.07	-0.15	-0.04	-0.08
C20	-0.04	0.07	-0.09	**0.32**	0.18	-0.16	-0.02	-0.03	0.04	-0.05	0.18	0.02
C21	-0.02	0.07	0.04	-0.05	-0.08	-0.03	0.01	-0.01	0.06	0.07	0.13	0.11
C22	-0.01	-0.05	-0.06	0.12	0.14	0.01	0.01	-0.03	0.15	0.19	**0.27**	**0.27**
C23	-0.02	-0.03	0.02	0.12	0.18	0.05	-0.04	-0.03	-0.20	-0.20	**-0.25**	-0.11
C24	-0.04	-0.16	-0.20	0.01	0.01	-0.03	0.03	0.01	0.19	0.09	0.20	0.20
C25	-0.04	0.09	0.12	0.13	0.15	0.02	0.01	0.01	0.07	0.06	0.02	0.06
C26	-0.09	-0.05	0.04	0.14	**0.21**	0.10	0.02	0.03	0.15	0.06	0.06	0.15
C27	-0.11	-0.10	-0.12	**0.28**	**0.23**	-0.02	0.00	-0.02	-0.17	-0.19	-0.20	-0.16
C28	-0.14	-0.19	**-0.25**	**0.21**	0.13	-0.07	0.02	-0.03	-0.16	-0.20	-0.20	**-0.22**
C29	0.02	**-0.21**	-0.02	-0.11	0.07	0.20	-0.07	-0.04	-0.06	-0.03	**-0.25**	-0.05
C30	0.00	0.12	0.14	0.09	0.11	0.02	0.02	-0.02	-0.12	-0.08	-0.09	-0.07
C32	0.09	**0.25**	**0.29**	0.09	0.15	0.04	-0.05	-0.01	-0.04	0.04	0.07	0.13
C33	0.05	0.05	0.01	0.01	-0.02	-0.04	-0.03	-0.02	0.05	0.08	0.13	0.09
C34	0.05	0.16	0.19	0.01	0.01	0.02	-0.03	-0.03	-0.06	0.05	-0.02	-0.00
C35	0.03	**0.24**	**0.22**	0.07	0.04	-0.02	0.01	0.00	-0.07	0.06	0.01	0.01
C36	0.09	**0.40**	**0.49**	0.07	0.09	-0.01	0.01	0.00	-0.08	0.07	0.02	0.01
C37	0.11	**0.23**	**0.25**	**-0.29**	**-0.25**	0.03	0.00	-0.02	0.20	**0.30**	0.20	**0.34**
C38	0.11	0.06	-0.06	-0.05	-0.18	-0.13	0.09	0.10	-0.02	0.09	0.17	0.01
C39	0.16	-0.05	-0.10	**-0.35**	-0.40	-0.06	0.09	0.16	-0.13	0.04	-0.20	**-0.27**
C40	-0.14	**-0.26**	**-0.27**	0.04	0.01	-0.02	0.02	0.01	0.11	-0.03	-0.10	-0.12
C41	0.06	0.02	-0.02	-0.17	-0.19	-0.04	0.01	-0.05	0.01	0.04	0.03	-0.00
C42	0.06	0.06	-0.01	-0.13	-0.19	-0.07	0.02	-0.02	0.04	0.08	0.13	0.07
C43	-0.06	-0.10	-0.09	**-0.23**	**-0.23**	0.01	-0.02	0.03	-0.20	**-0.22**	**-0.31**	**-0.27**
C44	0.06	-0.06	-0.04	**-0.55**	**-0.47**	0.02	-0.02	0.05	-0.20	-0.07	**-0.25**	**-0.23**
C45	-0.05	**-0.25**	-0.19	0.14	0.19	0.06	-0.02	-0.02	0.20	0.09	0.17	**0.23**
C46	-0.12	**-0.22**	-0.06	-0.07	0.10	0.16	0.00	-0.01	0.19	0.03	0.13	0.19

3.4 Stability of the estimated BDS-3 DCBs

To further analyze the stability of the estimated DCBs for BDS-3 in this study, the daily series of the estimated DCBs for the types of C1P-C5P and C2I-C1X are depicted in Figures 15 to 16. It can be found that the daily DCB solutions for the type of C1P-C5P are between -40 ns~30 ns, and the daily DCB solutions for the type of C1P-C5P are between -6 ns~4 ns for all satellites. The daily DCB solutions for both two types exhibit good stability for each satellite. In addition, the monthly mean values of STDs for the estimated DCBs are listed in Table 4. The bolded values are the STDs of DCB combination products for DLR and CAS. It can be found that the monthly mean values of STDs for the estimated DCBs in this study are all smaller than 0.12 ns, which exhibits good stability. The differences of STDs between HHU and the other two institutions are within 0.02 ns, which further illustrates the consistency of stability between them. Besides, the STDs of the directly estimated DCBs for HHU are generally smaller than that of the DCB combination products for DLR and CAS. This also verifies the necessity to estimate the multi-frequency DCBs directly for BDS-3 satellites.

Table 3 Monthly mean deviations of DCBs between HHU and DLR for each satellite (unit: ns)

PRN	HHU-DLR-1						HHU-DLR-2			
	C1XC5X	C1XC7Z	C1XC8X	C5XC7Z	C5XC8X	C7ZC8X	C2IC1X	C2IC5X	C2IC7Z	C2IC8X
C19	0.01	**0.51**	**0.49**	**0.37**	**0.37**	0.00	-0.01	-0.03	0.01	0.00
C20	0.02	**0.41**	**0.39**	**0.31**	**0.29**	-0.02	-0.01	-0.01	0.06	0.04
C21	0.02	0.11	0.10	0.09	0.07	-0.02	-0.02	0.00	0.03	0.01
C22	0.06	**0.23**	**0.24**	0.00	0.00	0.02	-0.03	0.01	0.02	0.03
C23	0.02	**0.31**	**0.31**	**0.31**	**0.31**	0.01	-0.01	0.00	0.03	0.04
C24	0.12	0.16	0.13	-0.16	-0.19	-0.03	-0.02	0.07	-0.02	-0.04
C25	0.07	0.08	0.07	0.15	0.12	-0.03	-0.01	0.07	0.01	0.00
C26	0.12	**-0.24**	**-0.26**	**-0.33**	**-0.33**	0.00	-0.01	0.10	-0.06	-0.06
C27	0.04	-0.12	-0.16	-0.20	**-0.24**	-0.04	0.01	0.04	0.01	-0.02
C28	0.02	-0.17	**-0.21**	**-0.27**	**-0.31**	-0.04	0.03	0.04	0.01	-0.02
C29	0.00	**-0.26**	**-0.23**	**-0.28**	**-0.26**	0.02	0.02	0.01	-0.08	-0.04
C30	-0.03	-0.07	-0.08	0.05	0.04	-0.01	0.00	-0.03	0.00	0.00
C32	-0.09	**-0.24**	**-0.22**	-0.01	-0.01	0.00	0.09	0.00	-0.03	-0.03
C33	-0.03	0.03	0.02	0.03	0.01	-0.01	0.05	0.00	-0.01	-0.01
C34	-0.03	0.05	0.01	**0.21**	0.17	-0.04	0.00	-0.02	-0.02	-0.06
C35	-0.02	-0.13	-0.16	0.05	0.01	-0.03	0.01	0.01	0.00	-0.03
C36	-0.03	-0.17	-0.17	0.16	0.15	0.00	0.03	0.03	0.01	0.02
C37	0.01	0.03	0.04	0.14	0.15	0.01	0.03	0.05	-0.05	-0.03
C38	**0.24**	**0.32**	**0.39**	-0.06	0.07	0.10	0.03	**0.30**	0.11	0.12
C39	0.13	0.10	0.19	0.00	0.10	0.09	0.06	**0.24**	-0.02	0.02
C40	0.11	0.15	**0.21**	0.00	0.07	0.06	0.09	0.20	0.01	0.05
C41	-0.07	-0.02	-0.03	0.05	0.03	-0.02	0.05	-0.01	0.05	0.05
C42	-0.03	-0.13	-0.14	-0.14	-0.15	-0.02	0.06	0.01	-0.02	-0.03
C43	-0.02	0.00	-0.03	0.10	0.06	-0.06	0.01	-0.03	0.00	-0.03
C44	-0.03	0.17	0.14	**0.28**	**0.25**	-0.03	0.01	-0.03	0.00	-0.01
C45	**-0.72**	**-0.54**	**-0.50**	-0.02	0.02	0.04	**-0.50**	**-1.22**	-0.03	0.01
C46	0.10	**-0.57**	**-0.53**	**-0.83**	**-0. 80**	0.03	0.04	0.18	0.00	0.04

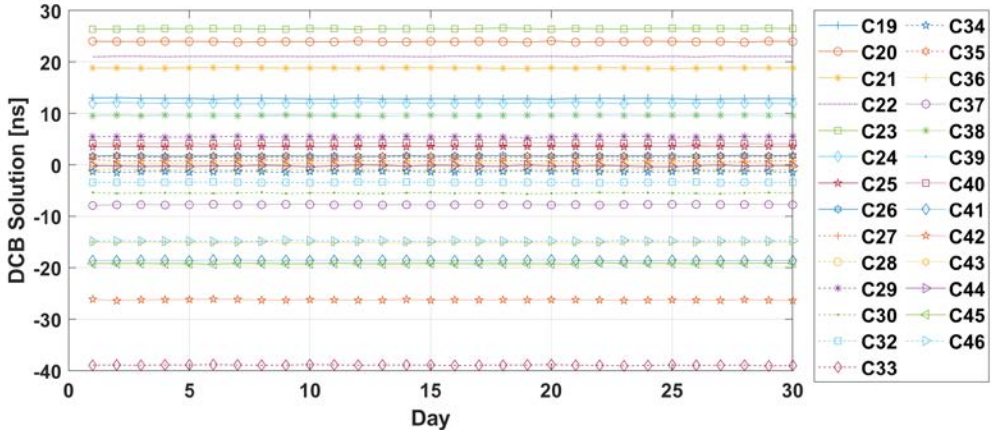


Figure 15 Daily solution series of the estimated DCB for the type of C1P-C5P

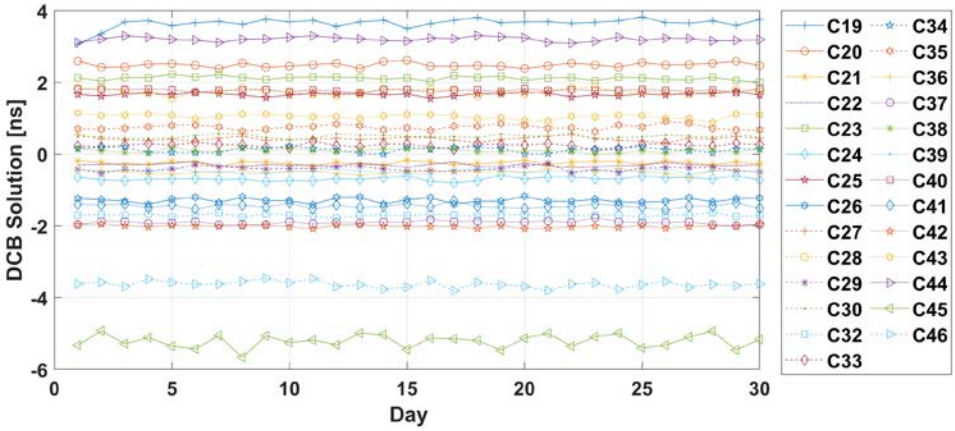


Figure 16 Daily solution series of the estimated DCB for the type of C2I-C1X

Table 4 Monthly mean values of STDs for the 19 types of the estimated DCBs (unit: ns)

Type	HHU	CAS	DLR
C1P-C2I	0.05	**0.06**	\
C1P-C5P	0.09	0.08	\
C1P-C6I	0.06	0.06	\
C1X-C5X	0.09	0.09	\
C1X-C6I	0.10	0.10	\
C1X-C7Z	0.11	0.09	\
C1X-C8X	0.11	0.10	\
C2I-C1X	0.06	**0.09**	0.07
C2I-C5X	0.11	**0.11**	0.12
C2I-C7Z	0.12	**0.10**	0.10
C2I-C8X	0.12	**0.12**	0.10
C5P-C2I	0.09	**0.08**	\
C5P-C6I	0.04	**0.05**	\
C5X-C6I	0.09	**0.10**	**0.11**
C5X-C7Z	0.05	**0.08**	**0.11**
C5X-C8X	0.03	**0.07**	**0.10**
C6I-C7Z	0.12	**0.14**	**0.11**
C6I-C8X	0.11	**0.11**	**0.11**
C7Z-C8X	0.03	**0.06**	**0.06**

4. Conclusions

In this paper, we focus on the estimation and analysis of BDS-3 multi-frequency differential code bias, especially for the new frequencies. The estimation method of BDS-3 multi-frequency DCBs is introduced first, and then the estimated DCBs in this study are carefully analyzed compared to the DCB products of both DLR and CAS in terms of inner consistency, external consistency, and stability.

For the results of inner consistency, most of these closure error series are within 0.2 ns, and the closure error series of each satellite fluctuate near zero and have no obvious systematic deviation. The monthly mean values of the closure errors for each satellite are within 0.2 ns, and most of them are within 0.1 ns. Hence, the inner consistency of the estimated BDS-3 DCBs in this study exhibits good performance. For the results of external consistency, the mean deviations for each satellite between HHU and CAS are mainly within 0.3 ns, and the mean deviations for each satellite between HHU and DLR are mainly within 0.2 ns for the directly estimated DCB products. This verifies a good external consistency for the estimated BDS-3 DCBs in this study. For the results of stability, the monthly mean values of STDs for the estimated BDS-3 DCBs are all smaller than 0.12 ns, which exhibits good stability. The differences of STDs between HHU and the other two institutions are within 0.02 ns, which illustrates their consistency of stability. The STDs of the directly estimated BDS-3 DCBs in this study are generally smaller than that of the DCB combination products of DLR and CAS. Therefore, the accuracy and stability of the directly estimated BDS-3 DCBs exhibit better performance than that of DCB combination products, which further verifies the necessity to estimate the multi-frequency DCBs directly for BDS-3 satellites.

Acknowledgment

This study was supported by The National Natural Science Foundation of China (41830110, 42004014), Natural Science Foundation of Jiangsu Province (BK20200530), Open Project Research Fund of Technology Innovation Center for Geological Environment Monitoring MNR (2022KFK1212002), China Postdoctoral Science Foundation

(2020M671324), Jiangsu Planned Projects for Postdoctoral Research Funds (2020Z412).

References

[1] M Hernández-Pajares, J.M. Juan, J. Sanz, R. Orus, A. Garcia-Rigo, J. Feltens, A. Komjathy, S.C. Schaer, A. Krankowski, The IGS VTEC maps: a reliable source of ionospheric information since 1998, J. Geod. 83 (2009) 263–275. https://doi.org/10.1007/s00190-008-0266-1.

[2] O. Montenbruck, A. Hauschild, Code biases in multi-GNSS point positioning, ION-ITM-2013, 2013.

[3] N. Wang, Y. Yuan, Z. Li, O. Montenbruck, B. Tan, Determination of differential code biases with multi-GNSS observations, J. Geod. 90 (2016) 209–228, doi.org/10.1007/s00190-015-0867-4.

[4] E., Sardón, A., Rius, N., Zarraoa, Estimation of the transmitter and receiver differential biases and the ionospheric total electron content from Global Positioning System observations, Radio Sci. (1994), doi.org/10.1029/94RS00449.

[5] Z. Li, Y. Yuan, L. Hui, J. Ou, X. Huo, Two-step method for the determination of the differential code biases of COMPASS satellites, J. Geod. 86 (2012) 1059–1076, doi.org/10.1007/s00190-012-0565-4.

[6] F. Guo, X. Zhang, J. Wang, Timing group delay and differential code bias corrections for BeiDou positioning, J. Geod. 89 (2015) 427–445, doi.org/10.1007/s00190-015-0788-2.

[7] B. Zhang, P. Teunissen, Y. Yuan, H. Zhang, M. Li, Joint estimation of vertical total electron content (VTEC) and satellite differential code biases (SDCBs) using low-cost receivers, J. Geod. 92 (2018) 401–413, doi.org/10.1007/s00190-017-1071-5.

[8] T. Liu, B. Zhang, Y. Yuan, M. Li, Real-Time Precise Point Positioning (RTPPP) with raw observations and its application in real-time regional ionospheric VTEC modeling, J. Geod. 92 (2018) 1–17, doi.org/10.1007/s00190-018-1118-2.

[9] A.J. Mannucci, BD Wilson, D.N. Yuan, C.H. Ho, U.J. Lindqwister, T.F. Runge, A global mapping technique for GPS-derived ionospheric total electron content measurements, Radio Sci. 33 (1998) 565–582, doi.org/10.1029/97RS02707.

[10] A. Krankowski, I.I. Shagimuratov, I.I. Ephishov, A. Krypiak-Gregorczyk, G. Yakimova, The occurrence of the mid-latitude ionospheric trough in GPS-TEC measurements, Adv. Space Res. 43 (2009) 1721–1731, doi.org/10.1016/j.asr.2008.05.014.

[11] O. Montenbruck, A. Hauschild, P. Steigenberger, Differential Code Bias Estimation using Multi-GNSS Observations and Global Ionosphere Maps, NAVIGATION. 61 (2014) 191–201. doi.org/10.1002/navi.64.

[12] W. Li, M. Li, C. Shi, R. Fang, W. Bai, GPS and BeiDou Differential Code Bias Estimation Using Fengyun-3C Satellite Onboard GNSS Observations, Remote Sens. 9 (2017) 1239, doi.org/10.3390/rs9121239.

[13] Y. Zhu, S. Tan, L. Feng, X. Cui, X. Jia, Estimation of the DCB for the BDS-3 New Signals Based on BDGIM Constraints, Adv. Space Res. (2020). doi.org/10.1016/j.asr.2020.05.019.

[14] J. Lin, X. Yue, S. Zhao, Estimation and analysis of GPS satellite DCB based on LEO observations, GPS Solut. (2016), doi.org/10.1007/s10291-014-0433-1.

[15] L. Yuan, S. Jin, M.M. Hoque, Estimation of LEO-GPS receiver differential code bias based on inequality constrained least square and multi-layer mapping function, GPS Solut. 24 (2020). https://doi.org/10.1007/s10291-020-0970-8.

[16] L. Yuan, M. Hoque, S. Jin, A new method to estimate GPS satellite and receiver differential code biases using a network of LEO satellites, GPS Solut. 25 (2021) 1–12. doi.org/10.1007/s10291-021-01109-y.

[17] X. Li, T. Ma, W. Xie, K. Zhang, J. Huang, X. Ren, FY-3D and FY-3C onboard observations for differential code biases estimation, GPS Solut. 23 (2019), doi.org/10.1007/s10291-019-0850-2.

[18] X. Li, W. Zhang, K. Zhang, Q. Zhang, Y. Yuan, GPS satellite differential code bias estimation with current eleven low earth orbit satellites, J. Geod. 95 (2021), doi.org/10.1007/s00190-021-01536-2.

[19] R. Jin, S. Jin, G. Feng, M_DCB: Matlab code for estimating GNSS satellite and receiver differential code biases, Gps Solut. 16 (2012) 541–548. doi.org/10.1007/s10291-012-0279-3.

[20] B. Shu, H. Liu, L. Xu, X. Gong, R. Zhang, Analysis of satellite-induced factors affecting the accuracy of the BDS satellite differential code bias, GPS Solut. (2016) 1–12. doi.org/10.1007/s10291-016-0577-2.

[21] X. Li, W. Xie, J. Huang, T. Ma, X. Zhang, Y. Yuan, Estimation and analysis of differential code biases for BDS3/BDS2 using iGMAS and MGEX observations, J. Geod. 93 (2019) 419–435. doi.org/10.1007/s00190-018-1170-y.

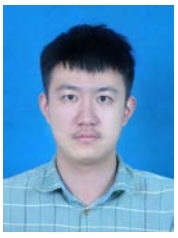
[22] Y. Yang, Y. Mao, B. Sun, Basic performance and future developments of BeiDou global navigation satellite system, Satell. Navig. 1 (2020) 1, doi.org/10.1186/s43020-019-0006-0.

[23] Y. Yang, X.U. Yangyin, L.I. Jinlong, C. Yang, Progress and performance evaluation of BeiDou global navigation satellite system: Data analysis based on BDS-3 demonstration system, Sci. China Earth Sci. 61 (2018) 614–624. doi.org/10.1007/s11430-017-9186-9.

[24] J. Sanz, J. Miguel Juan, A. Rovira-Garcia, G. González-Casado, GPS differential code biases determination: methodology and analysis, GPS Solut. 21 (2017) 1549–1561, doi.org/10.1007/s10291-017-0634-5.

Authors

Haijun Yuan received the M.S. degree from Hohai University, Nanjing, China in 2021. He is currently a Ph.D candidate at the School of Earth Sciences and Engineering, Hohai University. His current research mainly focuses on multi-frequency and multi-constellation GNSS precise positioning and navigation.



Zhuoming Hu received the M.S. degree from Hohai University, Nanjing, China in 2022. His research interests include the GNSS data processing, real-time precise positioning and navigation.

Xiufeng He received the Ph.D. degree from the Hong Kong Polytechnic University in 1998. She is currently a professor at the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. Her research interests include satellite geodesy, deformation monitoring, multi-source data fusion, and integrated navigation.

Zhetao Zhang received the Ph.D. degree from Tongji University, Shanghai, China in 2019. He is currently an associate professor at the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. His current research focuses on the GNSS precise positioning and navigation under the conditions including canyon environment, low-cost receiver, and multi-GNSS situation.

Journal of Global Positioning Systems (*2022*)

Vol. 18, No. 1: 28-40

DOI: 10.5081/jgps.18.1.28

A Geometry-based Ambiguity Validation (GBAV) Method for GNSS carrier phase observation

Wu Chen[1), 4)], Ying Xu[2)], Duojie Weng[1),4)], Shengyue Ji[3)]

[1)] Department of Land Surveying and Geo-informatics, Hong Kong Polytechnic University

[2)] College of Geomatics, Shandong University of Science and Technology, Qingdao, China

[3)] School of Geosciences, China University of Petroleum, Qingdao, China

[4)] Research Institute for Artificial Intelligence of Things, Hong Kong Polytechnic University

✉ **Corresponding author**: Wu Chen, lswuchen@polyu.edu.hk

Abstract: Integer ambiguity validation is an indispensable and critical step in GNSS carrier phase positioning for precise and reliable positioning applications. The crucial problems associated with any ambiguity validation methods are as follows. 1) The fixed ambiguity vector can be separated from all other ambiguity candidates under certain tests (separability). 2) The probability of fixing to wrong ambiguity combinations (mis-fixing) can be controlled to an acceptable level based on different application requirements. Traditional ambiguity validation methods, such as the *R*-ratio and the difference tests which use one statistical test to control both separability and mis-fixing rate, are widely used due to easier computation. The performances of these methods are generally acceptable. However, experiments show that these tests with a fixed threshold can cause either a small percentage of mis-fixing or overly conservative with long observation time. In this paper, we propose a new Geometry Based Ambiguity Validation (GBAV) method which uses two statistical tests to control geometry separability and mis-fixing probability separately. The thresholds for both tests can be strictly determined based on user requirements to control the quality of ambiguity resolution. Three 24-hour GNSS (GPS, BDS) datasets (two short baselines and one middle-range baseline) are processed using the proposed GBAV method, and compared with the popular *R*-ratio method. The results show that by giving proper control on the mis-fixing probability (<0.01%), there is no mis-fixing case in all three datasets.

Keywords: GNSS, Ambiguity, Carrier Phase

1. Introduction

The Global Navigation Satellite Systems (GNSS) are satellite navigation systems which provide space-based positioning, navigation and timing (PNT) services in all weather conditions, anywhere on or near the Earth (Leick 2004). GNSS provides two common types of measurements for positioning, namely pseudo-range and carrier phase. These measurements enable the determination of the ranges between the receiver antenna and the satellites. The carrier phase based positioning results in more precise range than those from pseudo-range, if the carrier phase ambiguity can be reliably resolved (Han and Rizos 1999). However, an incorrect integer ambiguity solution may cause severe biases in the position solution and in any other of the real-valued parameters and it is important to assess the probability of correct ambiguity estimation (Verhagen and Teunissen 2013). Thus, integer ambiguity validation is an indispensable and critical step in GNSS ambiguity resolution process. Over the past years, various ambiguity validation methods have been proposed, such as *F*-ratio test (Frei and Beutler 1990; Euler and Landau 1992), *R*-ratio test (Euler and Schaffrin 1991; Leick 2004; Teunissen and Verhagen 2009), difference test (Tiberius and De Jonge 1995), projector test (Wang et al. 1998; Han 1997). These methods are easy to compute and the performances are generally acceptable, if correct critical values are selected. However, there are some disadvantages for this type of ambiguity validation methods. Taking the most popular ratio test as an example, the critical values are normally selected empirically, as the statistic distributions for the tests are difficult to be established.

Editor in charge: Dr. Dongsheng Zhao

Therefore it is difficult to evaluate or to compare the performances of these empirical tests (Li and Wang 2012). In addition, the experiments results from (Teunissen and Verhagen 2009; Teunissen and Verhagen 2004; Teunissen 2013;Verhagen and Teunissen 2013) indicated that the traditional use of the ratio test with a fixed threshold often results in either unacceptably high failure rates or overly conservative. For the strong models, the fixed value ratio tests are often too conservative, so that the false alarm rates are unnecessarily high, while the failure rates are very close to zero. For weak models, on the other hand, the currently used fixed values are often too low, so that the fixed solution is often wrongly accepted, resulting in high failure rates. To overcome these problems, Verhagen and Teunissen (2006) proposed a model-driven ratio test with a fixed failure rate. Simulation results have shown that it is possible to describe the threshold values based only on the number of ambiguities and the failure rate. Besides the ambiguity validation test mentioned above, Ellipsoidal Integer Aperture (EIA) (Teunissen 2003), and Penalized Integer Aperture (PIA) (Teunissen 2004) based validation methods are dependent on lower bound (Teunissen 1998a) or upper bound (Teunissen 2000) of ambiguity resolution success rates. The advantage of these approaches is that critical values of the statistical tests are linked with user controlled fail rates. However, the critical values rely on satellite geometry and it is difficult to describe them mathematically, particularly for multiple epoch observations. Also, the sample size is important for these approaches, resulting that longer observation time is preferred for reliable solution (Li and Wang 2012). To reduce the time required for observation and to improve the reliability, Ji et al. (2010) proposed to combine *R*-ratio and EIA tests together for ambiguity validation. Through allowing slight overlap of pull-in region, the observation time for EIA could be significantly reduced. The *R*-ratio test was applied at the same time to discriminate the cases in the overlapping regions. Test results showed the combined validation method improved the ambiguity resolution reliability, and had similar efficiency to the *R*-ratio at the same time. A comprehensive review and evaluation of these tests can be found in Verhagen (2004; 2005), Verhagen and Teunissen (2006), and Li and Wang (2012).

Geometrically, ambiguity resolution tries to find an intersection point of all ambiguities with minimum residuals, compared with all the other ambiguity combinations in ambiguity space. If there are no errors in GNSS measurements, the ambiguities can be fixed to integers when there is only one intersection point. For a given GNSS datasets, if there were two intersection points, no validation method can distinguish them. On the other hand, the measurement errors may shift the correct intersection point significantly. It will cause the ambiguity resolution algorithms fixing to wrong ambiguity. As mentioned above, ratio tests were applied popularly and empirically. They use one statistical test to control both problems. As a result, these tests with a fixed threshold may cause some ambiguity mis-fixing casesor overly conservative with long observation time. In this paper, we proposed a new Geometry Based Ambiguity Validation (GBAV) method to separate the validation test into statistical tests, including the spatial separability and mis-fixing rate. Based on the statistical distributions of the two tests, we are able to determine the thresholds based on user requirements to control the spatial separability and mis-fixing rate separately. With this new method, we can efficiently control the mis-fixing probability to ensure the quality of ambiguity resolution.

In section 2, the concept spatial separability and mis-fixing condition are introduced and their associated probabilities are given. The proposed Geometry Based Ambiguity Validation (GBAV) method is summarized in section 3. Numerical tests and results with three GNSS 24-hour datasets are given in section 4. The discussions and conclusions are given in section 5.

2. Spatial Separability and Mis-fixing Condition for Ambiguity Validation

The general form of linear observation equation on GNSS carrier phase observation can be expressed as (Parkinson et al. 1996; Leick 2004; Hofmann-Wellenhof et al. 1993):

$$AX + BN + \varepsilon = L \tag{1}$$

where L denotes the double difference observation vector, N is double difference carrier phase integer ambiguity vector ($N \in Z^n$), X is the vector of the other unknown parameters (including position coordinates), ε is the random errors, and the matrices A and B are the corresponding design matrices.

The solution of Eq. (1) can be obtained by minimizing Eq. (2) (Verhagen 2004):

$$\min ||L - BN - AX||^2_{Q_L}, \quad N \in Z^n, X \in R^n \tag{2}$$

where $||*||^2_{Q_L} = (*)^T {Q_L}^{-1}(*)$, and Q_L is the variance-covariance matrix of observation vector L.

In general, the ambiguity fix solution can be divided into three steps (Teunissen 1995). In the first step, the integer constraints on the ambiguities are simply ignored. The unconstrained least-squares solution is referred to as the float solution of $\hat{N}$, $\hat{X}$. The corresponding variance-covariance (VC) matrix is as following,

$$\begin{bmatrix}\hat{N}\\ \hat{X}\end{bmatrix},\ \begin{bmatrix}Q_{\hat{N}} & Q_{\hat{N}\hat{X}}\\ Q_{\hat{X}\hat{N}} & Q_{\hat{X}}\end{bmatrix} \quad (3)$$

In the second step, the integer ambiguity estimation $\breve{\mathrm{N}}$ is computed from the 'float' ambiguity, subject to $\min \left|\left|\hat{N}-\breve{\mathrm{N}}\right|\right|^2_{Q_{\hat{N}}}$. Integer rounding, integer bootstrapping and integer least-squares are different methods for obtaining the integer solution. Integer least-square (ILS) is optimal, as it maximizes the probability of correct integer estimation (Teunissen 1999). In contrast to rounding and bootstrapping, an integer search is needed to compute the ILS solution. This can be efficiently done with the LAMBDA method (Teunissen 1995b). Finally, fixed solution is obtained by:

$$\breve{X} = \hat{X} - Q_{\hat{X}\hat{N}} {Q_{\hat{N}}}^{-1}(\hat{N} - \breve{\mathrm{N}}) \quad (4)$$

For relative positioning, if the double difference ambiguity vector is truly known as N_0, the double difference carrier phase observation equation can be written as,

$AX_0 = L_0 + \lambda N_0 + e$ with a weight matrix P (5)

where A is the design matrix, X_0 is the receiver position vector, L_0 is the double difference carrier phase measurement vector without noise, λ is the wavelength of carrier phase, and e is the true error vector of carrier phase measurement. The other errors, such as tropospheric delay and ionospheric delay are not considered here as the double difference process significantly reduces their effects on short baselines. For Long baselines, we can use GNSS measurements to estimate tropospheric and ionospheric delays, together with receiver position and clock error parameters.

Assuming N_0 is known, the residual vector V_0 and the weighted sum of squared residuals Z_0 of Eq. (5) can be expressed as Eqs. (6) and (7), when the least squares estimation method is used to estimate position vector X_0.

$$V_0 = (I - H)e \quad (6)$$

$$Z_0 = V_0^T P V_0 = e^T (I - H) P (I - H) e \quad (7)$$

where

$$H = A(A^T P A)^{-1} A^T P \quad (8)$$

The ambiguity validation problem can be generally described as follows.

Give a group of ambiguity candidates $(N_1, N_2, N_3, \dots\dots N_m)$, and $N_0 \epsilon (N_1, N_2, N_3, \dots\dots N_m)$, $\forall N_i \epsilon (N_1, N_2, N_3, \dots\dots N_m)$, check if $N_0 = N_i$ for all $i = 1, \dots\dots, m$.

If N_i is a selected candidate,

$$AX_i = L_0 + \lambda N_i + e = L_0 + \lambda N_0 + \lambda \Delta N_i + e \quad (9)$$

where

$$\Delta N_i = N_i - N_0 \quad (10)$$

The total error in Eq. (9) is

$$\Delta = \lambda \Delta N_i + e \quad (11)$$

The residual vector of Eq. (9) is

$$V_i = (I - H)\Delta = \lambda (I - H)\Delta N_i + (I - H)e \quad (12)$$

The weighted sum of squared residuals Z_i is

$$Z_i = V_i^T P V_i = \lambda^2 \Delta N_i^T (I - H) P (I - H) \Delta N_i + e^T (I - H) P (I - H) e + 2\lambda \Delta N_i^T (I - H) P (I - H) \mathrm{e} \quad (13)$$

On the other hand, differencing Eq. (5) and Eq. (9), we can have:

$$A \Delta X_i = \lambda \Delta N_i \quad (14)$$

where $\Delta X_i = X_i - X_0$ is the position shift due to the wrong ambiguity.

There is no measurement error in Eq. (14), its residual only reflects the coordinate difference caused by the ΔN_i. The residual and the weighted sum of squared residuals of Eq. (14) are,

$$V_{\Delta Ni} = \lambda (I - H) \Delta N_i \quad (15)$$

and

$$V_{\Delta Ni}^T P V_{\Delta Ni} = \lambda^2 \Delta N_i^T (I - H) P (I - H) \Delta N_i \quad (16)$$

In another word, the residuals of Eq. (14) only reflect whether N_0 and N_i can be separated geometrically or not. For instance, N_0 and N_i cannot be separable if $V_{\Delta Ni} = 0$. Furthermore, the weighted sum of squared residuals (Eq. 16) can also be used to describe the degree whether N_0 and N_i can be separated. When $V_{\Delta Ni}^T P V_{\Delta Ni}$ is too small compared to $V_0^T P V_0$; N_0 and N_i cannot be separated geometrically. Conversely, it is possible to separate N_0 from N_i while $V_{\Delta Ni}^T \mathrm{P} \mathrm{V}_{\Delta Ni}$ is relatively larger than the sum of the squares of noise $V_0^T P V_0$.

Inserting Eqs. (6), (7), (15) and (16) into (13),

$$Z_i = V_i^T P V_i = V_{\Delta Ni}^T P V_{\Delta Ni} + V_0^T P V_0 + 2 V_{\Delta Ni}^T P V_0 \quad (17)$$

The weighted sum of squared residuals for Eq. (9) consists of three terms. The first term is only determined by the difference of selected ambiguity vector and the true ambiguity vector. The second term is only related to the true error e, and the third term is affected by the projection of the true error to the direction of residual vector $V_{\Delta \mathrm{N} i}$.

If the ambiguity candidates $(N_1, N_2, N_3, \dots\dots N_m)$ of Eq. (9) is arranged based on the size of weighted sum of squared residuals (Eq. (13)), from smallest to the largest, the ambiguity N_1 is an optimal solution of Eq. (9). Thus the ambiguity validation problem can be

described as whether N_1 is the true ambiguity vector N_0 or not, which can be divided into two cases: 1) N_0=N_1, and 2) N_0=N_i $(i \neq 1)$.

Case 1: N_0=N_1,

Let us consider case 1 first, when N_1 is the true ambiguity vector, or,

$$N_0 = N_1 \tag{18}$$

and thus,

$$V_1 = (I - H)e \tag{19}$$

Select N_i as a possible candidate ($i \neq 1$), the difference of the sums of residuals between N_1 and N_i should be,

$$\Delta Z_{1i} = Z_i - Z_1 = V_{\Delta N1i}^T P V_{\Delta N1i} + 2V_{\Delta N1i}^T P(I - H)e = V_{\Delta N1i}^T P V_{\Delta N1i} + 2V_{\Delta N1i}^T P V_1 > 0 \tag{20}$$

where $\Delta N_{1i} = N_i - N_1$

In Eq. (20), the first term is determined only by the geometry of satellites and ambiguity difference ΔN_{1i}. As mentioned above, it can be used to describe whether N_0 (N_1 in this case) and N_i can be separated compared to $V_0^T P V_0$ ($V_1^T P V_1$ in this case). Now we can define the separability index of N_i and N_1 as:

$$S_{1i} = \frac{V_{\Delta N1i}^T P V_{\Delta N1i}}{V_1^T P V_1} \tag{21}$$

When S_{1i} is relatively large, N_1 and N_i are geometrically separable compared with a given noise level $V_1^T P V_1$. Thus, the geometrical separability condition of ambiguity resolution is:

$$S_{1i} = \frac{V_{\Delta N1i}^T P V_{\Delta N1i}}{V_1^T P V_1} > k1 \tag{22}$$

The next question is how to determine k1. Since the distribution of the GNSS satellites can be considered as random, $V_{\Delta N1i}$ should obey the normal distribution just like the V_1 (Teunissen 1998b). Although we are not able to prove this hypothesis at moment, simulation tests are carried out to test if the distribution of $V_{\Delta N1i}$ is Gaussian. In the simulation, we firstly calculate three data sets of $V_{\Delta N1i}$ and consider them as the samples using the observation collected for the experiments (see section 4) in this study. The numbers of the samples are 34648. The probability density of one set of $V_{\Delta N1i}$ is shown in Fig 1 (the blue line). The mathematical expectation (μ) and standard deviation (σ) of $V_{\Delta N1i}$ is 0.000 and 0.025 respectively. The black line shows the normal probability density function with the same μ and σ. As shown in Fig 1, these two lines are very closer to each other. Furthermore, we apply the Jarque-Bera test in Matlab, h = jbtest(x), which returns a test decision for the null hypothesis that the data in vector x comes from a normal distribution. The alternative hypothesis is that it does not come from such a distribution. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. Test shows that all the three samples come from a normal distribution. As the molecular and denominator of Eqs. (21) is independent, S_{1i} obeys F-distribution, $S_{1i} \sim F(d,d)$, d is the degree of freedom. Thus, by giving a significant level, $k1$ is uniquely determined by the distribution of $S_{1i} \sim F(d,d)$.

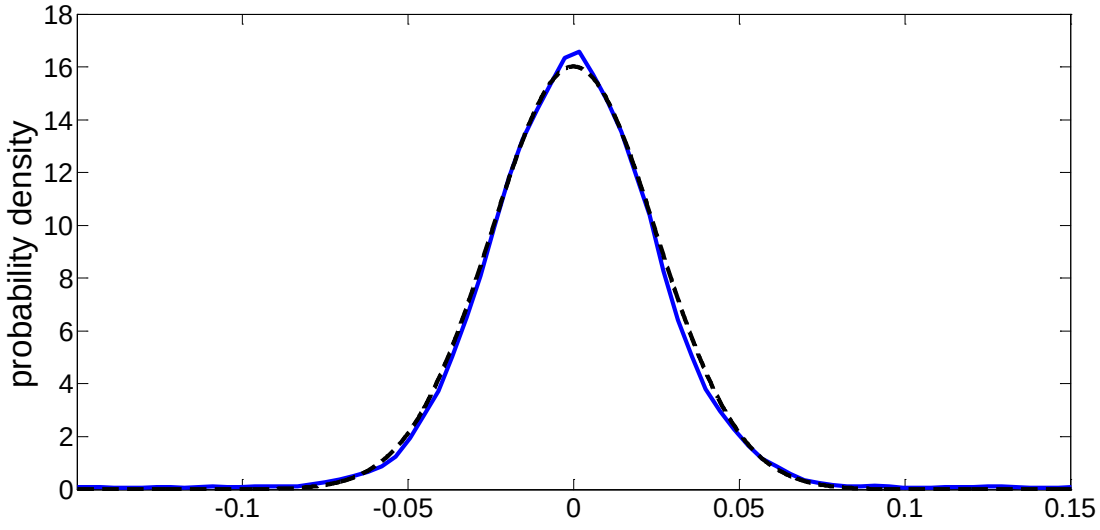


Fig 1 The distribution of $V_{\Delta N1i}$ (sample one, blue line), The Normal Distribution (black line)

The second term in Eq. (20) is determined by the size of residual of the true error e. From Eq. (20), we can get

$$\frac{2V_{\Delta N1i}^T P(I-H)e}{V_{\Delta N1i}^T PV_{\Delta N1i}} = \frac{2V_{\Delta N1i}^T PV_0}{V_{\Delta N1i}^T PV_{\Delta N1i}} > -1 \quad (23)$$

We define $M_i = m \cdot V_i, m = \frac{2V_{\Delta N1i}^T P}{V_{\Delta N1i}^T PV_{\Delta N1i}}$, and $M_0 = \frac{2V_{\Delta N1i}^T PV_0}{V_{\Delta N1i}^T PV_{\Delta N1i}} = \frac{2V_{\Delta N1i}^T P(I-H)e}{V_{\Delta N1i}^T PV_{\Delta N1i}}$.

In case 1, $V_1 = (I-H)e$, so

$$M_1 = M_0 = \frac{2V_{\Delta N1i}^T PV_1}{V_{\Delta N1i}^T PV_{\Delta N1i}} = \frac{2V_{\Delta N1i}^T PV_0}{V_{\Delta N1i}^T PV_{\Delta N1i}} > -1 \quad (24)$$

If the measurement error vector e obeys a zero mean normal distribution, M_0 also obeys normal distribution, as it is a linear combination of error vector e.

Case 2: N_0=N_i $(i \neq 1)$.

Let us consider the second case now, when N_i $(i \neq 1)$ is the true ambiguity vector $(N_0 = N_i)$. If N_1 is selected as an ambiguity solution in this case (with the smallest sum of residuals), an ambiguity mis-fixing happens.

In this case,

$$V_i = (I-H)e \quad (25)$$

$$V_1 = \lambda(I-H)\Delta N_{i1} + V_i = -V_{\Delta N1i} + V_i \quad (26)$$

$$\Delta Z_{1i} = Z_i - Z_1 = -V_{\Delta Ni1}^T PV_{\Delta Ni1} - 2V_{\Delta Ni1}^T P(I-H)e = -V_{\Delta N1i}^T PV_{\Delta N1i} + 2V_{\Delta N1i}^T P(I-H)e > 0 \quad (27)$$

where

$$\Delta N_{i1} = N_1 - N_i = -\Delta N_{1i} \quad (28)$$

Thus in this case,

$$M_0 = \frac{2V_{\Delta N1i}^T P(I-H)e}{V_{\Delta N1i}^T PV_{\Delta N1i}} > 1 \quad (29)$$

and,

$$M_1 = \frac{2V_{\Delta N1i}^T PV_1}{V_{\Delta N1i}^T PV_{\Delta N1i}} = \frac{2V_{\Delta N1i}^T P(-V_{\Delta N1i} + V_i)}{V_{\Delta N1i}^T PV_{\Delta N1i}} = -2 + M_0 > -1 \quad (30)$$

Comparing Eqs. (24) and (29), the reason for fixing to wrong ambiguity is clearly illustrated. When the projection of the residual vector of the true error e to the direction of $V_{\Delta N1i}$ is too large which causes $M_0 > 1$, a mis-fixing happens.

Now let us compare Eqs. (24) and (30). M_1 obeys a normal distribution for both cases. In case 1 when $N_0 = N_1$, M_1 ($M_1 = M_0$) obeys a zero mean normal distribution (representing a correction ambiguity vector). In case 2, M_1 obeys a normal distributions with a mean of -2 (representing an incorrect ambiguity vector), as shown in Fig. 2. With given M_1 calculated, a threshold $-1 + k2$ can be set up to decide if M_1 belongs to case 1 or case 2. When

$$M_1 > -1 + k2 \quad (31)$$

we consider M_1 belongs to case 1. Otherwise, we consider M_1 belongs to case 2. Thus M_1 can be considered as an index for ambiguity mis-fixing judgement.

For a given $k2$, the success probability P_s, the mis-fixing probability P_m, and undecided probability P_u can be calculated using Eq. (32), where P_{M1} is the probability distribution function of M_1.

$$\begin{cases} P_s = \int_{-1+k2}^{\infty} P_{M1} dx, & M_1 = M_0 \\ P_m = \int_{-1+k2}^{\infty} P_{M1} dx, & M_1 = -2 + M_0 \\ P_u = 1 - P_s - P_m \end{cases} \quad (32)$$

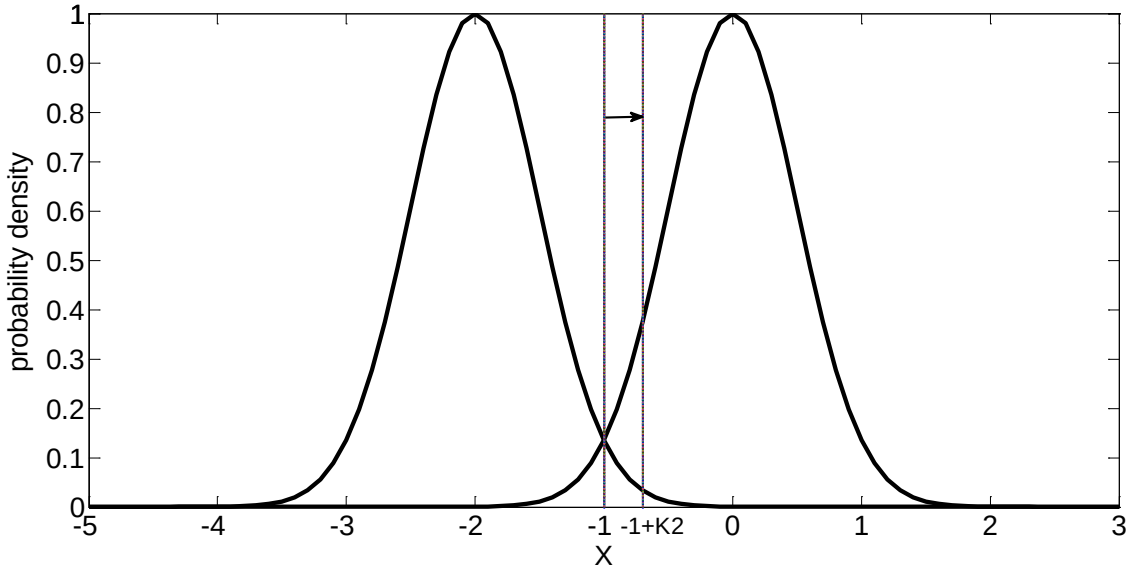


Fig 2 The distribution of M_1 in case 1 and case 2

As shown in Eq. (32), by giving the mis-fixing probability P_m, the threshold $k2$ is uniquely determined with the given variance of M_0. Assuming the true error e obeys a normal distribution, M_0 also obeys a normal distribution. The variance of M_0 can be estimated as,

$$\sigma_{M_0}^2 = \left(\frac{2V_{\Delta N1i}^T P(I-H)}{V_{\Delta N1i}^T P V_{\Delta N1i}}\right)(e \cdot e^T)\left(\frac{2V_{\Delta N1i}^T P(I-H)}{V_{\Delta N1i}^T P V_{\Delta N1i}}\right)^T \quad (33)$$

Since $e \cdot e^T = \sigma_0^2 P^{-1}$

$$\sigma_{M_0}^2 = \frac{4\sigma_0^2 V_{\Delta N1i}^T P(I-H)P^{-1}(I-H)^T P V_{\Delta N1i}}{\left(V_{\Delta N1i}^T P V_{\Delta N1i}\right)^2} = \frac{4\sigma_0^2 V_{\Delta N1i}^T P(I-H)V_{\Delta N1i}}{\left(V_{\Delta N1i}^T P V_{\Delta N1i}\right)^2} = \frac{4\sigma_0^2}{V_{\Delta N1i}^T P V_{\Delta N1i}} \quad (34)$$

Let $\sigma_0^2 \approx \frac{V_1^T P V_1}{d}$, where d is the degree of freedom of Eq. (9). Insert the equation above and Eq. (21) into (34),

$$\sigma_{M_0}^2 \approx \frac{4}{d \bullet S_{1i}} \quad (35)$$

According to Eq. (35), giving a S_{1i}, we can obtain the variance of M_0 ($\sigma_{M_0}^2$). With a given mis-fixing probability P_m and $\sigma_{M_0}^2$, the threshold $k2$ can be uniquely determined.

3. A Geometry Based Ambiguity Validation (GBAV) method

Based on the analysis in section 2, we propose a new ambiguity resolution method using both the geometrical separability condition (Eq. (22)) and the mis-fixing condition (Eq. (31)). Using both conditions, it enables to control the degree of spatial separability of the ambiguity candidates, and to control the probability of mis-fixing rates at the same time.

The ambiguity validation procedure based on the proposed GBAV method can be summarized as:

1) To avoid the big effect on the ambiguity validation of the pseudo-range noise, the GBAV method proposed in this paper is only based on the carrier phase observation. Theoretically, integer hypotheses should be followed from an ILS estimation based on Eq. (5). However, this equation is rank deficient with one epoch observation and thus long times observation will be required to get the float solutions and the ambiguity candidates. As a result, in this study, we calculate the "float" solution for the ambiguity vector with both pseudo-range and carrier phase observations, then determine the ambiguity candidate search space using the LAMBDA method. After the search range of ambiguity is determined, only carrier phase measurements will be used.
2) Check the data quality by examining residual V_1 with various receiver autonomous integrity monitoring (RAIM) fault detection and exclusion (FDE) methods (Feng et al. 2009) and remove measurements if a large error is detected. In this way, some outliers can be detected and removed from observation.

   Repeat 1) and 2) until no more errors can be found.
3) Confirm if the ambiguity vector associates with the smallest sum of residuals is the correct ambiguity, by checking ambiguity vectors with minimum and second minimum sum of residuals satisfying the separability condition (Eq. (22)), mis-fixing condition (Eq. (31)) or not,
4) If 3) are not satisfied, add one more epoch and then repeat 1) and 4).
5) When both separability condition and mis-fixing condition are satisfied, we fix the ambiguity $N_1 = N_0$.

For most conventional ambiguity validation methods (i.e. the ratio test), only one threshold is used. In GBAV algorithm, we applied two thresholds to control spatial separability and mis-fixing probability separately. The crucial issues for GBAV is the selection of the threshold $k1$ and $k2$.

For separability index S_{1i}, if we set the significant level $\alpha = 0.05$, we can estimate the threshold $k1$ with the degree of freedom d (or the number of observations v) using a F-distribution. Fig 3 gives the threshold values of $k1$ with different number of observation from 5 to 31. From Fig 3 we can see the thresholds decrease from 5.05 to 3.79 sharply when the number of observation changes from 5 to 7. When the number of observation is larger than 24, the value of F varied slowly, which is always below 2.0. Thus we simply set a table for $k1$ thresholds with different numbers of observations (Table 1). To balance the reliability and efficiency, the value of $k1$ we selected are all slightly higher than that of the corresponding values from the F distribution.

Table 1 $k1$ thresholds applied in this paper under the significant value of 0.05

v	5	6	7	8~9	10~13	14~23	>24
$k1$	5.5	4.5	4.0	3.5	3.0	2.5	2.0

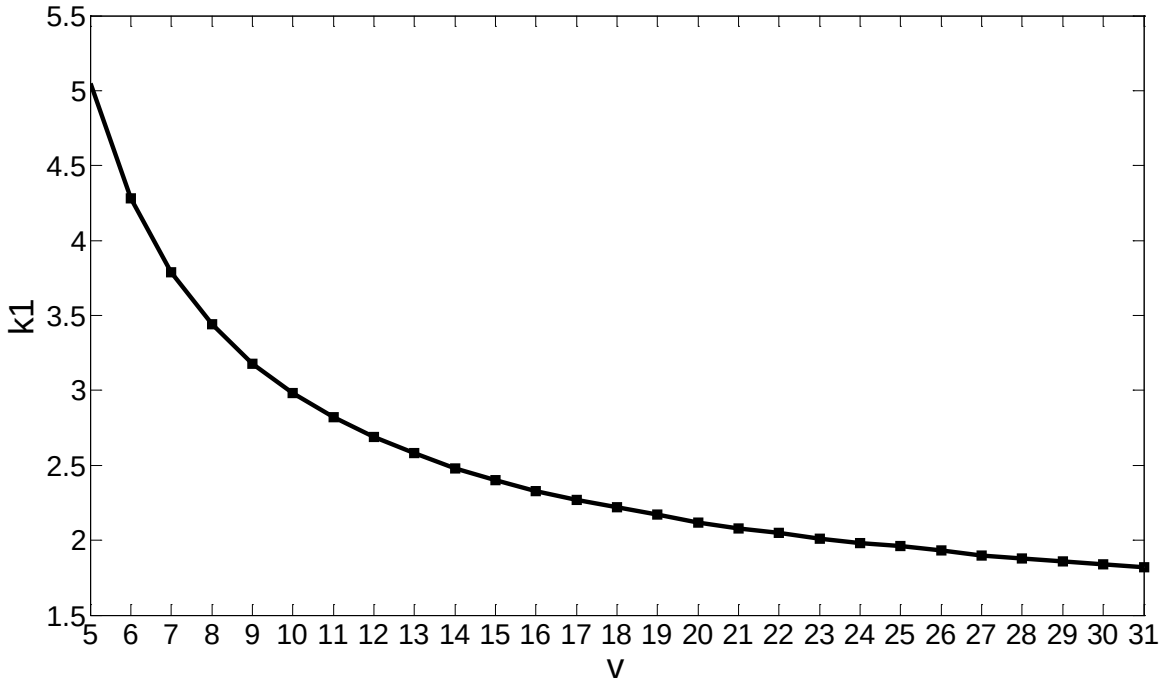


Figure 3 The thresholds of k1 with different number measurements under the significant value of 0.05

To determine the thresholds of $k2$, we need to estimate the variance of M_0 (Eq. (35)) first. We use $k1$ (low bound of S_{1i}) to replace S_{1i} in Eq. (35) and σ_M^2 for different number of observation are given in Table 2. From Table 2, we can find that σ_M^2 decrease steadily from 0.40 to 0.08 when number of observation increase from 5 to 31. Using the largest σ_M^2 which is 0.40 as an example, we can determine the threshold $k2$ with different mis-fixing probability. Table 3 illustrates the thresholds of $k2$, the probabilities of the success and undecided cases when $k2$ is equal to 0 or the mis-fixing probability is set to be 0.1% and 0.01%. As shown in Table 3, the estimation of $k2$ increases considerably from 0.00 to 0.49 when the mis-fixing rate declines. At the same time, the success rate decreases sharply from 99.38% to 90.00%, which means more time will be required to realize the ambiguity resolution when the mis-fixing probability reduces. It should be noted that the values provided here give the upper limits of mis-fixing probability. If the same thresholds are used, with lower value of σ_M^2, the mis-fixing probability will be less than that listed in Table 3.

Since the R-ratio test is very popular and widely used, we compare the GBAV method with it. The R-ratio test is defined as:

$$Ratio = \frac{V_2^T P V_2}{V_1^T P V_1} > k \tag{36}$$

Substituting Eqs. (22) and (24) into (36) yields

$$Ratio = 1 + S_{12} + S_{12} M_1 \tag{37}$$

Table 2 σ_M^2 for different numbers of observation

v	σ_M^2	v	σ_M^2	v	σ_M^2
5	0.40	14	0.15	23	0.10
6	0.31	15	0.14	24	0.10
7	0.26	16	0.13	25	0.09
8	0.23	17	0.13	26	0.09
9	0.21	18	0.12	27	0.09
10	0.19	19	0.12	28	0.09
11	0.18	20	0.11	29	0.08
12	0.17	21	0.11	30	0.08
13	0.16	22	0.10	31	0.08

It can be seen from Eq. (37) that the ratio test is a mixed parameter of spatial separation and mis-fixing index. By giving the thresholds of $k1$ and $k2$, we can obtain ratio test threshold k, which is a function of number of observation and mis-fixing probability. If we select $k1$ in Table 1, and $k2$ =0.0, 0.24 and 0.49 respectively, the thresholds for the R-ratio test varies from 1.00 to 3.20 for the number of observation from 5 to 24 (Table 4).

Table 3 The threshold of $k2$, success and undecided probabilities with $\sigma_M^2 = 0.4$

σ_M	$k2$	-1+$k2$	Success	Undecided	Mis-fixing
0.4	**0.00**	-1.00	99.38%	0.00%	0.62%
0.4	**0.24**	-0.76	97.19%	2.71%	**0.10%**
0.4	**0.49**	-0.51	90.00%	9.99%	**0.01%**

Table 4 The relationship between the thresholds for V-Ratio and GBAV test

$k2$	5	6	7	8~9	10~13	14~23	>24
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.24	2.32	2.08	1.96	1.84	1.72	1.60	1.48
0.49	3.20	2.80	2.60	2.40	2.20	2.00	1.80

Even we can use the variable thresholds of the ratio test for ambiguity validation, the GBAV method will be better that the ratio test. For example, in the case when S_{1i} is too small but M_1 is sufficiently large, the result will pass the ratio test. However, in this case, the ambiguity vectors are not spatially separable. Also, when is S_{1i} is very large, but M_1 is too small, the result will also pass the ratio test. But a mis-fixing case would be found with the GBAV method.

4. Numerical Examples

In this section, to evaluate the performance of the GBAV method proposed in this paper, three GNSS data sets with 24-hour observations are used for ambiguity resolutions. According to analyze in Ji and Xu (Ji et al.), we found that a better ambiguity resolution performance will be adopted when the cut off angle of BDS GEOs is set to 20°, and the cut off angle of IGSOs, MEOs is set to 15°. As a consequence, we set the cutoff angle for GEOs to be 20°, and the IGSOs, MEOs, as well as the GPS satellites to be 15°. In addition, the full ambiguity resolution rather than the partial ambiguity resolution is applied. Three ambiguity validation methods are used for ambiguity resolution, namely constant threshold for the ratio test, variable threshold for the ratio test (Eq. (37)), and the GBAV method. The quality of an ambiguity validation method is described by two factors, i.e. time required for ambiguity resolution and ambiguity mix-fixing rate. The first factor indicates the efficiency and the second factor represents the reliability of the validation method.

4.1 Data and Data processing methods

Two short baselines with GPS observation (GODE – GODN, 40 m baseline) and (HARB – HRAO, 1.24 km baseline) from the International GNSS Services (IGS) network, and a middle-range baseline (GS01 – GS02, 30.6 km in Beijing, China) with GPS/BDS observation were used for the evaluation of the GBAV method. For all stations, dual frequency geodetic receivers were installed at the stations. And the observation periods for all baselines are 24 hours. For the two short baselines, the update rate is 30s and for the middle range baseline the update rate is 1s.

To evaluate the performance of the proposed ambiguity validation method, we started from every epoch in the data sets until all ambiguities were fixed to their integers. In data processing, the ambiguity-fix rate (AFR) (Ji et al. 2010) is used to quantify the efficiency performance of ambiguity resolution with the following definition,

$$AFR = \frac{\text{Number of epoch with ambiguity fixed to integer}}{\text{Total number of epochs observed in the data sets}} \quad (38)$$

Also, all mis-fixing cases were recorded and quantified as the percentage of total observed epochs during the 24 hour observation period. In the data processing, we did not estimate σ_M^2 every epoch. Instead, we used the largest value of 0.4 for all processing. To set a baseline for comparison, we used the fixed R-ratio of 1, 1.5, 2.0, 2.5, 3.0, and 3.5. Then for the GBAV method, the thresholds of $k1$, and $k2$, are adopted from Tables 1 and 3. For the variable R-ratio test (Eq. (37)), the thresholds are given in Table 4. It is worth mentioning that, the observation of both GPS and Beidou for all the baselines is double-frequency signals. So the number of the observation in Tables 1 and 4 is generally larger than 8.

4.2 Test results

Baseline 1 (GODE – GODN, 40m)

This is a very short baseline and most of the errors can be effectively cancelled by double differencing. Table 5 gives the ambiguity resolution results with the fixing threshold ratio test. As shown in the table, when the ratio is large than 3.0, there is no mis-fixing case, and the time required for all epochs ambiguity fixed are 5-8 epochs or 3-4 min. More than 97% of epochs the ambiguities can be fixed within one epoch. On the other hand, with the ratio threshold less than 2.5, there are some mis-fixing cases.

The results with variable ratio test and the GBAV method are given in Table 6. When $k_2 = 0.00$, the threshold of R-ratio should be 1.0, and only 1 epoch is needed to fix the ambiguities, which is more efficient than that of the GBAV test. However, the mis-fixing rate (0.14%) is significant higher than that of the GBAV method (0.07%). Since there are 2880 epochs, 4 of them are mis-fixed shown as Table 7. The values of S and M at the first epoch when the Ratio suffer from a mis-fixing are shown in Table 8. It is shown that the

values of S in the first, third and last cases are smaller than K1, which means the ambiguity candidates of these cases cannot be spatial separated. When the GBAV test is involved, $k2 = 0.00$, the ambiguities of the last two cases in Table 8 are fixed to the right ones. When $k2$ =0.24, the first two cases in Table 8 also achieve the ambiguity resolution correctly, since the M controls the mis-fixing probability. On the other hand, there are 2 epochs (0.07%) of mis-fixing cases using the variable ratio test when $k2$ =0.24 and 0.49 respectively. Take $k2$ =0.49 for instance, the time required for all epoch's ambiguity fixed are 5 to 8 epochs or 2.5 to 4.5 min with the GBAV method. And the variable ratio only need 5 epochs. If we just check one epoch data, the variable ratio method can fix 99% of epochs while the GBAV method can fix around 91% when $k2$ =0.49. Thus, the ratio methods are more efficient than the GBAV method on ambiguity resolution.

Table 5 AFR and Mis-fixing rate of *R*-ratio test with certain values of threshold (GODE – GODN, 40m)

T_f	***Fix R*-ratio AFR (%)**					
	$k = 1$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 3.0$	$k = 3.5$
1	100.00	99.69	99.24	98.72	97.53	96.56
2		100.00	99.90	99.76	99.41	98.89
3			99.93	99.90	99.65	99.31
4			100.00	99.97	99.69	99.44
5				100.00	99.72	99.48
>5 & ≤ 8					100.00	100.00
Mis-fixing rate(%)						
	0.14	0.14	0.07	0.03	0.00	0.00

Note: T_f stands for Time to fix (epoch)

Table 6 AFR and mis-fixing rate of GBAV and variable *R*-ratio test with varying threshold (GODE – GODN, 40m)

T_f	***Variable R*-ratio AFR (%)**			**GBAV AFR (%)**		
	$k2 = 0.00$	$k2 = 0.24$	$k2 = 0.49$	$k2 = 0.00$	$k2 = 0.24$	$k2 = 0.49$
1	100.00	99.51	99.24	99.48	98.23	91.74
2		99.93	99.86	99.76	99.62	97.67
3		99.97	99.93	100.00	99.72	98.26
4		100.00	99.97		99.93	99.10
5			100.00		100.00	99.38
>5 & ≤ 8						100.00
Mis-fixing rate(%)						
	0.14	0.10	0.07	0.07	0.00	0.00

Table 7 Mis-fixing cases for ratio test (GODE – GODN, 40m)

GPS Time			$k2 = 0.00$ / 1epoch		$k2 = 0.24$		$k2 = 0.49$	
h	m	s	K	ratio	K	ratio	K	ratio
13	25	30	1.00	1.502	1.72	2.382	2.20	3.780
13	29	30	1.00	1.533	1.84	2.163	2.40	5.620
13	55	30	1.00	3.654	1.84	3.654	2.40	3.654
17	25	00	1.00	2.786	1.84	2.786	2.40	2.786

Table 8 Mis-fixing cases for GBAV test (GODE – GODN, 40m)

GPS Time			K1	1 epoch			$k2 = 0.00$		$k2 = 0.24$	
h	m	s		Ratio	S	M	S	M	S	M
13	25	30	3.0	1.502	2.488	-0.798	5.603	-0.863	9.068	-0.119
13	29	30	3.5	1.533	7.966	-0.933	7.966	-0.933	4.484	-0.616
13	55	30	3.5	3.654	2.904	-0.086	9.297	0.086	9.297	0.086
17	25	00	3.5	2.786	2.480	-0.280	7.307	-0.380	7.307	-0.380

Baseline 2 (HARB – HRAO, 1.2 km).

Again, we applied the fix ratio test first and the results are given in Table 9. In this example, for the fix ratio method, when the ratio is larger than 3.5, there are no mis-fixing cases. The time required for all epochs fixed are less than 25 epochs or 12.5 min.

When we use the GBAV method with $k2$ =0.49, there are no mis-fixing case and the time required for all epochs fixed are less than 15 epochs or 7.5 min (Table 10). The variable ratio test can fix ambiguity for all epochs during the same period, but there are 3 epochs of mis-fixing (0.10%). On the other hand, for the efficiency of ambiguity fixing, variable ratio test is slightly better.

Baseline 3 (GS01 – GS02, 30.6 km).

As this dataset include both GPS and Chinese BeiDou data, we consider two cases here: GPS only and GPS/BDS data. When we only use GPS data, the results are summarized in Tables 11 and 12. With the fix ratio test, the threshold with no mis-fixing cases is 3.5, and 15 min to resolve ambiguities for all epochs. When $k2$=0.24 and 0.49, the GBAV method can fix ambiguity with no mis-fixing cases. However, the variable ratio test suffers from 7.35% and 1.35% mis-fixing rate when $k2$=0.24 and 0.49. Again, for the efficiency of ambiguity fixing, the variable ratio test is slightly better.

Table 9 AFR and Mis-fixing rate of *R*-ratio test with certain values of threshold (HARB – HRAO, 1.24Km)

T_f	***Fix R*-ratio AFR (%)**					
	k = 1.0	k = 1.5	k = 2.0	k = 2.5	k = 3.0	k = 3.5
≤ 1	100.00	89.97	87.85	81.35	78.09	71.18
≤2		93.47	93.19	88.68	88.44	82.26
≤3		95.03	94.65	91.70	91.39	86.73
≤4		95.66	95.42	93.54	92.78	89.13
≤10		99.41	99.34	98.85	98.68	97.12
≤25		100.00	100.00	100.00	100.00	100.00
Mis-fixing rate(%)						
	8.33	3.47	0.49	0.10	0.03	0.00

Table 10 AFR and mis-fixing rate of GBAV and *R*-ratio test with varying threshold (HARB – HRAO, 1.24Km)

T_f		***Variable R*-ratio AFR (%)**			**GBAV AFR (%)**		
		$k2$ = 0.0	$k2$ = 0.2	$k2$ = 0.4	$k2$ = 0.0	$k2$ = 0.2	$k2$ = 0.4
≤1	100.00	89.03	85.97	90.91	80.10	72.36	
≤2		93.51	89.03	99.79	88.23	83.37	
≤3		95.03	89.03	100.00	91.84	86.67	
≤4		95.63	93.58		92.95	88.72	
≤10		99.38	99.06		98.37	95.83	
≤15		100.00	100.00		100.00	100.00	
Mis-fixing rate(%)							
	8.33	**0.31**	**0.10**	**4.20**	**0.07**	**0.00**	

Table 11 AFR and Mis-fixing rate of *R*-ratio test with certain values of threshold (GS01 – GS02, 30.6 km, GPS only)

T_f		***Fix R*-ratio AFR (%)**					
		k = 1	k = 1.5	k = 2.0	k = 2.5	k = 3.0	k = 3.5
0	1	100.00	26.51	7.24	2.28	0.62	0.20
≤1	60		68.32	32.68	11.62	2.72	0.40
≤3	180		81.88	65.72	51.34	33.07	23.96
≤5	300		90.66	77.03	65.65	58.41	45.00
≤10	600		100.00	100.00	99.36	86.16	71.21
≤13	780				100.00	99.32	83.26
≤15	900					100.00	100.00
Mis-fixing rate(%)							
		49.98	22.38	6.16	0.15	0.05	0.00

Table 12 AFR and mis-fixing rate of GBAV and *R*-ratio test with varying threshold (GS01 – GS02, 30.6 km, GPS only)

T_f		***Variable R*-ratio AFR (%)**			**GBAV AFR (%)**		
		$k2$ = 0.00	$k2$ = 0.24	$k2$ = 0.49	$k2$ = 0.00	$k2$ = 0.24	$k2$ = 0.49
0	1	100.00	25.83	5.33	15.70	11.37	0.18
≤1	60		62.76	22.69	72.96	15.43	0.29
≤3	180		72.65	60.56	92.21	47.63	23.02
≤5	300		86.30	70.38	96.16	63.21	42.25
≤10	600		100.00	99.62	100.00	86.33	70.23
≤ 13	780			100.00		100.00	82.06
≤ 15	900						100.00
Mis-fixing rate(%)							
		49.98	7.35	1.56	40.49	0.00	0.00

For the same baseline, the ambiguity resolution performance is much better when using both GPS/BeiDou data (Table 13 and 14). With the fix ratio test (Table 13), when the threshold is larger than 2.5, there are no mis-fixing cases and the time required for 100% ambiguity fixing is only about 30 epochs or 0.5 min. When the variable ratio test and the GBAV method apply (Table 14), there are no mis-fixing cases when $k2$ =0.49. The time require for 100% epoch ambiguity fixing is only 15s.

Table 13 AFR and Mis-fixing rate of R-ratio test with certain values of threshold (GS01 – GS02, 30.6 km, GPS+ BDS)

T_f		***Fix R*-ratio AFR (%)**					
		$k = 1$	$k = 1.5$	$k = 2.0$	$k = 2.5$	$k = 3.0$	$k = 3.5$
0	1	100.00	90.42	90.31	86.55	75.12	57.20
≤1/30	2		96.08	96.08	92.65	81.71	63.71
≤1/12	5		99.35	99.35	96.52	87.15	70.99
≤1/6	10		100.00	100.00	98.26	90.20	76.14
≤1/4	15				99.18	92.11	79.20
≤1/2	30				100.00	94.99	84.29
≤4	240					100.00	100.00
Mis-fixing rate(%)							
		8.63	0.12	0.05	0.00	0.00	0.00

Table 14 AFR and mis-fixing rate of GBAV and *R*-ratio test with varying threshold (GS01 – GS02, 30.6 km, GPS+ BDS)

T_f		***Variable R*-ratio AFR (%)**			**GBAV AFR (%)**		
		$k2 = 0.00$	$k2 = 0.24$	$k2 = 0.49$	$k2 = 0.0$	$k2 = 0.24$	$k2 = 0.49$
0	1	100.00	90.53	87.72	72.08	85.38	60.26
≤ 1/30	2		96.08	94.33	90.71	90.46	81.13
≤ 1/12	5		99.35	97.58	100.00	93.88	89.68
≤ 1/6	10		100.00	99.69		100.00	92.57
≤ 1/4	15			100.00			100.00
		8.63	0.05	0.00	4.31	0.00	0.00

From the above examples, we can see that when the threshold is high enough, the fix ratio test can achieve no mis-fixing case for all the test data. However, for different datasets, the thresholds vary from 2.5-3.5. If we use 3.5 for all the cases, it required almost 8 times more observation time for fixing ambiguity for all epochs than that with the threshold of 2.5 in the GPS/BDS case (Table 13). When the GBAV method is used, with $k2$=0.49, there is no mis-fixing case for all the datasets tested. This demonstrates that the GBAV method can effectively control mis-fixing probability. With the variable ratio method, the ambiguity fixing efficiency is generally better that that of the GBAV method, but there are a number of cases of mis-fixing on the three baselines with GPS only observation.

5. Conclusions

In this paper, we introduced two new concepts for ambiguity validation, i.e. spatial separability condition S_{1i} and mis-fixing condition M_1. By using these two concepts, we can understand why ambiguity mis-fixing occurs. If the satellite geometry is not strong enough, there may be a few ambiguity combinations which are not be able to be separated under the existing measurement noise level. Moreover, if the projection of true measurement error residuals to the direction of $V_{\Delta N1i}$ is too large which causes M_0>1, an ambiguity mis-fixing happens. The conventional ambiguity validation methods, such as ratio test and difference test, are the combinations of spatial separability condition S_{1i} and mis-fixing condition M_1. The distributions of S_{1i} and M_1 can be strictly defined which are the functions of the measurement quality, the number of observed satellites, and the satellite geometry. This enables us to set up the thresholds based on user requirements for the quality control the quality of ambiguity resolution.

Based on these concepts, we proposed a new geometry based ambiguity validation (GBAV) method which will ensure different ambiguity combinations to be both geometrically separable and mis-fixing probability controlled. The distributions and threshold computation methods for S_{1i} and M_1 are given in the paper, with given a significant value for S_{1i} and a mis-fixing probability for M_1.

The thresholds for the traditional ratio and difference tests are normally determined empirically as the statistical distributions are difficult to obtain. In this paper, we have shown that the traditional ratio and difference tests are the mixture of spatial separability condition and mis-fixing condition. By applying the same concepts, we can calculate the variable thresholds for the both methods, with given observation number and mis-fixing probability. However, with these single threshold methods, it is possible to have some mis-fixing cases when two ambiguity vectors are not geometrically separable, or M_0 is too small.

To evaluate the performance of the proposed GBAV method, three GNSS datasets with 24-hour observation are processed, using the fix and variable threshold ratio tests as a comparison. It is found that to achieve no mis-fixing for all epochs, the thresholds for different datasets are different. If the thresholds are increased too high, the efficiency of ambiguity resolution can drop significantly. Using the concepts proposed by this paper, when we select the mis-fixing probability less than 0.01% (or $k2$=0.49), there is no mis-fixing case with the GBAV method for all three datasets. However, there are a few cases of mis-fixing for the variable ratio test. On the other hand, the ambiguity fixing efficiency for the variable ratio test is slightly better than that of the GBAV method.

Also, combining GPS/BDS systems, the ambiguity resolution performance can be significantly improved

for medium-range baselines. For a 30 km baseline, it requires 15 min for all epoch ambiguity fixed with GPS data only. With GPS/BDS data, the time for all epoch ambiguity fixed can be reduced to 15s.

Acknowledgments

This research was supported by the University Grants Committee of Hong Kong under the scheme of Competitive Earmarked Research Grant (Project No: 152223/18E) and the research fund from the Smart Cities Research Institute, Hong Kong Polytechnic University.

References

Euler HJ, Landau H (1992) Fast GPS ambiguity resolution on-the-fly for real-time applications. In: Proceedings of 6th Int. Geod. Symp. on satellite Positioning, Columbus, Ohio. pp 650–659

Euler H-J, Schaffrin B (1991) On a measure for the discernibility between different ambiguity solutions in the static-kinematic GPS-mode. In: Kinematic Systems in Geodesy, Surveying, and Remote Sensing. Springer, pp 285–295

Feng S, Ochieng W, Moore T, et al (2009) Carrier phase-based integrity monitoring for high-accuracy positioning. GPS Solut 13:13–22.

Frei E, Beutler G (1990) Rapid static positioning based on the fast ambiguity resolution approach FARA: theory and first results. Manuscripta Geod 15:325–356.

Han S (1997) Quality-control issues relating to instantaneous ambiguity resolution for real-time GPS kinematic positioning. J Geod 71:351–361.

Han S, Rizos C (1999) The impact of two additional civilian GPS frequencies on ambiguity resolution strategies. In: 55th National Meeting US Institute of Navigation, “Navigational Technology for the 21st Century”, Cambridge, Massachusetts. pp 28–30

Hofmann-Wellenhof B, Lichtenegger H, Collins J (1993) Global Positioning System. Theory and Practice. Springer-verlag

Ji S, Chen W, Ding X, et al (2010) Ambiguity validation with combined ratio test and ellipsoidal integer aperture estimator. J Geod 84(10):597–604.

Ji S, Wang X, Xu Y, et al (2014) First Preliminary Fast Static Ambiguity Resolution Results of Medium-Baseline with Triple-Frequency Beidou Wavebands. J Navig 67(6): 1109-1119

Leick A (2004) GPS satellite surveying. John Wiley & Sons

Li T, Wang J (2012) Some remarks on GNSS integer ambiguity validation methods. Surv Rev 44:230–238.

Parkinson B, Spilker JJ, Axelrad P, Enge P (1996) GPS: theory and applications, vols 1 and 2. American Institute of Aeronautics and Astronautics

Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geod 70(1-2):65–82.

Teunissen PJG (1999) An optimality property of the integer least-squares estimator. J Geod 73(11):587–593.

Teunissen PJG (2013) GNSS integer ambiguity validation: overview of theory and methods. Proc Inst Navig Pac PNT 673–684.

Teunissen PJG (2003) A carrier phase ambiguity estimator with easy-to-evaluate fail-rate. Artif Satell 38:89–96.

Teunissen PJG (2004) Penalized GNSS ambiguity resolution. J. Geod., 78(4-5):235–244.

Teunissen PJG (1998) On the integer normal distribution of the GPS ambiguities. Artif Satell 33:49–64.

Teunissen PJG (2000) ADOP based upperbounds for the bootstrapped and the least-squares ambiguity success rates. Artif Satell 35:171–179.

Teunissen PJG (2005) Penalized GNSS ambiguity resolution with optimally controlled failure-rate. Artif Satell 40:219–227.

Teunissen PJG, Verhagen S (2009) The GNSS ambiguity ratio-test revisited: a better way of using it. Surv Rev 41(312):138–151.

Teunissen PJG, Verhagen S (2004) On the foundation of the popular ratio test for GNSS ambiguity resolution. In: Proc. ION GNSS. pp 2529–2540

Tiberius C, De Jonge PJ (1995) Fast positioning using the LAMBDA method. Proc. DSNS 95, Norway, Paper No. 30

Verhagen S (2004) Integer ambiguity validation: An open problem? GPS Solut 8:36–43.

Verhagen S (2005) On the reliability of integer ambiguity resolution. Navigation 52:99–110.

Verhagen S, Teunissen PJ (2013) The ratio test for future GNSS ambiguity resolution. GPS Solut 17:535–548.

Verhagen S, Teunissen PJ (2006) New global navigation satellite system ambiguity resolution method compared to existing approaches. J Guid Control Dyn 29:981–991.

Wang J, Stewart MP, Tsakiri M (1998) Stochastic modeling for static GPS baseline data processing. J Surv Eng 124:171–181.

Xu PL, Cannon E, Lachapelle G (1995) Mixed integer programming for the resolution of GPS carrier phase ambiguities, presented at IUGG95 Assembly, Boulder, July 2–14.

Authors

Wu Chen is a Professor at the Department of Land Surveying and Geo-Informatics in the Hong Kong Polytechnic University. His interests include GNSS Applications on Transportation, Kinematic GPS, System Integration, GNSS Performance Evaluation, GPS Software Receiver, Regional GPS Network, Vehicle and Personal Navigation Systems, and Wireless Sensor Network.

Ying Xu is an Associate Professor Shandong University of Science and Technology. She received a PhD degree from the Hong Kong Polytechnic University in 2016. Her research focuses on the algorithm development for high precision GNSS.

Duojie Weng is a Research Assistant Professor with the Hong Kong Polytechnic University. His interests included GNSS, indoor positioning and precise positioning. He received the PhD from the Hong Kong Polytechnic University in 2016, and MSc and BSc degrees in Electrical Engineering from Hohai University in China.

Shengyue Ji is an Associate Professor in China University of Petroleum (East China). His research interests include precise GNSS positioning and the inonosphere in low latitude areas etc. He has published more than 30 papers in international famous journals as the first or corresponding author, such as Journal of Geophysical Research, Journal of Geodesy, GPS Solutions, etc. And he has made scientific contributions many GNSS aspects, such as precise Ocean navigation and positioning, ambiguity resolution and validation, cycle slip detection and repairing, precise point positioning, data quality control etc.

Journal of Global Positioning Systems (*2022*)

Vol. 18, No. 1: 41-55

DOI: 10.5081/jgps.18.1.41

Journal of Global Positioning Systems
2022
© CPGPS

Comparative analysis of data quality and performance index for BDS-3 constellation

Zhipeng Ding[1,2], Kaifei He[1,2,*], Ming Li[1,2], Yu Wu[1,2], Yue Zhang[1,2] and Jinquan Yang[1,2]

1. College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China
2. Technology Innovation Center for Maritime Silk Road Marine Resources and Environment Networked Observation, Ministry of Natural Resources, Qingdao, China

* Corresponding author: Kaifei He, kfhe@upc.edu.cn

Abstract: The global BeiDou-3 Navigation Satellite System (BDS-3) was completed in July 2020. In terms of data processing, the final positioning and baseline solving results will be affected by the quality of the raw observation data. Therefore, it is necessary to analyse and evaluate the data quality of the complete BDS-3 constellation and its service performance. Based on all observing satellites and the open signals from MGEX stations that can track BDS-3, improved software is used to analyse the complete BDS-3 constellation and signals. Moreover, the service performance of BDS-3 is evaluated using self-developed software. The geometric configuration of the complete BDS-3 constellation is found to be slightly better than that of GPS. However, the overall multipath error is about 10 cm higher than that of GPS, although the increased choke of the measured maritime data effectively weakens the multipath error. The pseudorange multipath error of each signal runs in the order B1I>B2a>B2b>B3I>B2a+B2b>B1C; other quality indicators exhibit little difference among bands. In terms of service performance, the carrier phase residuals are 0.17-0.48cm. After data convergence, the relative positioning performance fluctuates around 5 cm of the "true value", although the fluctuations in the vertical direction are up to 10 cm.

Key words: complete BDS-3 constellation; data quality; service performance evaluation

1. Introduction

The first BeiDou-3 Navigation Satellite System (BDS-3) modules were officially launched in 2009, and the last satellite of the network was added in 2020. The 30 satellites that constitute BDS-3 include 24 medium-circle earth orbit (MEO) satellites, three geostationary earth orbit (GEO) satellites, and three inclined geosynchronous orbit (IGSO) satellites [1-3]. BDS-3 provides satellite signals at various frequencies, with the two open-service signals of B1I (1561.098 Hz) and B3I (1268.52 Hz) in the B1 and B3 frequency bands, the B1C frequency band centred on 1575.420 MHz (the same as GPS L1 and Galileo E1), and the B2a frequency band centred on 1176.450 Hz (the same as GPS L5, QZSS L5, IRNSS L5, and Galileo E5a). The most recently launched satellite added the B2a+B2b (1191.795 Hz) signal [4-7]. On December 27, 2019, the B2b signal interface document was published, disclosing two B2b (1207.14 Hz) signals, which provide basic navigation services, and the PPP-B2b signal, which provides precision single-point positioning services [8].

As the basis of BDS data processing, the quality of the original data directly affects the final positioning results or baseline solution. Extensive analysis of the BDS data quality has been conducted, including comparisons with other systems. In terms of quality evaluation, Cai *et al.* (2016) analysed the

Editor in charge: Dr. Baocheng Zhang

noise and multipath effect of BDS-2 based on the zero-baseline double difference, and found that the noise of the B1 band has the largest pseudorange multipath error (mean error of 0.36 cm), whereas the noise of the carrier phase in different frequency bands varies from 0.9-1.5 mm [9]. In 2018, several BDS-3 satellites were launched. Yang *et al.* (2019) studied the basic performance of BDS-3, and reported that the post-processing of the orbit can reach centimetre-level accuracy. Additionally, the average satellite clock offset uncertainty of 18 MEO satellites was found to be 1.55 ns, and the mean ranging error of space signals was about 0.474 m. A method to improve the positioning, navigation, and timing services was subsequently developed [10]. In 2019, some BDS-3 satellites added new signals. Dai *et al.* (2019) studied the noise and multipath level of these BDS-3 signals and satellites, and showed that the standard deviation (STD) of the pseudorange noise in the B1I, B3I, B1C, and B2a bands was 7.4, 6.7, 14, and 13 cm, respectively, whereas the STD of the carrier phase noise was 1.84, 1.85, 1.85, and 1.85 cm, respectively. The STD of the pseudorange multipath errors in bands B1I, B3I, B1C, and B2a was 0.34, 0.21, 0.48, and 0.33 m, respectively [11]. In terms of data service performance, the pseudo-single point positioning and precise single-point positioning (PPP) were tested by Mu *et al.* (2020), who showed that BDS-3 has a slightly lower positioning accuracy than GPS and Galileo, but performs better than GLONASS [12]. Zhang *et al.* (2019) combined the BDS-3 measurement data and showed that the ambiguity resolution efficiency of RTK could be improved by incorporating the BDS-3 measurements, where by the success rate increased from 88.5 to 91.4%. The convergence time of the PPP algorithm was shortened from about 1 h to less than 30 min, and the positioning accuracy was significantly enhanced. Both BDS-3 and GPS can provide centimetre-level dynamic positioning accuracy [13]. Different signal frequency bands exhibit different data service performance. Zhu *et al.* (2021) analysed the new B1C and B2a signals of BDS-3, and found that the positioning performance was comparable to that of GPS and Galileo [14].

Most previous research has focused on BDS-2 or subsets of the BDS-3 satellites. Given the limitations of existing software and the scarcity or incompleteness of signal data, the data quality and basic evaluation methods of BDS-3 have not been systematically tested since the completion of the network. Existing studies have only examined a few BDS-3 satellites and certain signal bands; in particular, the data quality of satellites above C37 has rarely been studied. In this paper, based on all the observable MGEX in-orbit satellite data of BDS-3 and the data quality of the associated signals, indicators related to observation data quality and data service performance are studied and compared with GPS. Comparing and evaluating the complete BDS-3 data quality not only provides a systematic summary of the complete global BDS-3 network, but also lays the foundation for studying the application of BDS-3 in the production process.

2. Architecture of BDS-3 and Dataset Description

Since February 2022, more than 500 IGS tracking stations have been in operation around the world. GNSS multi-mode tracking stations provided the GNSS experimental data for the study of BDS-3 data quality and evaluation of the data performance. The data used in the experiment were divided into static and dynamic data. The static data were taken from the MGEX IGS stations, which can receive all BDS-3 signals. The specific station information is listed in Table 1. The selected observation data were from day of year (DOY) 33-42, 2021. To analyse the global data service performance of BDS-3, data were selected from two stations located in China (WUH2 and URUM) and five stations in other countries (POTS, SGOC, SUTM, ULAB, and WIND).

The satellites that participated in the calculations at each station are listed in Table 2. The GNSS data for the shipboard dynamic experiment were collected from the offshore waters of Tangdao Bay, Qingdao, China, near to the China University of Petroleum (East China), on December 16, 2021, over an observation duration of 2 h. A choke coil was installed on the antenna, as shown in Figure 1.

Table 1 Selected IGS station information

Site	Location	latitude /N	longitude /E	Height/m	Receiver	Antenna
POTS	Germany	52.379	13.066	144.4	JAVAD TRE_3	JAVRNGANT_G5T-NONE
SGOC	Sri Lanka	6.892	79.874	-78.5		
SUTM	South Africa	-32.381	20.811	1797.6		
ULAB	Mongolia	47.865	107.052	1575.7		
URUM	Urumqi	43.808	87.601	858.9		
WIND	Namibia	-22.575	170.189	1734.7		
WUH2	Wuhan	30.532	114.357	25.8		

Table 2 BDS-3 satellites used in the calculations

Site	Satellites involved in the solution
POTS	C19-C30 C32-C46 C60
SGOC	
SUTM	
ULAB	
URUM	
WIND	C20 C21 C23-C30 C32 C33 C34 C36 C37 C38 C40-C46 C60
WUH2	C19-C30 C32-C46

Fig. 1 Dynamic station setup

3. BDS-3 Data Quality Evaluation

3.1 Multipath Error

The multipath combination is a geometry-free and ionosphere-free combination formed by a one-frequency code and two-frequency phase measurements. It contains the combined noise and multipath errors of the code and phase measurements. However, because the noise and multipath errors of the phase are somewhat smaller than those of the code, the combination series mainly reflects the code noise and its multipath error. The GNSS dual-frequency multipath error is usually evaluated using a linear combination of the pseudorange and carrier [15]. The specific calculation can be written as:

$$MP_1 = P_1 - \left(1 + \frac{2}{\alpha - 1}\right)L_1 + \frac{2}{\alpha - 1}L_2 = M_1 + \mathrm{C}_1 - \left(1 + \frac{2}{\alpha - 1}\right)m_1 + \frac{2}{\alpha - 1}m_2 \quad (1)$$

$$MP_2 = P_2 - \left(\frac{2\alpha}{\alpha - 1}\right)L_1 + \left(\frac{2\alpha}{\alpha - 1} - 1\right)L_2 = M_2 + \mathrm{C}_2 - \frac{2}{\alpha - 1}m_1 + \left(\frac{2\alpha}{\alpha - 1} - 1\right)m_2 \quad (2)$$

$$\mathrm{C}_1 = -\left(1 + \frac{2}{\alpha - 1}\right)n_1\lambda_1 + \left(\frac{2}{\alpha - 1}\right)n_2\lambda_2 \quad (3)$$

$$C_2 = -\left(\frac{2\alpha}{\alpha-1}\right)n_1\lambda_1 + \left(\frac{2\alpha}{\alpha-1}-1\right)n_2\lambda_2 \quad , \quad \alpha = \frac{f_1^2}{f_2^2} \tag{4}$$

where *MP* represents the multipath error, *M* is the pseudorange multipath error, and *m* is the carrier phase multipath error. When there is no cycle skip, *C* is a constant. *MP* is mainly affected by the pseudorange multipath, because the values of m_1 and m_2 are much smaller than M_1 and M_2, which are used to measure the multipath effect. The multipath errors of BDS-3 visual satellites in all frequency bands were analysed, and the results from each observation station were averaged over 10 consecutive days and compared with GPS. The results are shown in Figures 2 and 3.

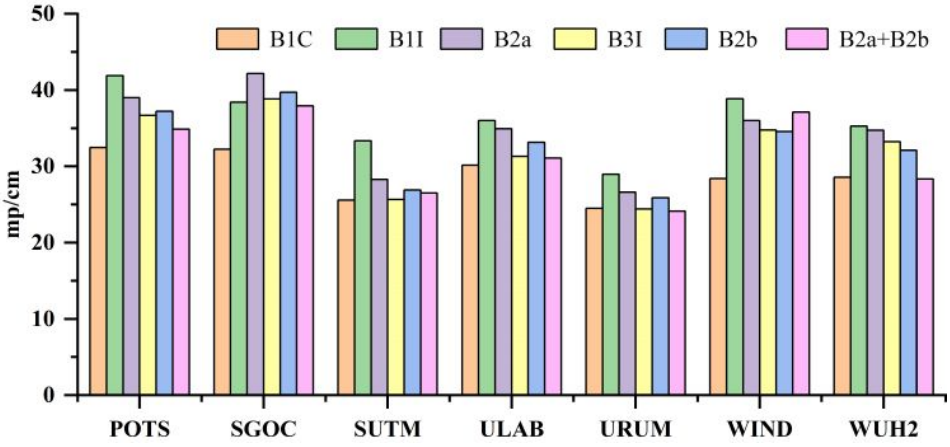


Fig. 2 BDS-3 pseudorange multipath error of station

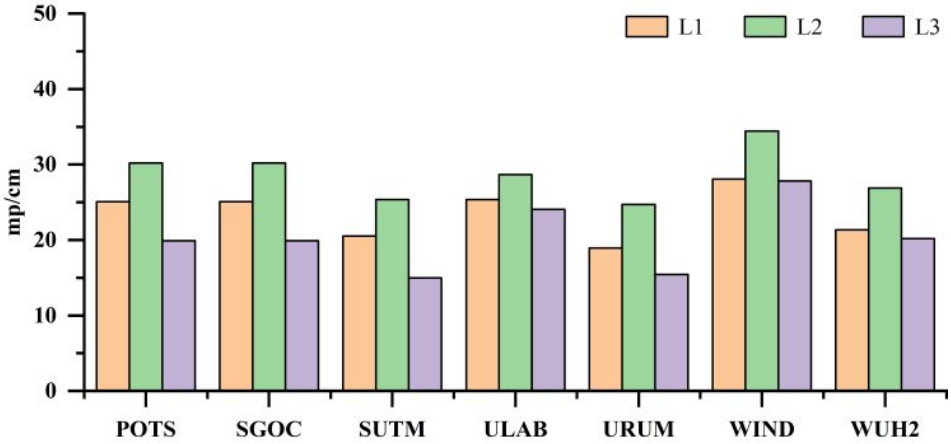


Fig. 3 GPS pseudorange multipath error of station

(1) The multipath errors of each frequency band are slightly different, and the pseudorange multipath errors of each frequency band of BDS-3 run in the order B1I>B2a>B2b>B3I>B2a+B2b>B1C. The largest multipath errors are in bands B1I and B2a, and the maximum value of 42.16 cm appears in band B2a at station SGOC. The smallest multipath error is in band B1C, and the minimum value is 24.10 cm at station URUM.

(2) The multipath error of each frequency band of BDS-3 is worse than that of the GPS data from the same station over the same period. The maximum pseudorange multipath error of GPS is 34.44 cm in band L2 at station WIND, some 7.62 cm smaller than the equivalent value for BDS-3. The minimum GPS pseudorange multipath error of 15.00 cm occurs in band L5 at station URUM. This value is 9.10 cm smaller than the equivalent for BDS-3. Therefore, the multipath error of BDS-3 requires further study.

(3) The pseudorange multipath errors of BDS-3 and GPS exhibit similar trends. For instance, the multipath errors at stations SUTM and URUM are

small, whereas those at stations POTS, SGOC, and WIND are large. These trends are related to the environment and observation conditions of a particular station at that time.

To solve the problem of insufficient research on BDS-3 satellites above C37, four satellites were selected for further analysis: C40 (IGSO), C44 (MEO), C59 (GEO), and G30 (MEO). For these BDS-3 satellites, Figure 4 shows the variation in the multipath effect with respect to elevation angle for each frequency at station WUH2. The multipath errors of the IGSO and MEO satellites exhibit opposite trends with respect to elevation angle. This is mainly because, as the satellite enters or leaves the observation field of view, serious multipath errors and noise occur. The MEO satellite loses significant amounts of data through signal interruption; because the GEO satellite experiences little change in elevation angle, the corresponding multipath error is small and varies gently. Comparing all frequency bands, B1I has the most serious multipath effect among the different orbital types; the GPS satellite and MEO satellite of BDS-3 exhibit the same trend.

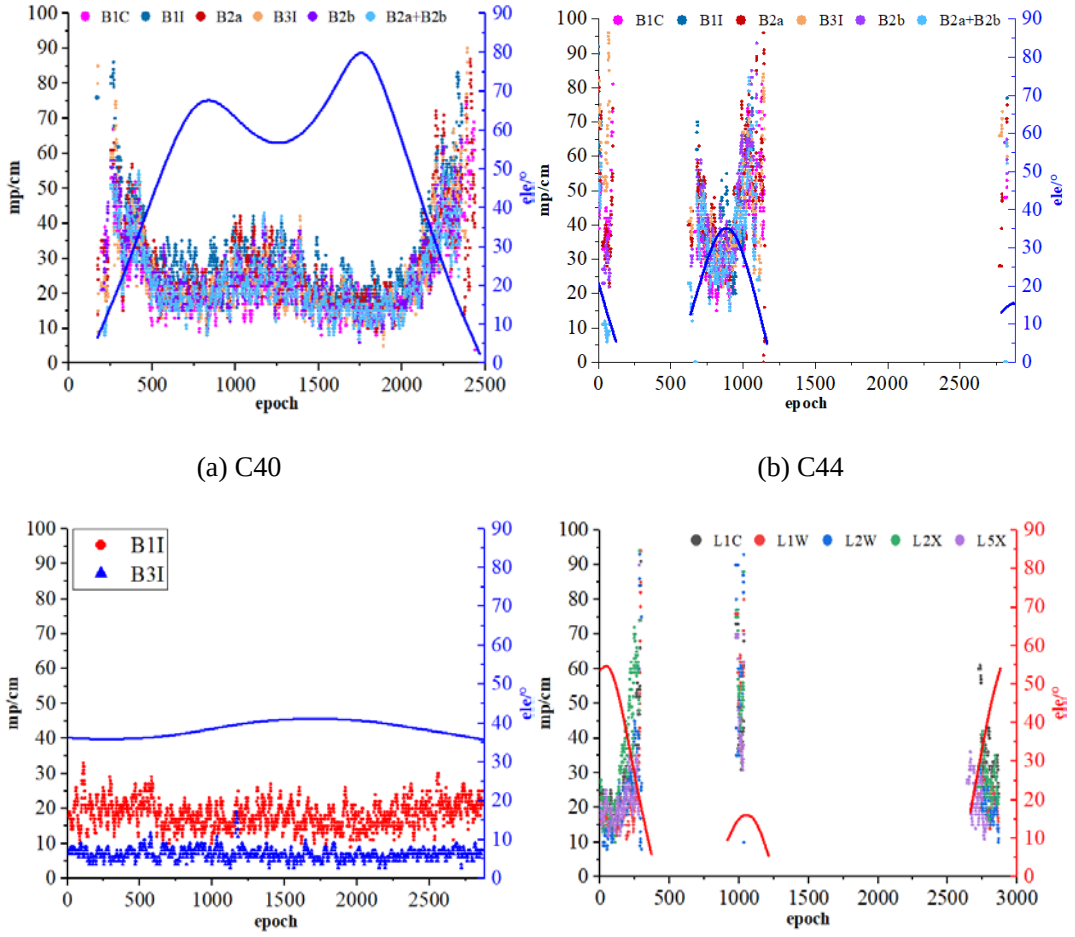


Fig. 4 C40/C44/C59/G30 pseudorange multipath error and elevation angle variation

Next, we analysed the multipath effects of two common frequency bands collected from the sea. Figure 5 shows that, after adding the choke, the multipath error of the two GEO satellites (C59 and C60) is greatly weakened and basically fluctuates around 1 cm. The multipath error of some IGSO and MEO satellites is also weakened, fluctuating around 10 cm. Therefore, the marine dynamic platform effectively reduces the impact of multipath error on BDS-3 service performance through the appropriate placement of choke coils.

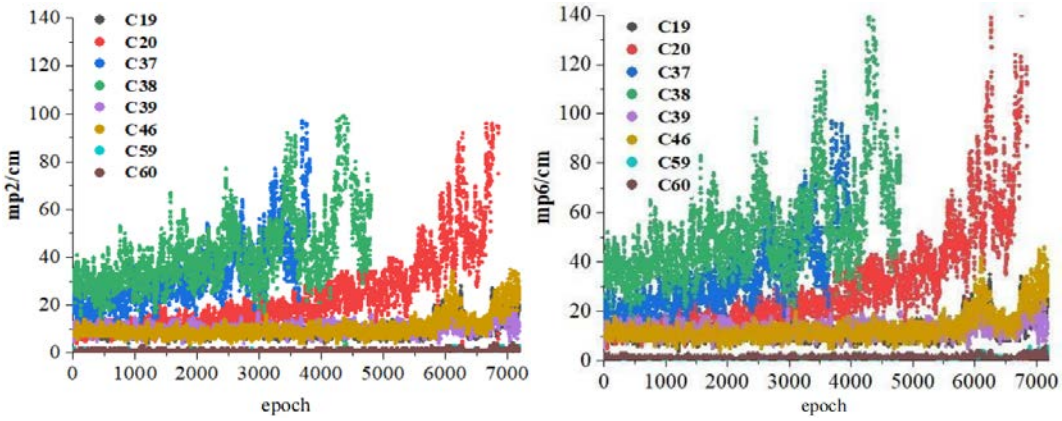


Fig. 5 BDS-3 multipath time series diagram of B1I/B3I

3.2 Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) is the ratio of carrier signal intensity to noise intensity. The SNR level is mainly affected by antenna gain parameters, the state of the correlator in the receiver, and multipath effects. It is one of the indicators reflecting the observation quality of the carrier phase, and is expressed as the ratio of the average power of the signal to the average power of the noise [16]. The SNR of each frequency band at seven stations over 10 consecutive days was averaged and compared with GPS data. The results are shown in Figures 6 and 7.

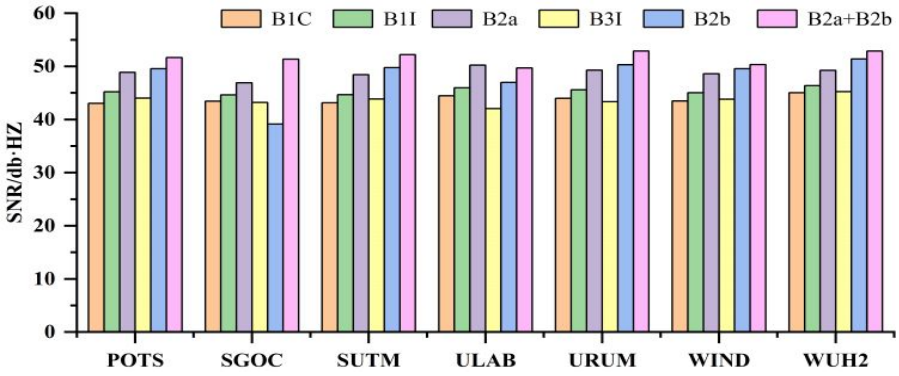


Fig. 6 BDS-3 SNR of stations

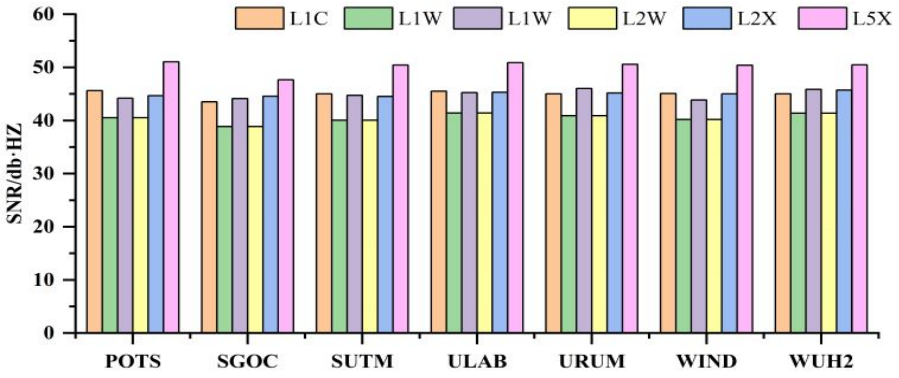


Fig. 7 GPS SNR of stations

(1) Although the SNR of each station in each frequency band is slightly different, the difference between the SNR of BDS-3 in all frequency bands is generally small. However, the SNR of the B2a+B2b frequency band is large. The minimum value of 39.14 dB·Hz in frequency band B2b occurs at station SGOC; by comparison, frequency band B2a+B2b has a maximum value of 52.88 dB·Hz at station WUH2,

a difference of 13.74 dB·Hz.

(2) The SNR of BDS-3 in each band is comparable to that of GPS at the same station. The maximum SNR of GPS is 51.04 dB·Hz in the L5 band at station POTS, which is 1.84 dB·Hz lower than the maximum value of BDS-3. The minimum SNR of GPS is 38.87 dB·Hz in the L1W and L2W bands at station SGOC, some 0.27 dB·Hz larger than the minimum value of BDS-3. The BDS-3 values are basically consistent with the average SNR of each GPS station, and the average SNRs basically fluctuate from 40-50 dB·Hz.

The relationship between SNR and satellite elevation angle is now analysed by examining the C40 (IGSO), C44 (MEO), C59 (GEO), and G30 (MEO) satellites. The results are shown in Figure 8.

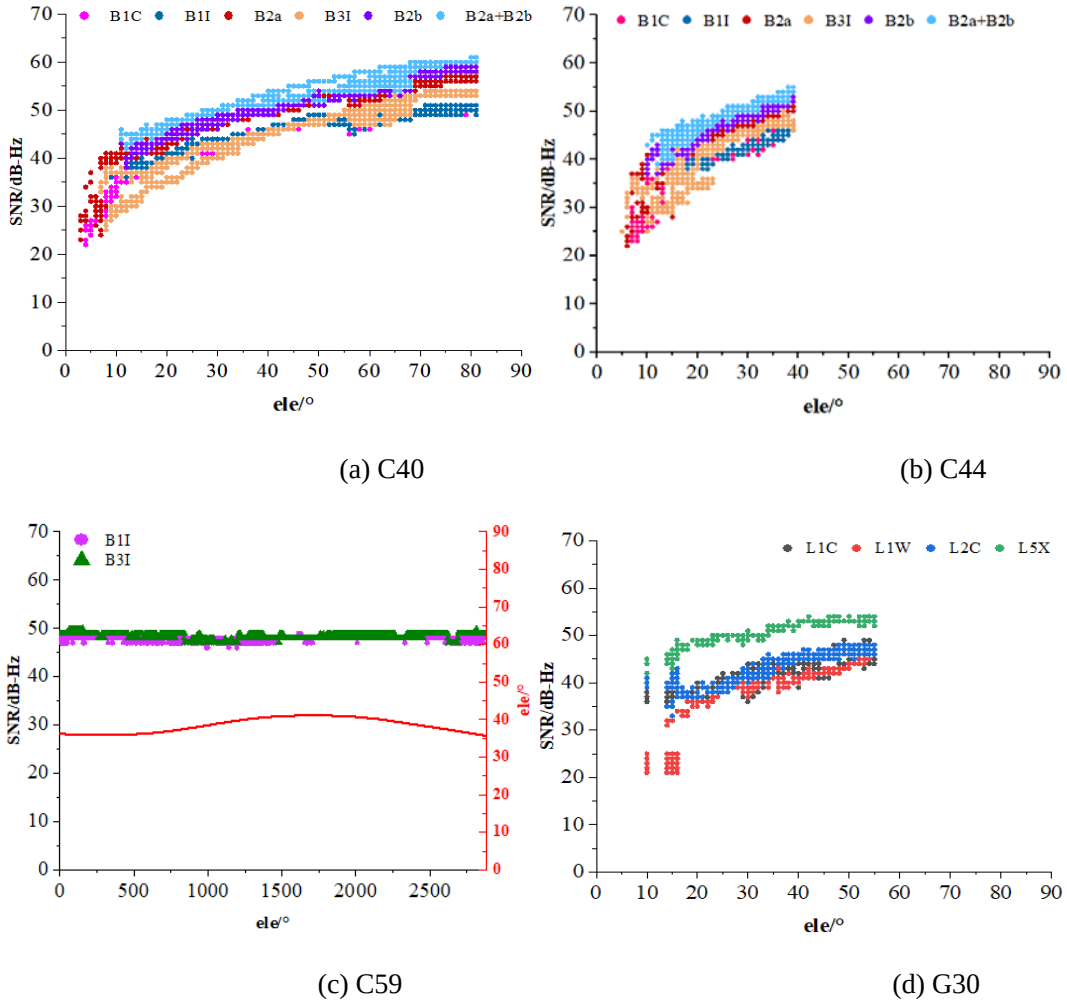


Fig. 8 C40/C44/C59/G30 SNR with respect to elevation angle

Figure 8 shows that the SNR of the IGSO and MEO satellites gradually increases with increasing elevation angle, and varies within the range 22-60 dB·Hz. Because the elevation angle of GEO satellites remains stable, the SNR of the two frequency bands does not change significantly, varying within the range 46-48 dB·Hz. For the IGSO and MEO satellites of BDS-3, the SNR of band B1I is lower than that of the other bands, and the SNR of B2b and B2a+B2b is the highest.

Analysing the SNRs of the measured marine data, Figure 9 indicates that, except for the GEO satellites, the SNR exhibits the opposite trend to that of the multipath effect and contrasts with the observations from static stations. The SNR does not change significantly, basically fluctuating in the range 30-55 dB·Hz.

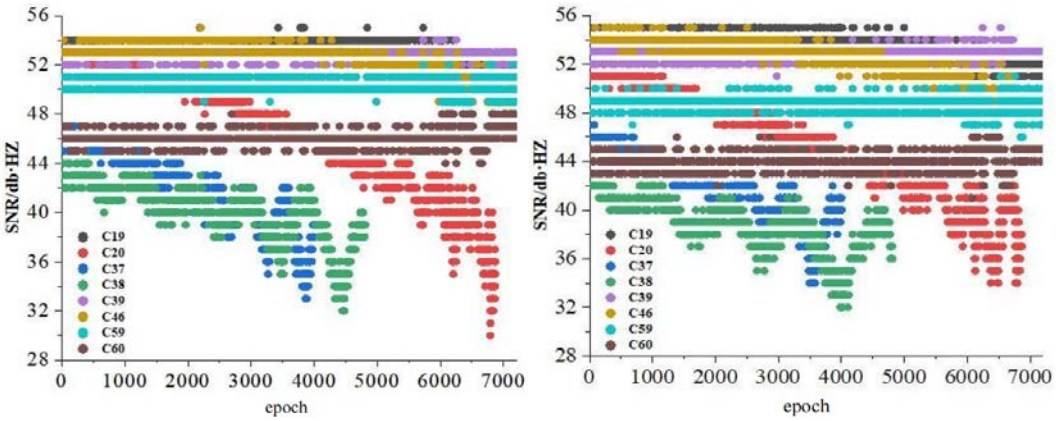


Fig. 9 BDS-3 SNR time series diagram of B1I/B3I

4. BDS-3 Data Service Performance Indicators

4.1 BDS-3 Precision Factor

The service performance of GNSS is related to the spatial geometric configuration of the satellite. In terms of measurement, the dilution of precision (DOP) is often used to describe the basic structure of the spatial geometric distribution of the satellite [17], as shown in the following expressions:

$$Q_X = (B^T P B)^{-1} = \begin{bmatrix} q_{xx} & q_{xy} & q_{xz} & q_{xt} \\ q_{yx} & q_{yy} & q_{yz} & q_{yt} \\ q_{zx} & q_{zy} & q_{zz} & q_{zt} \\ q_{tx} & q_{ty} & q_{tz} & q_{tt} \end{bmatrix} \quad (5)$$

$$GDOP = \sqrt{q_{xx} + q_{yy} + q_{zz} + q_{tt}} \quad (6)$$

$$PDOP = \sqrt{q_{xx} + q_{yy} + q_{zz}} \quad (7)$$

$$HDOP = \sqrt{q_{xx} + q_{yy}} \quad (8)$$

$$VDOP = \sqrt{q_{zz}} \quad (9)$$

The geometric DOP (GDOP) is the distance vector amplification factor between the receiver and the satellite caused by system ranging errors. If the spatial distribution of the satellite is not concentrated in one region and can be evenly distributed in different directions, the positioning accuracy of the satellite is higher than that of the uneven distribution in the same case. The position DOP (PDOP) describes the error caused by the influence of the geometric shape between the satellite and the receiver. A better geometric distribution of satellites in the sky gives a smaller PDOP value and higher positioning accuracy of the satellite system. The horizontal DOP (HDOP) is the square root of the sum of the squares of errors in latitude and longitude, which describes the positioning accuracy in the horizontal direction. The vertical DOP (VDOP) describes the positioning accuracy in the vertical direction. HDOP is consistent with VDOP in that lower values indicate higher positioning accuracy [18].

The mean DOP values of BDS-3 and GPS over 10 consecutive days at all seven stations are shown in Figure 10.

Figure 10(a) shows that, except for stations WIND and SUTM, the PDOP value of BDS-3 is slightly lower or equivalent to that of GPS. The largest difference appears at station WUH2, where the BDS-3 PDOP is 0.46 lower than that of GPS. This indicates that the distribution of BDS-3 in the sky is slightly better than that of GPS satellites. From Figure 10(b), it is apparent that the GDOP of BDS-3 is equivalent to that of GPS, except at stations ULAB and WUH2. At these stations, BDS-3 has a smaller GDOP than GPS, and the maximum difference reaches 0.78 at ULAB. The horizontal and vertical accuracy factors exhibit similar variation trends as GDOP and PDOP. BDS-3 has slightly higher values than GPS at three stations, and lower values than GPS at the other four stations. The largest differences occur at station WUH2, where the HDOP and VDOP values of BDS-3 are 0.28 and 0.34 lower than those of GPS, respectively. Thus, in general, the error

magnification of BDS-3 is smaller than that of GPS.

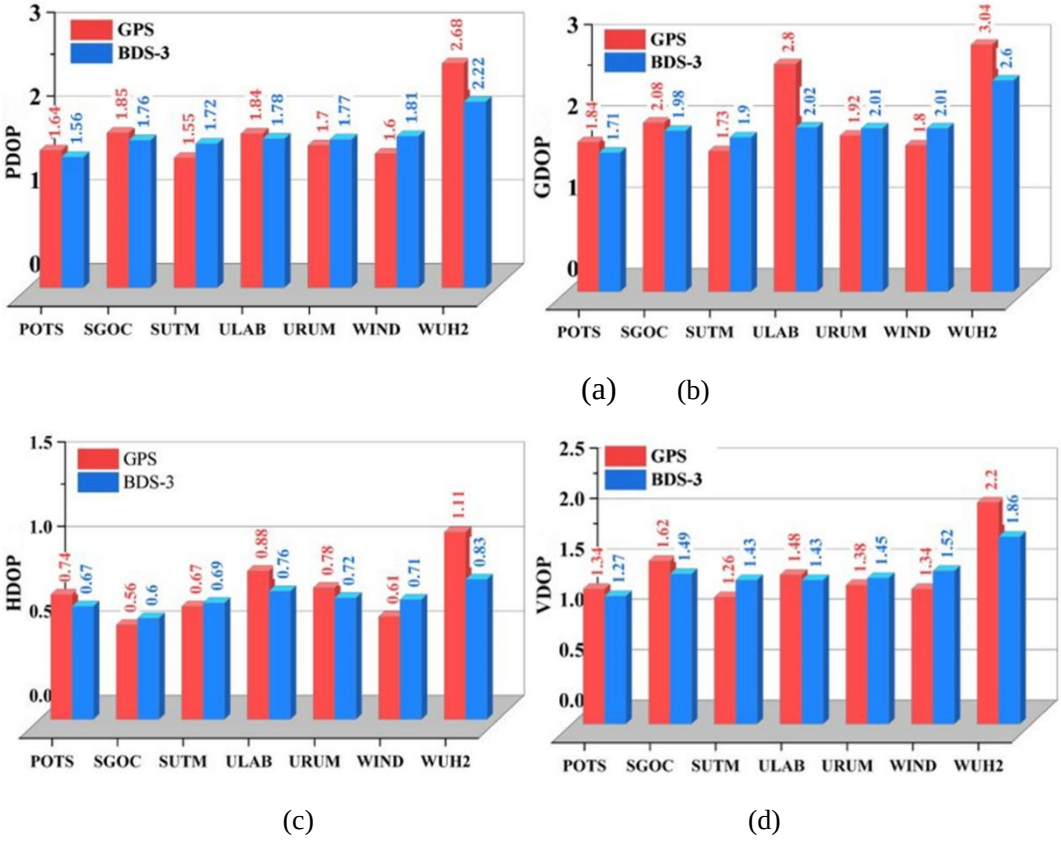


Fig. 10 BDS-3/GPS precision factor at each station

The horizontal precision factor of BDS-3 and GPS is smaller than the vertical precision factor, indicating that the horizontal accuracy is greater than the vertical accuracy. The maximum HDOP and VDOP of BDS-3 at station WUH2 are 1.11 and 2.20, respectively.

4.2 Three-Difference Residual of Carrier Phase

The precision of carrier phase observations can often reach the millimetre level. These observations can be combined with pseudorange observations to detect cycle slip. Because the observation values do not vary much over short periods of time, the precision of carrier phase observations can be expressed using the three-difference method. The adjacent epochs between each frequency band are changed three times to obtain L^3, which is used to evaluate the accuracy of the carrier phase observations. The evaluation method can be written as:

$$\begin{aligned} \Delta L_i^1 &= L_i - L_{i-1} \\ \Delta L_i^2 &= \Delta L_i^1 - \Delta L_{i-1}^1 \\ \Delta L_i^3 &= \Delta L_i^2 - \Delta L_{i-1}^2 \end{aligned} \tag{10}$$

where i denotes the observation epoch and ΔL_i^1, ΔL_i^2, ΔL_i^3 denote the first, second, and third differences. The accuracy is evaluated by calculating the mathematical expectation and variance of the third difference between the desired epochs. The mathematical expectation and variance are calculated as:

$$E(L) = \frac{1}{n}\sum_{i=1}^{n} \Delta L_i^3 \qquad \sigma_L^2 = \frac{1}{n-1}\sum_{i=1}^{n} [\Delta L_i^3 - E(L)]^2 \tag{11}$$

where n represents the number of observed epochs, $E(L)$ is the mathematical expectation, and σ_L^2 is the variance.

Taking the 2021 DOY 33 data from station WUH2 with a sampling rate of 1 s and no cycle skip after prior inspection, the carrier phase accuracy of each frequency band of BDS-3 and GPS was calculated. The results are presented in Tables 3 and 4.

Table 3 GPS carrier phase accuracy in each frequency band

system	GPS				
band	L1C	L1W	L2W	L2X	L5X
accuracy /mm	5.66	5.63	5.59	5.57	5.63

Table 4 BDS-3 carrier phase accuracy in each frequency band

system	GPS					
band	L1X	L2I	L5X	L6I	L7Z	L8X
accuracy /mm	7.89	8.14	8.13	8.19	8.3	8.29

The carrier phase accuracy of each band of BDS-3 is almost unchanged and fluctuates within a range of 0.5 mm. The largest carrier phase error of 8.3 mm appears in band B2b, and the smallest error of 7.89 mm occurs in band B1C. The carrier phase observation accuracy of BDS-3 is lower than that of GPS. The minimum error of GPS occurs in frequency band L2W, and is 2.3 mm smaller than that of BDS-3; the maximum error appears in frequency band L1C, and is 2.64 mm smaller than that of BDS-3. In the case of no cycle skip, the carrier phase positioning accuracy of BDS-3 in each frequency band exhibits little difference to that of GPS, and the difference is within 3 mm.

4.3 Performance Analysis for Relative Positioning

4.3.1 Dynamic shipborne experiment

In a real dynamic environment, the satellite signal is often out of lock or blocked. For the special constellation configuration of BDS-3, some high-latitude areas of China may suffer from frequently blocked signals from geostationary satellites because of their low elevation angle. Once the signal becomes occluded, the station will start to search for other satellites and solve the data again, which will increase the positioning error in this period. Therefore, it is necessary to conduct real-time positioning analysis using BDS-3 data through dynamic carrier experiments to study the dynamic positioning performance.

The first set of calculation examples use dynamic shipborne experimental data, collected on September 26, 2020, from 04:00:00-06:57:00 UTC. The observation environment of the whole measurement area is good, and there is basically no obstacle occlusion. As special hardware is required to receive BDS-3 satellite signals, the GAMG measurement station was selected as the reference receiver. This mobile station is located in the sea near Weihai City, China (37°35 '8.88 "N, 122°6' 41.67" E) and can receive most BDS-3 satellite signals. The distribution of the running track for this mobile station is shown in Figure 11, and the distribution of the onboard data installation instruments and equipment is shown in Figure 12: the receiver models on this dynamic station are Septentrio PolaRx5 and Trimble R9. The receiver had a sampling interval of 1 s and the baseline length was approximately 563 km. The LAMBDA method was used to fix the ambiguity.

In the actual dynamic environment, it is difficult to obtain the “true value” of the anchor point. Therefore, based on the principle of absolute distance between two antennas, the coordinate position of each antenna was calculated for every epoch, and then the distance between the two antennas was calculated as the evaluation standard[19].

Fig. 11 Trajectory diagram of shipboard

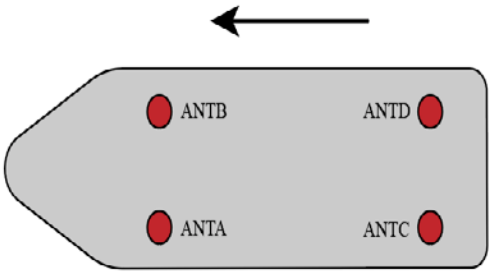


Fig. 12 Antenna position of experiment

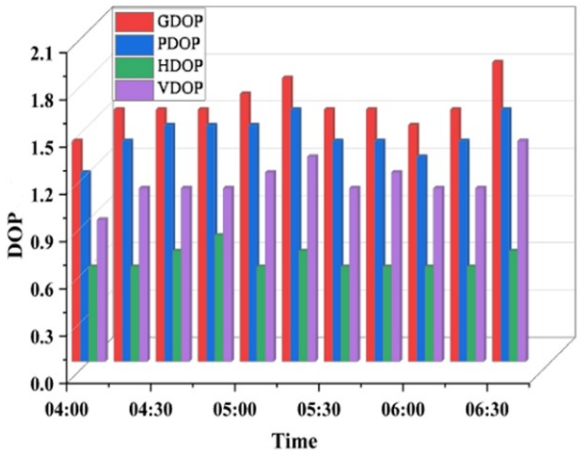


Fig. 13 Changes in satellite number and DOP

Figure 13 shows the changes in DOP values at antenna ANTA. The number of visible BDS-3 satellites ranged from 16-19. A higher number of satellites greatly improves the geometric structure of the observation satellites. From the DOP values in the figure, we find that PDOP and GDOP are higher than 1.8, HDOP is higher than 1.5, and VDOP is higher than 1.0.

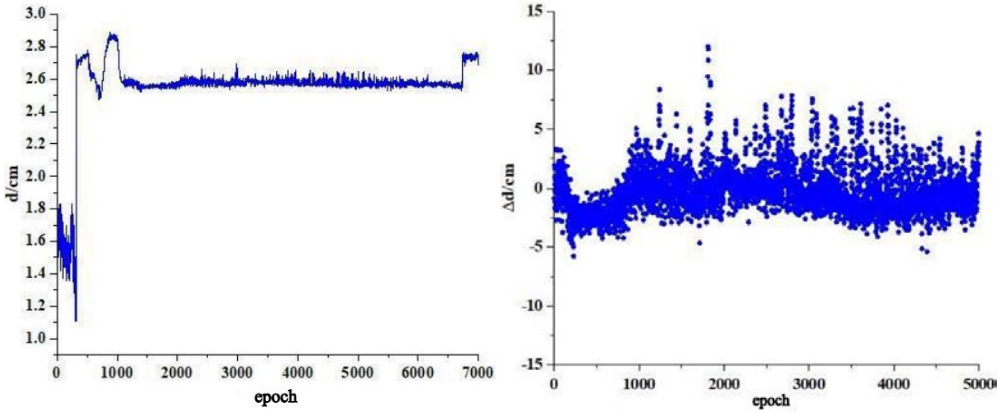


Fig. 14 ANTA-ANTD variations (left) and their differences from the true value (right)

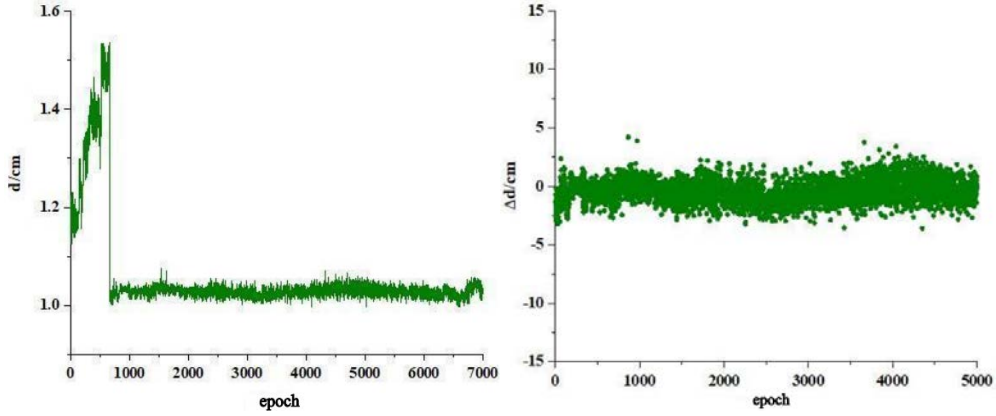


Fig. 15 ANTA-ANTB variations (left) and their difference from the true value (right)

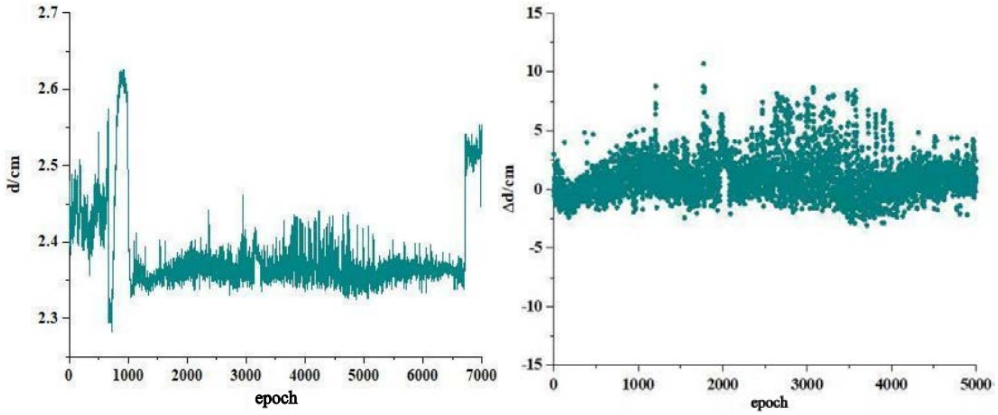


Fig. 16 ANTB-ANTD variations (left) and their difference from the true value (right)

Figures 14-16 (left) illustrate the fluctuations in the distance between antennas ANTA, ANTB, and ANTD over time. Under normal circumstances, the distance between them should be fixed. The right-hand panels in these figures represent the differences between the distance in each epoch and the true value (taken as the mean of multiple measurements). It can be seen from the diagram, three experiments of overall data calculating wave is small, only during the first half of the baseline weight appear larger deviation. Therefore, when using BDS-3 for real-time dynamic data positioning, the positioning accuracy between antennas is good after data convergence, basically fluctuating within a range of 5 cm.

Table 5 Shipboard data processing results

baseline	Min/cm	Max/cm	Mean/cm	Sdev/cm
ANTA-ANTD	-5.77	11.98	-0.28	1.89
ANTA-ANTB	-3.61	4.19	-0.41	0.94
ANTB-ANTD	-3.06	10.69	1.06	1.75

Table 5 presents statistics from processing the data of 5000 epochs after convergence. The largest deviation occurs between ANTA and ANTD, and the largest difference after convergence is 11.98 cm. The highest average value is for the distance between ANTB and ANTD. The mean distance between ANTA and ANTD and between ANTA and ANTB is less than 0.5 cm. In each case, the STD between any two antennas is less than 2 cm. Overall, the real-time dynamic positioning accuracy of BDS has reached the centimetre level, which is basically equivalent to the simulation experiment accuracy of static stations.

4.3.2 BDS-3 static experiment

To evaluate the difference in dynamic relative positioning performance between BDS-3 and GPS, a set of static data was dynamically processed using self-developed software. The true values for the experiment were taken from coordinates provided by ITRF. The specific parameters of the station are listed in Table 6.

Figure 17 shows that the real-time dynamic relative positioning of BDS-3 requires some time to locate the ambiguity. The positioning time is about 25 min. During this period, the data fluctuate significantly, while the ambiguity is fixed. The data fluctuations during the fixed ambiguity time are relatively small, and are related to the number of participating satellites. The dual-frequency GPS observations have an RMS of 6 mm, whereas those of BDS-3 have an RMS of 4–5 mm. This is related to the lack of BDS-3 satellite data observed during this period.

Table 6 Information of LEIJ-HUEG data

declaration	parameter
Antenna	LEIAR25.R4 LEIT
Receiver	JAVAD TRE_3
Time	2021-9-25 14:00:00-14:59:59
Sampling	1s
Baseline distance	520km
Elevation mask	10°
BDS-3 satellite	C24/C25/C26/C27/C33/C37/C38/C41/C42/C43/C46/C60

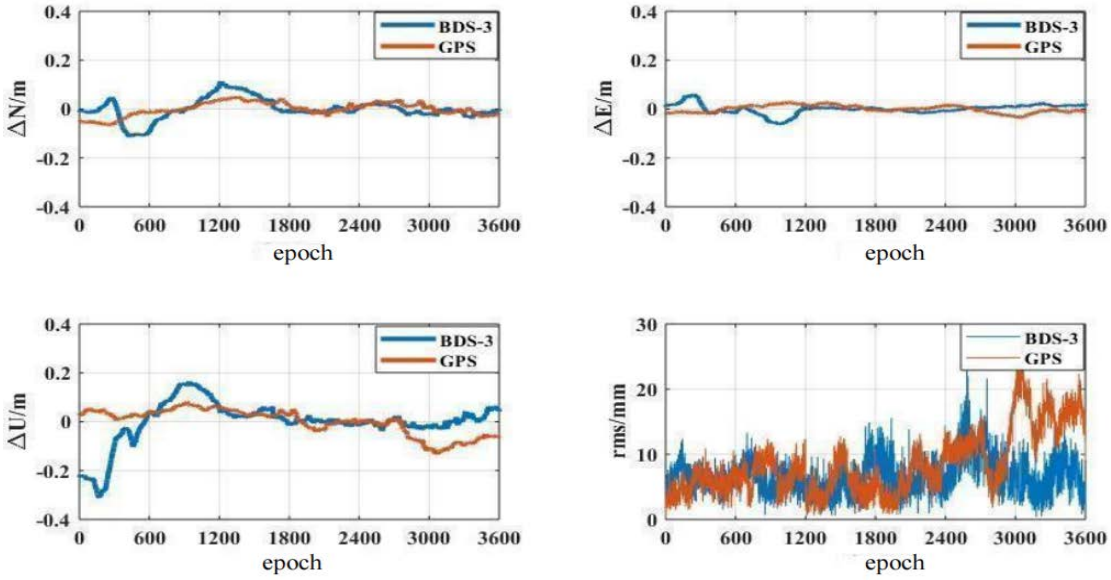


Fig. 17 Time series for the bias variation of position based on BDS-3/GPS

Table 7 Statistics of GPS/BDS-3 data processing results

System	Direction	Min/cm	Max/cm	Mean/cm	STD/cm
GPS	ΔN	-0.32	4.91	0.93	1.01
	ΔE	-0.36	2.58	-0.17	1.41
	ΔU	-13.21	6.02	-2.17	5.17
BDS-3	ΔN	-3.91	11.29	0.86	0.72
	ΔE	-1.93	2.47	0.33	2.24
	ΔU	-0.38	6.36	0.77	3.23

Table 7 summarizes the data of 2000 epochs after convergence. The largest deviation occurs in the vertical direction of GPS. In addition to the large deviation values, the real-time dynamic positioning accuracy of BDS-3 has reached the centimetre level. Once the ambiguity has been fixed, the data converge around 0, with horizontal fluctuations of 2-3 cm and vertical fluctuations of ~10 cm. The accuracy of real-time dynamic relative positioning after data convergence is basically equivalent to that of GPS.

5. Conclusions

(1) This study has analysed the complete BDS-3 data quality. The overall multipath error of BDS-3 is larger than that of GPS by around 5-10 cm. In terms of SNR, BDS-3 is basically consistent with each GPS station, and the average SNR basically fluctuates from 40-50 dB·Hz.

(2) In terms of service performance, in the case of no cycle skips, the carrier phase positioning accuracy of BDS-3 in each frequency band is similar to that of GPS, with the difference being within 3 mm. The four indices of DOP at most stations are lower than those of GPS. The maximum difference appears at station ULAB, where the GDOP of BDS-3 is 0.78 m lower than that of GPS. The double difference pseudorange residuals of each frequency band from the zero baseline range from 0.2-0.3 m, and the carrier phase residuals range from 0.17-0.48 cm. After data convergence, the relative positioning error fluctuates within 5 cm of the true value, although the fluctuations in the vertical direction are larger (within 10 cm).

(3) The quality indicators of existing BDS-3 signals were also evaluated. In terms of the pseudorange multipath error, the average value of the observation data at each station over 10 consecutive days was found to run in the order B1I>B2a>B2b>B3I>B2a+B2b>B1C. Frequency bands B2a+B2b and B2a have large SNRs, while the other bands have SNRs of 40-45 dB·Hz. The carrier phase accuracy of BDS-3 varies little in each frequency band, with fluctuations of only 0.5 mm.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (42174021 and 42074028) and the Shandong Provincial Natural Science Foundation, China (ZR2021MD060).

References

[1] CSNO (China Satellite Navigation Ofce) (2018a) Development of the BeiDou navigation satellite system (version 3.0)

[2] CSNO (China Satellite Navigation Ofce) (2018b) BeiDou navigation satellite system open service performance standard (version 2.0)

[3] CSNO (China Satellite Navigation Ofce) (2019) BeiDou navigation satellite system signal in space interface control document open service signal B1I (version 3.0)

[4] Lu M, Li W, Yao Z, et al. (2019) Overview of BDS III new signals. Navigation - Journal of The Institute of Navigation, 66(4):19-35.

[5] Liu L, Zhang T, Zhou S, et al. (2019) Improved design of control segment in BDS-3. Navigation, 66(1):37-47.

[6] Yang Y, Gao W, Guo S, et al. (2019) Introduction to BeiDou-3 navigation satellite system. Navigation, 66(01): 7-18.

[7] Betz J W, Lu M, Morton Y, et al. (2019) Introduction to the special issue on the BeiDou navigation system. Annual of Navigation. 66 (1):3-5.

[8] CSNO (China Satellite Navigation Ofce) (2019) BeiDou navigation satellite system signal in space interface control document open service signal B2b (Beta version)

[9] Cai C, He C, Santerre R, et al. (2016) A Comparative Analysis of Measurement Noise and Multipath for Four Constellations: GPS, BeiDou, GLONASS and Galileo. Survey Review, 48(349): 287-295.

[10] Yang Y, Xu Y, Li J, et al. (2018) Progress and Performance Evaluation of BeiDou Global Navigation Satellite System: Data Analysis Based on BDS-3 Demonstration System. Science China Earth Sciences, 61(5): 614-624.

[11] Dai P, Ge Y, Qin W, et al. (2019) BDS-3 Time Group Delay and Its Effect on Standard Point Positioning. Remote Sensing, 11(15):1819-1848.

[12] Mu R, Dang Y, Xu C. (2020) BDS-3/GNSS Data Quality and Positioning Performance Analysis[C]. China Satellite Navigation Conference (CSNC) 2020 Proceedings: Volume I. Springer Singapore, Singapore, 368-379.

[13] Zhang Z, Li B, Nie L, et al. (2019) Initial assessment of BeiDou-3 global navigation satellite system: signal quality, RTK and PPP. GPS Solutions, 23(4):111-123.

[14] Zhu Y, Zheng K, Cui X, et al. (2021) Preliminary Analysis of the Quality and Positioning Performance of BDS-3 Global Interoperable Signal B1C&B2a. Advances in Space Research, 67(8): 2483-2490.
[15] Wu X, Zhou J, Wang G, et al. (2012) Multipath error detection and correction for GEO/IGSO satellites. Science China Physics, Mechanics and Astronomy, 55(7): 1297-1306.
[16] Bu J, Zuo X, Li X, Chang J, Zhang X. (2019) Evaluation and analysis on positioning performance of BDS/QZSS satellite navigation systems in Asian-Pacific region. Advances in Space Research, 63(7): 2189-2211.
[17] Teng Y, Wang J. (2016) A closed-form formula to calculate geometric dilution of precision (GDOP) for multi-GNSS constellations. GPS Solutions, 20(3): 331-339.
[18] Jian Z, Jaa C, Zwa C, et al. (2022) Signal quality and positioning performance of GPS/BDS-3/GLONASS/Galileo in polar regions. Advances in Space Research, 69(6): 2541-2554.
[19] Teunissen P. (1995) The least-square ambiguity decorrelation adjustment: a method for fast GPS ambiguity estimation. Journal of Geodesy, 70:65-82.
[20] He K (2015) GNSS kinematic position and velocity determination for airborne gravimetry. Technische Universität Berlin

Authors

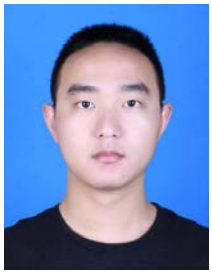
Zhipeng Ding received his Master degree in Geodesy and Survey Engineering from China University of Petroleum (East China) in 2022. His research is focused on BDS-3 precise kinematic relative positioning.

Kaifei He is a professor at the China University of Petroleum (East China). He received his Ph.D. degree in Geodesy from the Technical University of Berlin (TU Berlin) in 2015 and studied in GFZ as a Ph.D. student from 2010 to 2014. Currently, his research focuses on GNSS navigation and positioning algorithms, underwater positioning, geodesy data processing, and software development.

Ming Li is a graduate student of China University of Petroleum (East China). His research focuses on BDS-3 high-precision real-time dynamic positioning.

Yu Wu is a graduate student of China University of Petroleum (East China). His research focuses on underwater navigation and positioning

Yue Zhang is a graduate student of China University of Petroleum (East China). His research focuses on retrieval of sea surface wind field from quad-polarization SAR.



Jinquan Yang is a graduate student of China University of Petroleum (East China). His research focuses on the theory and method of GNSS/INS integrated navigation.

Journal of Global Positioning Systems (*2022*)
Vol. 18, No. 1: 56-69
DOI: 10.5081/jgps.18.1.56

Journal of Global Positioning Systems
2022
© CPGPS

Innovative Formulation in Discrete Kalman Filtering with Constraints - A Generic Framework for Comprehensive Error Analysis

Jianguo Wang, Benjamin Brunson, Baoxin Hu

Department of Earth and Space Science and Engineering
Lassonde School of Engineering, York University, Toronto, Canada
Emails: jgwang@yorku.ca, brunson@yorku.ca and baoxin@yorku.ca

Corresponding author: Benjamin Brunson

Abstract: This manuscript establishes a generic framework for comprehensive error analysis in discrete Kalman filtering with constraints, which systematically provides a complete set of algorithmic formulas along with demonstrating an alternative process of theoretical analytics of discrete Kalman filter. This constructive work aims extensively to standardize the formulation of Kalman filter with constraints. In analogy to the similar framework for standard discrete Kalman filter (without any constraints), the proposed framework specifically considers: model formulation vs. the error sources, the solution of the state and process noise vectors, the residuals for the process noise vector and the measurement noise vector, the redundancy contribution of the predicted state vector, process noise vector and measurement vector, and other relevant essential aspects, of which some of the features are essential to comprehensive error analysis, but are nonexistent yet in the primary algorithm in Kalman filtering with constraints. Besides, the algorithmic form of the Extended Kalman filter with constraints is also provided for practical purpose. At the end, specific remarks about the developed framework are given to emphasize on its usage to a certain extend.

KEY WORDS: Kalman filter, state constraint, error analysis, generic framework, redundancy contribution.

1. INTRODUCTION

The Kalman filter is a recursive estimator that provides estimates of a group of selected states on the ground of a specific system model and measurements that are acquired over time. Its applications have steadily expanded in sciences and engineering since the 1960s.

Usually, the Kalman filter consists of a system model associated with its modeling errors as process noises and a measurement model associated with measurement noise. However, there are also many circumstances under which *a priori* knowledge of a dynamic system leads to equality constraints that may be imposed on the system states in Kalman filtering. Examples of this include path-constrained motion along roadways [Yang et al, 2005; Hasberg et al 2012] and constant velocity motion of tracking targets [Alouani and Blair, 1991]. In multisensor integrated navigation, the states representing the attitude commonly involve specific constraints, e.g., the elements of the direction cosine matrix have to conform to orthonormality conditions and the elements in a quaternion vector or rotation vector have to be in unit norm. Apparently, the formulation of indirect observation (Least Squares) adjustment with constraints in Geodesy and Geomatics has been generally standardized [Mikhail, 1970; Rao and Toutenburg, 1999; Wang, et al, 2019]. By contrast, the formulation on states-constrained Kalman filter is far from being standardized to the same degree.

Constrained Kalman filtering by augmentation was first proven by Doran [1992], which has been considered as a seminal paper on the subject [Pizzinga, 2012]. There exist several dominant strategies to impose constraints on the system states in Kalman filtering, which are generally divided into three categories [Simon, 2010; Khabbazi and Esfanjani, 2014]:

Reparameterization: this technique incorporates any system state constraints by reducing the parameterization of the system, through which the physical meaning of the system states may be lost [Simon, 2010].

"Perfect" Observations: this technique treats the system state constraints as pseudo-observations with zero variance. Without further simplification, it may

cause numerical instability [Doran, 1992; Alouani and Blair, 1991].

Projection: this technique transforms the estimate of the system states onto a constraint surface [Khabbazi and Esfanjani, 2014]. Such transformation may be accomplished through projection of the system state estimate [Simon and Chia, 2002], projection of the system itself [Ko and Bitmead, 2007], or projection of the Kalman gain matrix [Teixeira et al, 2008]. State projection is the most commonly used method of imposing constraints on the system states in Kalman filtering [Khabbazi and Esfanjani, 2014]. The Kalman gain projection has been generalized for non-linear constraints [Xu et al, 2017]. These techniques may also apply their constraints less strictly by taking a weighted average between the constrained and the unconstrained solution [Baker and Thennadil, 2019], or by taking model uncertainty into account in the gain projection approach [Khabbazi and Esfanjani, 2015].

Besides, some other techniques have also been used to impose equality constraints in Kalman filtering that do not fit under the above mentioned three broad categories. Xu et al [2013] considered constraints *a priori* information that should also be incorporated into a system's dynamic models. Ghanbarpourasl and Zobar [2022] utilized singular value decomposition to separate the system state into a deterministic (i.e. fully constrained) and a stochastic component. Pizzinga [2012] framed the constrained Kalman filter as a recursive least-squares problem.

Unfortunately, there is still a lack of generic algorithmic formulas directly for the standard form of the discrete Kalman filter with constraints in literature for conducting comprehensive error analysis. This motivates the authors to develop a complete set of the generic formulas for it, so that one can easily adapt to theoretical development and practical implementation.

Following this introduction, this manuscript first summarizes the innovative alternate formulation of standard Kalman filter originally deduced by Wang [1997] and also specifically detailed and applied in [Caspary and Wang, 1998; Wang, 1997; Wang, 2008, 2009; Wang et al, 2009; Wang et al, 2009; Gopaul et al, 2010; Wang et al, 2010; Qian, 2017; Qian, et al, 2015, 2016; Wang et al, 2015, 2021; Zhang et al, 2017]. Then, as the core of this manuscript, Section 3 systematically develops the theoretical aspects and practical algorithm in discrete Kalman filtering with constraints, which innovatively promote the comprehensive error analysis. Section 4 further delivers the proposed algorithm in the form of Extended Kalman filter with constraints. The manuscript ends with concluding remarks in Section 5.

2. ALGORITHMIC FORMULATIONS OF STANDARD KALMAN FILTER

In general, a Kalman filter estimates the state vector by minimizing its mean squared errors after the minimum variance principle or equivalently its weighted sum of the residuals squared after the Principle of Least Squares, on the basis of operating system and measurement models recursively.

2.1 Standard form of Discrete Kalman filter

Let us define the standard form of Kalman filter first. Consider a linear or linearized system described in state space and the data are made available over a discrete time series $t_0, t_1, ..., t_k, ..., t_N$, of which each time instant corresponds to an observation epoch and is simply depicted as $0, 1, ..., k, ..., N$. Without loss of generality, the formulation here omits the deterministic system input.

At an arbitrary observation epoch k ($1 \le k \le N$), the system and measurement models are given as follows [Wang et al, 2021]:

$$\boldsymbol{x}(k) = \boldsymbol{A}(k,k-1)\boldsymbol{x}(k-1) + \boldsymbol{B}(k,k-1)\boldsymbol{w}(k) \quad (2.1)$$

(or simply $\boldsymbol{x}(k) = \boldsymbol{A}(k)\boldsymbol{x}(k-1) + \boldsymbol{B}(k)\boldsymbol{w}(k)$ (2.1a))

$$\boldsymbol{z}(k) = \boldsymbol{C}(k)\boldsymbol{x}(k) + \boldsymbol{\Delta}(k) \quad (2.2)$$

wherein $\boldsymbol{x}(k)$, $\boldsymbol{z}(k)$, $\boldsymbol{w}(k)$, and $\Delta(k)$ are the n-dimensional state-vector, the p-dimensional observation vector, the m-dimensional process noise vector, and the p-dimensional measurement noise vector, respectively, while $\boldsymbol{A}(k,k-1)$, $\boldsymbol{B}(k,k-1)$, and $\boldsymbol{C}(k)$ are the $n \times n$ coefficient matrix of $\boldsymbol{x}(k)$, the $n \times m$ coefficient matrix of $\boldsymbol{w}(k)$, and the $p \times n$ coefficient matrix of $\boldsymbol{z}(k)$, respectively. About the relevant stochastic information, $\boldsymbol{w}(k) \sim N(\boldsymbol{o}, \boldsymbol{Q}(k))$ and $\boldsymbol{\Delta}(k) \sim N(\boldsymbol{o}, \boldsymbol{R}(k))$ are assumed, where $N(a,b)$ represents a normal distribution with a and b as its expectation (vector) and variance (matrix). Between two different observation epochs, it is presumed to have $Cov(\boldsymbol{w}(i), \boldsymbol{w}(j)) = \boldsymbol{O}$ and $Cov(\boldsymbol{\Delta}(i), \boldsymbol{\Delta}(j)) = \boldsymbol{O}$ for ($i \ne j$), and $Cov(\boldsymbol{w}(i), \boldsymbol{\Delta}(j)) = \boldsymbol{O}$ for any i and j.

Besides, the initial state vector is given as $\boldsymbol{x}(0)$ with its variance matrix $\boldsymbol{D}_{xx}(0)$ and is independent of $\boldsymbol{w}(k)$ and $\Delta(k)$ for any k, i.e., $Cov(\boldsymbol{w}(k), \boldsymbol{x}(0)) = \boldsymbol{O}$ and $Cov(\boldsymbol{\Delta}(k), \boldsymbol{x}(0)) = \boldsymbol{O}$.

2.2 The Solution after Minimum Variance Principle

Without any unnecessary repetition of the solution derivation, the Kalman filtering algorithm at k from k-1 upon the definition in Section 2.1 after

the minimum variance principle is directly summarized below:

$$\hat{\boldsymbol{x}}(k) = \hat{\boldsymbol{x}}(k/k-1) + \boldsymbol{G}(k)\boldsymbol{d}(k) \tag{2.3}$$

with its associated variance matrix

$$\begin{aligned}\boldsymbol{D}_{xx}(k) = &[I - \boldsymbol{G}(k)\boldsymbol{C}(k)]\boldsymbol{D}_{xx}(k/k-1) \\ &\cdot[I - \boldsymbol{G}(k)\boldsymbol{C}(k)]^T + \boldsymbol{G}(k)\boldsymbol{R}(k)\boldsymbol{G}^T(k)\end{aligned} \tag{2.4}$$

wherein $\boldsymbol{I}$ is a *n*x*n* identity matrix and $\boldsymbol{G}(k)$ is a *n*x*p* Kalman gain matrix:

$$\boldsymbol{G}(k) = \boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k) \tag{2.5}$$

The predicted state vector (from the time update) and its variance matrix are as follows:

$$\hat{\boldsymbol{x}}(k/k-1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}(k-1/k-1) \tag{2.6}$$

$$\begin{aligned}\boldsymbol{D}_{xx}(k/k-1) = &\boldsymbol{A}(k)\boldsymbol{D}_{xx}(k-1/k-1)\boldsymbol{A}^T(k) \\ &+ \boldsymbol{B}(k)\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\end{aligned} \tag{2.7}$$

The system innovation vector and its variance matrix are computed after:

$$\boldsymbol{d}(k) = \boldsymbol{z}(k) - \boldsymbol{C}(k)\hat{\boldsymbol{x}}(k/k-1) \tag{2.8}$$

$$\boldsymbol{D}_{dd}(k) = \boldsymbol{C}(k)\boldsymbol{D}_{xx}(k-1/k-1)\boldsymbol{C}^T(k) + \boldsymbol{R}(k) \tag{2.9}$$

Essentially, the system innovation vectors: $\boldsymbol{d}(1), \boldsymbol{d}(2), ..., \boldsymbol{d}(k), ...$ are independent of each other [Chui & Chen, 1987], i.e., $Cov(\boldsymbol{d}(i), \boldsymbol{d}(j)) = \boldsymbol{O}$ ($i \neq j$). However, the elements in $\boldsymbol{d}(k)$ at epoch *k* are not only correlated, but also blend all of the separate error sources. Traditionally, the error analysis has been centered on the system innovation series. In addition, it is proved that the estimate of the state vector $\boldsymbol{x}(k)$ and the system innovation vector $\boldsymbol{d}(k)$ are independent of each other based on (2.3) and (2.8), i.e.,

$$\boldsymbol{D}_{xd}(k) = O \tag{2.10}$$

2.3 Alternate Formulation for Comprehensive Error Analysis

Obviously, $\boldsymbol{d}(k)$ is originated from the process noise series $\boldsymbol{w}(1), ..., \boldsymbol{w}(k), ...$, the measurement noise series $\Delta(1), ..., \Delta(k), ...$ along with the initial state vector $\boldsymbol{x}(0)$. Therefore, as a matter of fact, the system and measurement models in (2.1) and (2.2) are associated with three groups of independent stochastic information that is propagated into the state solution from time to time. Specifically at *k*, the system is contaminated by (i) the measurement noise vector $\Delta(k)$, (ii) the process noise vector $\boldsymbol{w}(k)$, and (iii) the noise associated with the predicted state vector from $\boldsymbol{A}(k, k-1)\boldsymbol{x}(k-1)$, into which $\Delta(1), ..., \Delta(k-1)$ and $\boldsymbol{w}(1), ..., \boldsymbol{w}(k-1)$ starting with $\boldsymbol{x}(0)$ are propagated through the recursive mechanism as in (2.1) and (2.2) from the past.

Along two different paths, either after the Minimum Variance Principle or Least Squares Principle, the Kalman filtering algorithm is equivalently derived. A widely repeated derivation is to deliver the equivalent solution on the ground of the predicted state vector $\boldsymbol{x}(k/k-1)$, as a pseudo-measurement vector by merging (ii) and (iii) as in (2.1) in Least Squares approach, and the measurement vector $\boldsymbol{z}(k)$ from (i). An apparent drawback to this formulation is that two groups of the independent stochastic information in (ii) and (iii) are blended into $\boldsymbol{x}(k/k-1)$ and are no more separable in error analysis.

To enhance the error analysis in discrete Kalman filtering, Wang [1997] proposed an innovative alternate formulation. Innovatively, the system state prediction in (2.1) was further split into two pseudo-measurement vectors:

$$\boldsymbol{l}_x(k) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}(k-1) = \hat{\boldsymbol{x}}(k/k-1) \quad \boldsymbol{D}_{l_x l_x}(k) \tag{2.11}$$

$$\boldsymbol{l}_w(k) = \boldsymbol{w}_0(k) \qquad \boldsymbol{Q}(k) \tag{2.12}$$

with $\boldsymbol{w}_0(k) = \boldsymbol{o}$ (zero mean presumed) and

$$\boldsymbol{D}_{l_x l_x}(k) = \boldsymbol{A}(k)\boldsymbol{D}_{xx}(k-1)\boldsymbol{A}^T(k) \tag{2.13}$$

The real measurement vector $\boldsymbol{z}(k)$ remains as in (2.2) and denoted by $\boldsymbol{l}_z(k) = \boldsymbol{z}(k)$.

The residual equations corresponding to (2.11), (2.12) and (2.2) are as follows:

$$\boldsymbol{v}_{l_x}(k) = \hat{\boldsymbol{x}}(k) - \boldsymbol{B}(k)\hat{\boldsymbol{w}}(k) - \boldsymbol{l}_x(k) \tag{2.14}$$

$$\boldsymbol{v}_{l_w}(k) = \hat{\boldsymbol{w}}(k) - \boldsymbol{l}_w(k) \tag{2.15}$$

$$\boldsymbol{v}_{l_z}(k) = \boldsymbol{C}(k)\hat{\boldsymbol{x}}(k) - \boldsymbol{l}_z(k) \tag{2.16}$$

with $\boldsymbol{D}_{l_x l_x}(k)$, $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ as their measurement variance matrices, respectively, in which the state vector is extended to include the process noise vector $\boldsymbol{w}(k)$ being estimated together with $\boldsymbol{x}(k)$.

In seeking for a Least Squares solution for $\boldsymbol{x}(k)$ and $\boldsymbol{w}(k)$, the cost function is constructed

$$\begin{aligned}\boldsymbol{min}: g(k) = &\boldsymbol{v}_{l_x}^T(k)\boldsymbol{D}_{l_x l_x}^{-1}(k)\boldsymbol{v}_{l_x}(k) + \\ &\boldsymbol{v}_{l_w}^T(k)\boldsymbol{Q}^{-1}(k)\boldsymbol{v}_{l_w}(k) + \boldsymbol{v}_z^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_z(k)\end{aligned} \tag{2.17}$$

In (2.14), (2.15) and (2.16), there are (*n* + *m*) states and (*n* + *m* +*p*) measurements. The number of the redundant measurements remains unchanged, namely, *p*. It is not in question about the identity of $\boldsymbol{x}(k)$ derived after (2.17) and the one in (2.3) [Wang, 1997]. The beauty of this formulation lies in the feasibility for the direct analysis of the three original error sources at any epoch *k*. Especially, it allows for reliability analysis in discrete Kalman filtering [Wang, 1997, 2009]. For the benefit of the

presentation in next section, the outcome from this alternate formulation is summarized below:

1) The solution of the state vector

First, equations (2.3) – (2.9) in Section 2.2 remain unchanged to form the basis of the solution. Alternatively, (2.3) is also given as follows

$$\boldsymbol{x}(k/k) = \boldsymbol{l}_x(k) + \boldsymbol{B}(k)\boldsymbol{l}_w(k) + \boldsymbol{K}(k)\{\boldsymbol{l}_z(k) - \boldsymbol{C}(k)[\boldsymbol{l}_x(k) - \boldsymbol{B}(k)\boldsymbol{l}_w(k)]\} \quad (2.18)$$

Importantly, the process noise vector is estimated by

$$\hat{\boldsymbol{w}}(k) = \boldsymbol{w}_0(k) + \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k) \quad (2.19)$$

with its variance matrix

$$\boldsymbol{D}_{ww}(k) = \boldsymbol{Q}(k) - \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\boldsymbol{Q}(k) \quad (2.20)$$

and its covariance matrix with the estimated state vector

$$\boldsymbol{D}_{xw}(k) = \boldsymbol{B}(k)\boldsymbol{Q}(k) - \boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\boldsymbol{Q}(k) \quad (2.21)$$

2) The residual vectors

$$\boldsymbol{v}_{l_x}(k) = \boldsymbol{D}_{l_x l_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{G}(k)\boldsymbol{d}(k) \quad (2.22)$$

$$\boldsymbol{v}_w(k) = \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{G}(k)\boldsymbol{d}(k) \quad (2.23)$$

$$\boldsymbol{v}_z(k) = [\boldsymbol{C}(k)\boldsymbol{G}(k) - \boldsymbol{I}]\boldsymbol{d}(k) \quad (2.24)$$

with ***their variance matrices***

$$\boldsymbol{D}_{v_{l_x}v_{l_x}}(k) = \boldsymbol{D}_{l_x l_x}(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k-1)\boldsymbol{C}(k)\boldsymbol{D}_{l_x l_x}(k) \quad (2.25)$$

$$\boldsymbol{D}_{v_w v_w}(k) = \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\boldsymbol{Q}(k) \quad (2.26)$$

$$\boldsymbol{D}_{v_z v_z}(k) = [\boldsymbol{I} - \boldsymbol{C}(k)\boldsymbol{G}(k)]\boldsymbol{R}(k) \quad (2.27)$$

3) ***The redundancy contributions*** in measurement groups corresponding to (2.11), (2.12) and 2.2):

$$r_{l_x}(k) = tr[\boldsymbol{A}(k)\boldsymbol{D}_{xx}(k-1)\boldsymbol{A}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)] \quad (2.28)$$

$$r_{l_w}(k) = tr[\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k-1)\boldsymbol{C}(k)\boldsymbol{B}(k)] \quad (2.29)$$

$$r_z(k) = tr[\boldsymbol{I} - \boldsymbol{C}(k)\boldsymbol{G}(k)] \quad (2.30)$$

For the entire system either after (2.1) and (2.2), or after (2.11), (2.12) and (2.2), the total redundancy number at epoch *k* satisfies [Wang, 1997; 2009, 2021; etc]

$$r(k) = r_{l_x}(k) + r_{l_w}(k) + r_z(k) = p(k) \quad (2.31)$$

wherein $p(k)$ is the number of the real measurements or the dimension of $\boldsymbol{z}(k)$.

4) The individual redundancy indexes

In practice, $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ are commonly diagonal so that the individual redundancy indexes in components for the process noise vector are

$$r_{w_i}(k) = [\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)]_{ii} \quad (i = 1, 2, ..., m(k)) \quad (2.32)$$

and for the measurement vector

$$r_{z_i}(k) = [\boldsymbol{I} - \boldsymbol{C}(k)\boldsymbol{G}(k)]_{ii} \quad (i = 1,2,..., p(k)) \quad (2.33)$$

Indeed, as $\boldsymbol{D}_{l_x l_x}(k)$ in (2.11) is not a diagonal matrix in general, no individual redundancy indexes in components become meaningful here for $\boldsymbol{l}_x(k)$.

5) ***The variance of unit weight*** (the variance factor)

$$\hat{\sigma}_0^2(k) = \boldsymbol{d}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{d}(k)/p(k) \quad (2.34)$$

or

$$\hat{\sigma}_0^2(k) = [\boldsymbol{v}_{l_x}^T(k)\boldsymbol{D}_{l_x l_x}^{-1}(k)\boldsymbol{v}_{l_x}(k) + \boldsymbol{v}_{l_w}^T(k)\boldsymbol{Q}^{-1}(k)\boldsymbol{v}_{l_w}(k) + \boldsymbol{v}_z^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_z(k)]/p(k) \quad (2.35)$$

6) ***The posteriori variance matrix*** of the estimated state vector

$$\hat{\boldsymbol{D}}_{xx}(k) = \hat{\sigma}_0^2(k)\boldsymbol{D}_{xx}(k) \quad (2.36)$$

which directly reflects the latest available residuals due to the modeling and measurement errors. For the usage in (2.36), one can apply the epochwise variance factor as in (2.34) or (2.35), a regional variance factor, i.e., an average over a specific time period, or even a global variance factor from the entire data period [Wang, 1997, 2009; Wang et al, 2021]. ***However, it is noticed that plenty of the applications with applying Kalman filter have inappropriately considered (2.4), instead of (2.36), as their posteriori state variance matrix***.

Refer to [Wang, 1997, 2008, 2009; Caspary and Wang, 1998; Wang et al, 2021] for more details about this alternate formulation and its advantages for error analysis in discrete Kalman filtering.

3. GENERIC FORMULATION OF DISCRETE KALMAN FILTER WITH CONSTRAINTS

This section provides readers with our original development of a generic formula set, which meaningfully serves as an innovative framework for comprehensive error analysis in discrete Kalman filtering with constraints in parallel with the one summarized in Section 2.3, and also describes their connections. In this work, the constraints are restricted to the equality constraints,

3.1 The Functional and Stochastic Models

Upon the models of the standard Kalman filter defined in Section 2.1, a Kalman filter with constraints indicates that there exist the following additional constraints among the states

$$\boldsymbol{H}^T(k)\boldsymbol{x}(k) - \boldsymbol{h} = \boldsymbol{o} \quad (3.1)$$

wherein $\boldsymbol{H}(k)$ is a *n*×*h*-dimensional coefficient matrix and $tr[\boldsymbol{H}(k)] = h$ ($h < n$), which is either originally linear or linearized from nonlinear constraints and $\boldsymbol{h}$ is the *h*-dimensional constant

vector. Hence, the equations (2.1), (2.2) and (3.1) together represent the system model, the measurement model, and the constraints among the states in discrete Kalman filtering.

In analogy to the alternate formulation summarized in Section 2.3, the Principle of Least Squares is straightforwardly applied epochwise hereinafter to result the solution for discrete Kalman filter with constraints. To demonstrate the flexibility in dealing with the available functional and stochastic models, three different ways that deliver an identical estimate of the state vector $\boldsymbol{x}(k)$ are introduced in Sections 3.2, 3.3 and 3.4, respectively, of which Section 3.4 is the focus of attention of this manuscript.

3.2 Approach One

The measurement equation system is here structured as follows

1) ***a pseudo-measurement vector*** $\boldsymbol{l}'_x(k)$ is given directly by using the solution of the state vector from the standard Kalman filter (without any constraints) as in Section 2, which establishes the following residual equation:

$$\boldsymbol{v}_x(k) = \boldsymbol{x}_h(k) - \boldsymbol{l}'_x(k) \tag{3.2}$$

wherein $\boldsymbol{x}_h(k)$ is the state estimate subject to the constraints as in (3.1) while the pseudo-measurement vector and its variance matrix are:

$$\boldsymbol{l}'_x(k) = \boldsymbol{x}(k/k) \quad \text{(refer to (2.3))} \tag{3.3}$$

$$\boldsymbol{D}_{l'_x l'_x}(k) = \boldsymbol{D}_{xx}(k) \quad \text{(refer to (2.4))} \tag{3.4}$$

2) ***a group of*** *h* ***linear constraints*** as in (3.1) are applied.

The equations (3.2) and (3.1) together compose the model in the form of indirect observations with constraints. So, the Principle of Least Squares is applied to the following cost function at epoch *k*:

$$\begin{aligned}&\boldsymbol{min}: g(\boldsymbol{x}_h(k)/\boldsymbol{l}'_x(k), \boldsymbol{z}(k)) = \\ &\boldsymbol{v}_{l'_x}^T(k)\boldsymbol{D}_{l'_x l'_x}^{-1}(k)\boldsymbol{v}_{l'_x}(k) + 2\boldsymbol{k}_h^T[\boldsymbol{H}^T(k)\boldsymbol{x}_h(k) - \boldsymbol{h}]\end{aligned} \tag{3.5}$$

which was called the Mean Square Method in [Simon and Chia, 2000]. To seek for the (minimum) extreme value of (3.5), its first order derivative with respect to $\boldsymbol{x}_h(k)$ is assigned to 0:

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{x}_h(k)} g(\boldsymbol{x}_h(k)/\boldsymbol{l}'_x(k), \boldsymbol{z}(k)) \\ &= 2\boldsymbol{v}_{l'_x}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k) + 2\boldsymbol{k}_h^T\boldsymbol{H}^T(k) = \boldsymbol{o}\end{aligned} \tag{3.6}$$

which yields

$$\boldsymbol{D}_{xx}^{-1}(k)\boldsymbol{x}_h(k) + \boldsymbol{H}(k)\boldsymbol{k}_h(k) - \boldsymbol{D}_{xx}^{-1}(k)\boldsymbol{l}'_x(k) = \boldsymbol{o} \tag{3.7}$$

The equations (3.7) and (3.1) together compose a normal equation system:

$$\begin{pmatrix} \boldsymbol{D}_{xx}^{-1}(k/k) & \boldsymbol{H}(k) \\ \boldsymbol{H}^T(k) & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_h(k) \\ \boldsymbol{k}_h \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{xx}^{-1}(k/k)\boldsymbol{l}'_x(k) \\ \boldsymbol{h} \end{pmatrix} \tag{3.8}$$

which possesses two unknown parameter vectors: the state vector $\boldsymbol{x}_h(k)$ and the Lagrange multiplier vector $\boldsymbol{k}_h(k)$ brought by the constraints.

To solve (3.8), one can first derive $\boldsymbol{x}_h(k)$ from the first equation:

$$\begin{aligned}\boldsymbol{x}_h(k) &= \boldsymbol{D}_{xx}(k)[\boldsymbol{D}_{xx}^{-1}(k/k)\boldsymbol{l}'_x(k) - \boldsymbol{H}(k)\boldsymbol{k}_h(k)] \\ &= \boldsymbol{x}(k/k) - \boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{k}_h(k)\end{aligned} \tag{3.9}$$

wherein $\boldsymbol{x}(k/k)$ is the minimum variance estimate of the state vector given in (2.3). Substituting (3.9) into the second equation of (3.8) delivers the Lagrange multiplier vector $\boldsymbol{k}_h(k)$:

$$\boldsymbol{k}_h(k) = \boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k) - \boldsymbol{h}] \tag{3.10}$$

where a helping matrix $\boldsymbol{N}_{hh}(k)$ is defined to simplify the notation

$$\boldsymbol{N}_{hh}(k) = \boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k) \tag{3.11}$$

The substitution of (3.10) into (3.9) gives $\boldsymbol{x}_h(k)$

$$\begin{aligned}&\hat{\boldsymbol{x}}_h(k) = \boldsymbol{x}(k/k) \\ &- \boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\{\boldsymbol{H}^T(k)\boldsymbol{x}(k/k) - \boldsymbol{h}\}\end{aligned} \tag{3.12}$$

in which the overhead symbol ^ is commonly ignored wherever no confusion may occur.

The associated variance matrix with $\boldsymbol{x}_h(k)$ is derived based on (3.12):

$$\begin{aligned}&\boldsymbol{D}_{\boldsymbol{x}_h\boldsymbol{x}_h}(k) = \boldsymbol{D}_{xx}(k/k) \\ &- \boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\end{aligned} \tag{3.13}$$

wherein $\boldsymbol{D}_{xx}(k/k)$ is the variance matrix of $\boldsymbol{x}(k/k)$ in (2.4).

This solution is indeed identical with the one after Maximum Probability Method and Projection Method presented in Simon and Chia [2000].

3.3 Approach Two

Differently from Approach One in Section 3.2, the measurement equation system is here structured as follows

1) ***A pseudo-measurement vector*** $\boldsymbol{l}''_x(k)$ is defined by the predicted state vector $\boldsymbol{x}(k/k-1)$ from t_{k-1} to t_k (i.e., time update) from the standard Kalman filter as in Section 2, which establishes the following residual equation:

$$\boldsymbol{v}_{l''_x}(k) = \boldsymbol{x}_h(k) - \boldsymbol{l}''_x(k) \tag{3.14}$$

with

$$\boldsymbol{l}''_x(k) = \boldsymbol{x}(k/k-1) = \boldsymbol{A}(k)\boldsymbol{x}(k-1) \tag{3.15}$$

$$\boldsymbol{D}_{\boldsymbol{l}_x''\boldsymbol{l}_x''}(k) = \boldsymbol{D}_{\boldsymbol{xx}}(k/k-1) = \boldsymbol{A}(k)\boldsymbol{D}_{\boldsymbol{xx}}(k-1)\boldsymbol{A}^T(k) + \boldsymbol{B}(k)\boldsymbol{Q}(k)\boldsymbol{B}^T(k) \quad (3.16)$$

2) ***A measurement vector*** $\boldsymbol{l}_z(k)$ is adapted from the real measurement vector $\boldsymbol{z}(k)$ at t_k as in (2.2), which yields the following residual equation:

$$\boldsymbol{v}_{l_z}(k) = \boldsymbol{v}_z(k) = \boldsymbol{C}(k)\boldsymbol{x}_h(k) - \boldsymbol{l}_z(k) \quad (3.17)$$

with

$$\boldsymbol{l}_z(k) = \boldsymbol{z}(k) \quad (3.18)$$

$$\boldsymbol{D}_{l_z l_z}(k) = \boldsymbol{R}(k) \quad (3.19)$$

wherein $\boldsymbol{C}(k)$ is the same as in (2.2).

3) ***a group of*** *h* ***linear constraints*** as in (3.1) are applied, wherein ***h*** (**bold** and *italic*) is the constant vector in the constraints.

Now, the equations (3.14), (3.17) and (3.1) together compose another model in the form of indirect observations with constraints at epoch *k*. Accordingly, the cost function for applying the Principle of Least Squares is as below:

$$\min: g(\boldsymbol{x}_h(k)/\boldsymbol{l}_x''(k), \boldsymbol{z}(k)) = \boldsymbol{v}_{l_x''}^T(k)\boldsymbol{D}_{l_x''l_x''}^{-1}(k)\boldsymbol{v}_{l_x''}(k) + \boldsymbol{v}_{l_z}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_{l_z}(k) + 2\boldsymbol{k}_h^T(k)[\boldsymbol{H}^T(k)\boldsymbol{x}_h(k) - \boldsymbol{h}(k)] \quad (3.20)$$

The same as with (3.5), the 1st order derivative of (3.20) with respect to $\boldsymbol{x}(k)$ is assigned to 0:

$$\frac{\partial}{\partial \boldsymbol{x}_h(k)} g(\boldsymbol{x}_h(k)/\boldsymbol{l}_x''(k), \boldsymbol{z}(k)) = 2\boldsymbol{v}_{l_x''}^T(k)\boldsymbol{D}_{l_x''l_x''}^{-1}(k) + 2\boldsymbol{v}_{l_z}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k) + 2\boldsymbol{k}_h^T\boldsymbol{H}^T(k) = \boldsymbol{o} \quad (3.21)$$

which gives

$$\{\boldsymbol{D}_{l_x l_x}^{-1}(k) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\}\boldsymbol{x}_h(k) + \boldsymbol{H}(k)\boldsymbol{k}_h - \boldsymbol{D}_{l_x l_x}^{-1}(k)\boldsymbol{l}_x(k) - \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k) = \boldsymbol{o} \quad (3.22)$$

From (3.22) and (3.1), the normal equation system goes as follows:

$$\begin{pmatrix} \boldsymbol{D}_{l_x''l_x''}^{-1}(k) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k) & \boldsymbol{H}(k) \\ \boldsymbol{H}^T(k) & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_h(k) \\ \boldsymbol{k}_h \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{l_x''l_x''}^{-1}(k)\boldsymbol{l}_x''(k) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k) \\ \boldsymbol{h} \end{pmatrix} \quad (3.23)$$

which is identical to (3.8) because it can be proved

$$[\boldsymbol{D}_{l_x''l_x''}^{-1}(k) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)]^{-1} = \boldsymbol{D}_{xx}(k/k) \quad (3.24)$$

and

$$\boldsymbol{D}_{l_x''l_x''}^{-1}(k)\boldsymbol{l}_x''(k) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k) = \boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{x}(k/k-1) + \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k) = \boldsymbol{D}_{xx}^{-1}(k)\boldsymbol{x}(k) \quad (3.25)$$

This implies that (3.8) and (3.23) result in the identical solution for the state vector.

3.4 Approach Three

Furthermore, differently from Approaches One and Two, Approach Three here develops the proposed framework for comprehensive error analysis in discrete Kalman filtering with constraints, which is particularly an extension of (2.14) – (2.16) by adding the constraints among the states. The measurement equation system is hereto structured as follows:

1) ***The first pseudo-measurement vector*** $\boldsymbol{l}_x(k)$ is here defined by the predicted state vector exclusive of the effect of the process noise vector. Its residual equation is (refer to (2.14):

$$\boldsymbol{v}_{l_x}(k) = \boldsymbol{x}_h(k) - \boldsymbol{B}(k)\boldsymbol{w}_h(k) - \boldsymbol{l}_x(k) \quad (3.25)$$

$$\boldsymbol{l}_x(k) = \boldsymbol{A}(k)\boldsymbol{x}(k-1) \quad (3.26)$$

$$\boldsymbol{D}_{l_x l_x}(k) = \boldsymbol{A}(k)\boldsymbol{D}_{\boldsymbol{xx}}(k-1)\boldsymbol{A}^T(k) \quad (3.27)$$

2) ***The second pseudo-measurement vector*** $\boldsymbol{l}_w(k)$ is defined by the process noise vector, which gives the residual equation below (refer to (2.15):

$$\boldsymbol{v}_{l_w}(k) = \boldsymbol{w}_h(k) - \boldsymbol{l}_w(k) \quad (3.28)$$

$$\boldsymbol{l}_w(k) = \boldsymbol{w}_0(k) \quad (\text{usually } \boldsymbol{w}_0(k) = \boldsymbol{o}) \quad (3.29)$$

$$\boldsymbol{D}_{l_w l_w}(k) = \boldsymbol{Q}(k) \quad (3.30)$$

3) ***A measurement vector*** $\boldsymbol{l}_z(k)$ is adapted from the real measurement vector $\boldsymbol{z}(k)$ at t_k as in (2.2). So, the residual equation is as (3.17) alongside with (3.18) and (3.19).

4) a group of *h* ***linear or linearized constraints*** are as in (3.1).

Essentially, one must give one's attention to (3.26), $\boldsymbol{l}_x(k) \neq \boldsymbol{x}(k/k-1)$ because

$$\boldsymbol{x}(k/k-1) = \boldsymbol{l}_x(k) + \boldsymbol{B}(k)\boldsymbol{l}_w(k) = \boldsymbol{A}(k)\boldsymbol{x}(k-1) + \boldsymbol{B}(k)\boldsymbol{w}_0(k) \quad (3.31)$$

Writing four equations (3.25), (3.28), (3.17), and (3.1) together gives the entire residual equation system with constraints as below:

$$\begin{pmatrix} \boldsymbol{v}_x^h(k) \\ \boldsymbol{v}_w^h(k) \\ \boldsymbol{v}_z^h(k) \end{pmatrix} = \begin{pmatrix} \boldsymbol{E}_x & -\boldsymbol{B}(k) \\ \boldsymbol{O} & \boldsymbol{E}_w \\ \boldsymbol{C}(k) & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_h(k) \\ \boldsymbol{w}_h(k) \end{pmatrix} - \begin{pmatrix} \boldsymbol{l}_x(k) \\ \boldsymbol{l}_w(k) \\ \boldsymbol{l}_z(k) \end{pmatrix} \quad (3.32)$$

$$\boldsymbol{H}^T(k)\boldsymbol{x}_h(k) - \boldsymbol{h} = \boldsymbol{o}$$

alongside with the blockwise covariance matrix of three independent measurement vectors $\boldsymbol{l}_x(k)$, $\boldsymbol{l}_w(k)$ and $\boldsymbol{z}(k)$ as in (3.27), (3.30), (3.19). The main difference of (3.32) from Approach One in Section 3.2 and Approach Two in Section 3.3 lies in

directly modeling three originally independent random vectors as the measurement vectors. Accordingly, the unknown parameters have been extended from $\boldsymbol{x}_h(k)$ to both of $\boldsymbol{x}_h(k)$ and $\boldsymbol{w}_h(k)$, This modeling strategy allows estimating the process noise vector epochwise and also the residual vector of $\boldsymbol{l}_w(k)$, which has been of scarcely any mention in literature, except initially modeled in Wang [1997].

Frankly, (3.32) allows specifying the following cost function for applying the Principle of Least Squares:

$$\begin{aligned}&\boldsymbol{min}: g(\boldsymbol{x}_h(k),\boldsymbol{w}_h(k)/\boldsymbol{l}_x(k),\boldsymbol{l}_w(k),\boldsymbol{l}_z(k))\\&=\boldsymbol{v}_{l_x}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{v}_{l_x}(k)+\boldsymbol{v}_{l_w}^T(k)\boldsymbol{Q}^{-1}(k)\boldsymbol{v}_{l_w}(k)\\&\quad+\boldsymbol{v}_{l_z}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_{l_z}(k)\\&\quad+2\boldsymbol{k}_h^T(k)[\boldsymbol{H}^T(k)\boldsymbol{x}_h(k)-\boldsymbol{h}(k)]\end{aligned}\tag{3.33}$$

which yields two 1st order partial derivatives for $\boldsymbol{x}_h(k)$ and $\boldsymbol{w}_h(k)$, respectively:

$$\begin{aligned}&\frac{\partial g(\boldsymbol{x}_h(k),\boldsymbol{w}_h(k)/\boldsymbol{l}_x(k),\boldsymbol{l}_w(k),\boldsymbol{l}_z(k))}{\partial\boldsymbol{x}_h(k)}\\&=2\boldsymbol{v}_{l_x}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\\&\quad+2\boldsymbol{v}_{l_z}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)+2\boldsymbol{k}_h^T(k)\boldsymbol{H}^T(k)=\boldsymbol{o}\end{aligned}\tag{3.34}$$

$$\begin{aligned}&\frac{\partial g(\boldsymbol{x}_h(k),\boldsymbol{w}_h(k)/\boldsymbol{l}_x(k),\boldsymbol{l}_w(k),\boldsymbol{l}_z(k))}{\partial\boldsymbol{w}(k)}\\&=-2\boldsymbol{v}_{l_x}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{B}(k)+2\boldsymbol{v}_{l_w}^T(k)\boldsymbol{Q}^{-1}(k)=\boldsymbol{o}\end{aligned}\tag{3.35}$$

Together with (3.1), (3.34) and (3.35) build up the corresponding normal equation system:

$$\begin{pmatrix}\boldsymbol{D}_{l_xl_x}^{-1}(k)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k) & -\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{B}(k) & \boldsymbol{H}(k)\\ -\boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k) & \boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{B}(k)+\boldsymbol{Q}^{-1}(k) & \boldsymbol{O}\\ \boldsymbol{H}^T(k) & \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\cdot\begin{pmatrix}\boldsymbol{x}_h(k)\\ \boldsymbol{w}_h(k)\\ \boldsymbol{k}_h(k)\end{pmatrix}=\begin{pmatrix}\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{l}_x(k)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{l}_z(k)\\ -\boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{l}_x(k)+\boldsymbol{Q}^{-1}(k)\boldsymbol{l}_w(k)\\ \boldsymbol{h}(k)\end{pmatrix}\tag{3.36}$$

Although deducing an explicit solution of (3.36) affirmatively seems complicated because the coefficient matrix of (3.36) is in the form of a 3×3 partitioned block matrix, we have successfully accomplished the algorithmic formulation of the solution for $\boldsymbol{x}_h(k)$, $\boldsymbol{w}_h(k)$ and $\boldsymbol{k}_h(k)$ inclusive of some further relevant contents, e.g., the residual vectors and redundancy contribution and redundant indexes of the measurements etc.

Before the solution is delivered, the equivalency of (3.36) to (3.8) and (3.23) is first proved. With the 2nd equation in (3.36), three specifics need readers' attention for the benefit of further derivation:

i) The coefficient matrix of $\boldsymbol{x}_h(k)$ in the 1st equation of (3.26) is $\boldsymbol{D}_{xx}^{-1}(k/k)$ (refer to (3.24)).

ii) The inverse of the coefficient matrix of $\boldsymbol{w}_h(k)$ in the 2nd equation of (3.26) gives

$$\begin{aligned}&[\boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{B}(k)+\boldsymbol{Q}^{-1}(k)]^{-1}\\&=\boldsymbol{Q}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned}\tag{3.37}$$

iii) Solving for $\boldsymbol{w}_h(k)$ from the 2nd equation in (3.26) gives

$$\begin{aligned}\boldsymbol{w}_h(k)&=[\boldsymbol{Q}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\boldsymbol{Q}(k)]\cdot\\&\quad\cdot\boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{x}_h(k)\\&+[\boldsymbol{Q}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\boldsymbol{Q}(k)]\cdot\\&\quad\cdot[-\boldsymbol{B}^T(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{l}_x(k)+\boldsymbol{Q}^{-1}(k)\boldsymbol{l}_w(k)]\end{aligned}\tag{3.38}$$

Substituting (3.38) into the 1st equation of (3.36) eliminates $\boldsymbol{w}_h(k)$

$$\begin{aligned}&\begin{pmatrix}\boldsymbol{D}_{xx}^{-1}(k/k) & \boldsymbol{H}(k)\\ \boldsymbol{H}^T(k) & \boldsymbol{O}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_h(k)\\ \boldsymbol{k}_h(k)\end{pmatrix}=\\&\begin{pmatrix}\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k)\\ \boldsymbol{h}\end{pmatrix}\end{aligned}\tag{3.39}$$

which proved that (3.36) is equivalent to (3.8) and (3.23) for $\boldsymbol{x}_h(k)$ and $\boldsymbol{k}_h(k)$ as

$$\begin{aligned}&\boldsymbol{D}_{xx}^{-1}(k/k-1)[\boldsymbol{l}_x(k)+\boldsymbol{B}(k)\boldsymbol{l}_w(k)]+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k)\\&=\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k)\\&=\boldsymbol{D}_{xx}^{-1}(k)\boldsymbol{x}(k)\end{aligned}\tag{3.40}$$

3.5 Solution

Now, without providing the lengthy intermediate steps, the solution of $\boldsymbol{x}_h(k)$, $\boldsymbol{w}_h(k)$ and $\boldsymbol{k}_h(k)$ is directly given below:

$$\begin{aligned}\boldsymbol{x}_h(k)&=\boldsymbol{x}(k/k-1)+\boldsymbol{K}(k)\boldsymbol{d}(k)\\&\quad-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{k}_h(k)\end{aligned}\tag{3.41}$$

$$\begin{aligned}\boldsymbol{w}_h(k)&=\boldsymbol{l}_w(k)+\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\&\quad-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\cdot\\&\quad\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned}\tag{3.42}$$

$$\begin{aligned}&\boldsymbol{k}_h(k)=\boldsymbol{N}_{hh}^{-1}(k)\{\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\cdot\\&\cdot[\boldsymbol{D}_{l_xl_x}^{-1}(k)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k)]-\boldsymbol{h}\}\end{aligned}\tag{3.43}$$

with the variance–covariance matrices of $\boldsymbol{x}_h(k)$ and $\boldsymbol{w}_h(k)$:

$$\begin{aligned}&\boldsymbol{D}_{\boldsymbol{x}_h\boldsymbol{x}_h}(k)=\boldsymbol{D}_{xx}(k/k-1)\\&-\boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k-1)\\&-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\end{aligned}\tag{3.44}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{w}_h\boldsymbol{w}_h}(k) &= \boldsymbol{Q}(k)\\ &-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\boldsymbol{Q}(k)\\ &-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\cdot\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned} \tag{3.45}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{x}_h\boldsymbol{w}_h}(k-1) &= [\boldsymbol{I}-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)]\cdot\\ &\cdot[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned} \tag{3.46}$$

3.6 Solutions with and without Constraints

The solution of $\boldsymbol{x}_h(k)$ and $\boldsymbol{w}_h(k)$ in discrete Kalman filtering with constraints is connected to the solution of $\boldsymbol{x}(k)$ and $\boldsymbol{w}(k)$ (without constraints) given in Section 2.3 as follows:

$$\begin{aligned}\boldsymbol{x}_h(k) &= \boldsymbol{x}(k/k)\\ &-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.47}$$

$$\begin{aligned}\boldsymbol{w}_h(k) &= \boldsymbol{w}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\cdot\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.48}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{x}_h\boldsymbol{x}_h}(k) &= \boldsymbol{D}_{xx}(k/k)\\ &-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\end{aligned} \tag{3.49}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{w}_h\boldsymbol{w}_h}(k-1) &= \boldsymbol{D}_{ww}(k)\\ &-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\cdot\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned} \tag{3.50}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{x}_h\boldsymbol{w}_h}(k-1) &= \boldsymbol{D}_{xw}(k)-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\cdot\\ &\cdot\boldsymbol{H}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned} \tag{3.51}$$

This group of formulas provides the opportunity to obtain the solution with constraints directly upon the solution from the standard Kalman filtering described in Section 2. A hard-won advantage of the solution expressions from (3.47) to (3.51) lies in first obtaining the solution after (2.3) (or (2.18)), (2.19), (2.4), (2.20) and (2.21) without considering the constraints and then utilizing $\boldsymbol{x}(k/k)$ to linearize the constraints, when they are nonlinear, and apply them towards the solution with constraints.

3.7 Residual Vectors and their Variance Matrices

For error analysis in Kalman filtering, $\boldsymbol{v}_{x_h}(k)$, $\boldsymbol{v}_{w_h}(k)$ (when $\boldsymbol{w}_0(k)=\boldsymbol{o}$) and $\boldsymbol{v}_{z_h}(k)$ with their associated covariance matrices are further derived below.

In general, they can directly be calculated according to (3.32) or individually after (3.25), (3.28) and (3.17). However, they are further detailed.

First, with the residual vector $\boldsymbol{v}_x^h(k)$ of $\boldsymbol{l}_x(k)$ in (3.25), substituting (3.41) and (3.42) or (3.47) and (3.48) into (3.25) gives

$$\begin{aligned}\boldsymbol{v}_{l_x}^h(k) &= \boldsymbol{x}(k/k)-\boldsymbol{l}_x(k)-\boldsymbol{B}(k)\boldsymbol{w}(k)\\ &-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\\ &+\boldsymbol{B}(k)\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\cdot\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.52}$$

or

$$\begin{aligned}\boldsymbol{v}_{l_x}^h(k) &= \boldsymbol{D}_{l_xl_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\ &-\boldsymbol{D}_{l_xl_x}(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\cdot\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.52a}$$

Based on (2.22), (3.52) is further simplified to

$$\begin{aligned}\boldsymbol{v}_{l_x}^h(k) &= \boldsymbol{v}_{l_x}(k)-[\boldsymbol{I}+\boldsymbol{B}(k)\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)]\\ &\cdot\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.53}$$

Second, with the residual vector $\boldsymbol{v}_{w_h}(k)$ of $\boldsymbol{l}_w(k)$ in (3.28), the substitution of (3.42) or (3.48) yields

$$\begin{aligned}\boldsymbol{v}_w^h(k) &= \boldsymbol{w}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]-\boldsymbol{l}_w(k)\end{aligned} \tag{3.54}$$

or

$$\begin{aligned}\boldsymbol{v}_{l_w}^h(k) &= \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\ &-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.54a}$$

After (2.23), (3.54) is further reformed to

$$\begin{aligned}\boldsymbol{v}_{l_w}^h(k) &= \boldsymbol{v}_{l_w}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.55}$$

Because the initial value of $\boldsymbol{l}_w(k)$ is usually assumed to be: $\boldsymbol{w}_0(k)=\boldsymbol{o}$ in practice, (3.54) becomes

$$\begin{aligned}\boldsymbol{v}_{l_w}^h(k) &= \boldsymbol{w}(k)-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.56}$$

Third, with the residual vector $\boldsymbol{v}_z^h(k)$ of $\boldsymbol{l}_z(k)=\boldsymbol{z}(k)$, substituting (3.41) or (3.47) into (3.17) delivers:

$$\begin{aligned}\boldsymbol{v}_z^h(k) &= \boldsymbol{C}(k)\{\boldsymbol{x}(k/k)-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\}-\boldsymbol{l}_z(k)\end{aligned} \tag{3.57}$$

or

$$\begin{aligned}\boldsymbol{v}_z^h(k) &= [\boldsymbol{C}(k)\boldsymbol{K}(k)-\boldsymbol{I}]\boldsymbol{d}(k)-\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.58}$$

According to (2.24), (3.58) is simplified to

$$\begin{aligned}\boldsymbol{v}_z^h(k) &= \boldsymbol{v}_z(k)-\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}^T(k)\boldsymbol{x}(k/k)-\boldsymbol{h}]\end{aligned} \tag{3.59}$$

The covariance matrix of the residual vectors for each of $\boldsymbol{v}_{x_h}(k)$, $\boldsymbol{v}_{w_h}(k)$ and $\boldsymbol{v}_{z_h}(k)$ are derived as follows:

(1) $\boldsymbol{D}_{v_{l_x}^h v_{l_x}^h}(k)$ is derived by applying the law of variance propagation to (3.52a)

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_x^h\boldsymbol{v}_x^h}(k) &= \boldsymbol{D}_{l_xl_x}(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{D}_{l_xl_x}(k)\\ &+\boldsymbol{D}_{l_xl_x}(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{D}_{l_xl_x}(k)\end{aligned}\quad(3.60)$$

as $\boldsymbol{D}_{\boldsymbol{xd}}(k)=\boldsymbol{O}$ in (2.10). Under the consideration of (2.25), (3.60) becomes

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_x^h\boldsymbol{v}_x^h}(k) &= \boldsymbol{D}_{\boldsymbol{v}_x\boldsymbol{v}_x}(k)+\boldsymbol{D}_{l_xl_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{l_xl_x}(k)\end{aligned}\quad(3.61)$$

(2) $\boldsymbol{D}_{v_{l_w}^h v_{l_w}^h}(k)$ is developed similarly by applying the law of variance propagation to (3.54a) :

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_w^h\boldsymbol{v}_w^h}(k) &= \boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\boldsymbol{Q}(k)\\ &+\boldsymbol{Q}(k)\boldsymbol{B}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned}\quad(3.62)$$

and further, based on (2.26),

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_w^h\boldsymbol{v}_w^h}(k) &= \boldsymbol{D}_{\boldsymbol{v}_w\boldsymbol{v}_w}(k)\\ &+\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\boldsymbol{Q}(k)\end{aligned}\quad(3.63)$$

(3) $\boldsymbol{D}_{v_z^h v_z^h}(k)$ is given by applying the law of variance propagation to (3.58)

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_z^h\boldsymbol{v}_z^h}(k) &= [\boldsymbol{I}-\boldsymbol{C}(k)\boldsymbol{K}(k)]\boldsymbol{R}(k)+\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{C}^T(k)\end{aligned}\quad(3.64)$$

and further according to (2.24),

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_z^h\boldsymbol{v}_z^h}(k) &= \boldsymbol{D}_{\boldsymbol{v}_z\boldsymbol{v}_z}(k)+\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\\ &\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{C}^T(k)\end{aligned}\quad(3.65)$$

3.8 Redundancy Contribution of Measurements

There are two levels of redundancy contribution: the total redundancy contribution of $\boldsymbol{l}_x(k)$, $\boldsymbol{l}_w(k)$ and $\boldsymbol{z}(k)$ together as well as the subtotal redundancy contribution of each of the groups, and the individual redundant indexes associated with each element in a group of the independent measurements, here specifically $\boldsymbol{l}_w(k)$ and $\boldsymbol{z}(k)$ because $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ are commonly diagonal in practice. The following discusses the redundancy contributions of $\boldsymbol{l}_x(k)$, $\boldsymbol{l}_w(k)$ and $\boldsymbol{l}_z(k)$ one by one:

(1) The redundancy contribution $r_{l_x}(k)$ of $\boldsymbol{l}_x(k)$

$$\begin{aligned}r_{l_x}(k) &= tr\{\boldsymbol{D}_{\boldsymbol{v}_x^h\boldsymbol{v}_x^h}(k)\boldsymbol{D}_{l_xl_x}^{-1}(k)\}\\ &= tr\{\boldsymbol{D}_{l_xl_x}(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\}\\ &+tr\{\boldsymbol{D}_{l_xl_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\}\end{aligned}\quad(3.66)$$

However, no individual redundant indexes will have the usual meaning for $\boldsymbol{l}_x(k)$ as its variance matrix of $\boldsymbol{D}_{\boldsymbol{l}_x\boldsymbol{l}_x}(k)=\boldsymbol{A}(k,k-1)\boldsymbol{D}_{\boldsymbol{xx}}(k-1)\boldsymbol{A}^T(k,k-1)$ will not be possibly diagonal in reality.

(2) The redundancy contribution $r_{l_w}(k)$ of $\boldsymbol{l}_w(k)$

$$\begin{aligned}r_{l_w}(k) &= tr\{\boldsymbol{D}_{\boldsymbol{v}_w^h\boldsymbol{v}_w^h}(k)\boldsymbol{D}_{l_wl_w}^{-1}(k)\}\\ &= tr\{\boldsymbol{D}_{\boldsymbol{v}_w\boldsymbol{v}_w}(k)\boldsymbol{Q}^{-1}(k)\}\\ &+tr\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\}\end{aligned}\quad(3.67)$$

or

$$\begin{aligned}r_{l_w}(k) &= tr\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\}\\ &+tr\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\}\end{aligned}\quad(3.68)$$

Besides, the individual redundant index associated with each component in $\boldsymbol{l}_w(k)$, when $\boldsymbol{Q}(k)$ is diagonal, is derived as follows

$$\begin{aligned}r_{l_w}^i(k) &= \{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\}_{ii}\\ &+\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\}_{ii}\\ &(i=1,2,...,m)\end{aligned}\quad(3.69)$$

(3) The redundancy contribution $r_{l_z}(k)$ (or $r_z(k)$) for $\boldsymbol{l}_z(k)$ (or $\boldsymbol{z}(k)$)

$$\begin{aligned}r_z(k) &= trace\{\boldsymbol{D}_{\boldsymbol{v}_z^h\boldsymbol{v}_z^h}(k)\boldsymbol{D}_{l_zl_z}^{-1}(k)\}\\ &= tr\{\boldsymbol{D}_{\boldsymbol{v}_z\boldsymbol{v}_z}(k)\boldsymbol{R}^{-1}(k)\}\\ &+tr\{\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\}\end{aligned}\quad(3.70)$$

in which the first item is

$$\begin{aligned}tr\{\boldsymbol{D}_{\boldsymbol{v}_z\boldsymbol{v}_z}(k)\boldsymbol{R}^{-1}(k)\} &= tr\{[\boldsymbol{I}-\boldsymbol{C}(k)\boldsymbol{K}(k)]\boldsymbol{R}(k)\boldsymbol{R}^{-1}(k)\}\\ &= tr[\boldsymbol{I}_{p\times p}-\boldsymbol{C}(k)\boldsymbol{K}(k)]\\ &= p(k)-tr[\boldsymbol{C}(k)\boldsymbol{K}(k)]\end{aligned}\quad(3.71)$$

When $\boldsymbol{R}(k)$ is diagonal, the individual redundant index with each component in $\boldsymbol{l}_z(k)$ is

$$\begin{aligned}r_z^i(k) &= \{\boldsymbol{D}_{\boldsymbol{v}_z\boldsymbol{v}_z}(k)\boldsymbol{R}^{-1}(k)\}_{ii}+\{\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\\ &\cdot\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\}_{ii}\end{aligned}\quad(3.72)$$

and further

$$r_z^i(k) = [\boldsymbol{I} - \boldsymbol{C}(k)\boldsymbol{K}(k)]_{ii} + [\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k) \cdot \boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)]_{ii} \quad (i = 1,2,..., p) \tag{3.73}$$

Finally, the total redundancy contribution of the three independent observation vectors together at epoch k, i.e., total redundancy number of $\boldsymbol{l}_x(k)$, $\boldsymbol{l}_w(k)$ and $\boldsymbol{l}_z(k)$ together is equal to

$$r(k) = r_{l_x}(k) + r_{l_w}(k) + r_z(k) \tag{3.74}$$

with the following specific detail,

$$\begin{aligned} r(k) = {} & tr\{\boldsymbol{D}_{l_x l_x}(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\} \\ & + tr\{\boldsymbol{D}_{l_x l_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k) \\ & \quad \cdot \boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\} \\ & + tr\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{C}^T(k)\boldsymbol{D}_{dd}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{B}(k)\} \\ & + tr\{\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k) \\ & \quad \cdot \boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{B}(k)\} \\ & + tr\{\boldsymbol{I} - \boldsymbol{C}(k)\boldsymbol{K}(k)\} \\ & + tr\{\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k) \\ & \quad \cdot \boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\} \end{aligned}$$

It can be proved that the total redundant index at epoch k is equal to

$$\begin{aligned} r(k) &= p(k) + tr\{\boldsymbol{N}_{hh}^{-1}(k)\boldsymbol{H}^T(k)\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\} \\ &= p(k) + tr\{\boldsymbol{I}_{h\times h}\} = p(k) + h(k) \end{aligned} \tag{3.75}$$

which means

$$r(k) = r_{l_x}(k) + r_{l_w}(k) + r_z(k) = p(k) + h(k) \tag{3.76}$$

with $p(k)$ and $h(k)$ being the number of the measurements in $\boldsymbol{z}(k)$ and the number of the constraints in (3.1).

3.9 Other Aspects

In addition, several algorithmic developments such as test statistics, variance factors and variance component estimation etc. may be further conducted, in analogy to the work in [Wang, 1997, 2008, 2009 etc.] and are excluded here due to the space restriction, except the following essential remark about the variance of unit weight:

(i) The variance of unit weight for Section 2 (standard discrete Kalman filter): the one in (2.34) is identical to the one in (2.35).

(ii) The variance of unit weight for Section 3 (discrete Kalman filter with constraints):

$$\hat{\sigma}_0^2 = [\boldsymbol{v}_{l_x}^{hT}(k)\boldsymbol{D}_{l_x l_x}^{-1}(k)\boldsymbol{v}_{l_x}^h(k) + \boldsymbol{v}_{l_w}^{hT}(k)\boldsymbol{Q}^{-1}(k)\boldsymbol{v}_{l_w}^h(k) + \boldsymbol{v}_z^{hT}(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_z^h(k)]/r(k) \tag{3.77}$$

as

$$\begin{aligned} \boldsymbol{d}^T(k)D_{dd}^{-1}(k)\boldsymbol{d}(k) \neq {} & [\boldsymbol{v}_{l_x}^{hT}(k)\boldsymbol{D}_{l_x l_x}^{-1}(k)\boldsymbol{v}_{l_x}^h(k) \\ & + \boldsymbol{v}_{l_w}^{hT}(k)\boldsymbol{Q}^{-1}(k)\boldsymbol{v}_{l_w}^h(k) + \boldsymbol{v}_z^{hT}(k)\boldsymbol{R}^{-1}(k)\boldsymbol{v}_z^h(k)] \end{aligned} \tag{3.78}$$

4. Algorithm in the Form of Extended Kalman Filter with Constraints

This section frames the relevant formulas in the form of Extended Kalman filter in accordance with the functional model defined in Section 3.1, but having them (i.e., (2.1), (2.2) and (3.1)) nonlinear.

The system model, the measurement model and constraint model appear nonlinear as follows:

$$\boldsymbol{x}(k) = \boldsymbol{A}(\boldsymbol{x}(k-1), k) + \boldsymbol{B}(k, k-1)\boldsymbol{w}(k-1) \tag{4.1}$$

$$\boldsymbol{z}(k) = \boldsymbol{C}(\boldsymbol{x}(k), k) + \boldsymbol{\Delta}(k) \tag{4.2}$$

$$\boldsymbol{H}(\boldsymbol{x}(k), k) - \boldsymbol{h} = \boldsymbol{o} \tag{4.3}$$

As for the variance propagation, three Jacobian matrices are derived here,

$$\boldsymbol{A}(k, k-1) = \boldsymbol{A}(k) = \frac{\partial \boldsymbol{A}(\boldsymbol{x}(k-1), k)}{\partial \boldsymbol{x}(k-1)} = \begin{pmatrix} \frac{\partial A_1(\boldsymbol{x}(k-1),k)}{\partial x_1(k-1)} & \frac{\partial A_1(\boldsymbol{x}(k-1),k)}{\partial x_2(k-1)} & \cdots & \frac{\partial A_1(\boldsymbol{x}(k-1),k)}{\partial x_n(k-1)} \\ \frac{\partial A_2(\boldsymbol{x}(k-1),k)}{\partial x_1(k-1)} & \frac{\partial A_2(\boldsymbol{x}(k-1),k)}{\partial x_2(k-1)} & \cdots & \frac{\partial A_2(\boldsymbol{x}(k-1),k)}{\partial x_n(k-1)} \\ \vdots & \vdots & & \vdots \\ \frac{\partial A_n(\boldsymbol{x}(k-1),k)}{\partial x_1(k-1)} & \frac{\partial A_n(\boldsymbol{x}(k-1),k)}{\partial x_2(k-1)} & \cdots & \frac{\partial A_n(\boldsymbol{x}(k-1),k)}{\partial x_n(k-1)} \end{pmatrix} \tag{4.4}$$

which is with respect to the estimated state vector $\boldsymbol{x}(k-1/k-1)$ at t_{k-1},

$$\boldsymbol{C}(k) = \frac{\partial \boldsymbol{C}(\boldsymbol{x}(k), k)}{\partial \boldsymbol{x}(k)} = \begin{pmatrix} \frac{\partial C_1(\boldsymbol{x}(k),k)}{\partial x_1(k)} & \frac{\partial C_1(\boldsymbol{x}(k),k)}{\partial x_2(k)} & \cdots & \frac{\partial C_1(\boldsymbol{x}(k),k)}{\partial x_n(k)} \\ \frac{\partial C_2(\boldsymbol{x}(k),k)}{\partial x_1(k)} & \frac{\partial C_2(\boldsymbol{x}(k),k)}{\partial x_2(k)} & \cdots & \frac{\partial C_2(\boldsymbol{x}(k),k)}{\partial x_n(k)} \\ \vdots & \vdots & & \vdots \\ \frac{\partial C_p(\boldsymbol{x}(k),k)}{\partial x_1(k)} & \frac{\partial C_p(\boldsymbol{x}(k),k)}{\partial x_2(k)} & \cdots & \frac{\partial C_p(\boldsymbol{x}(k),k)}{\partial x_n(k)} \end{pmatrix} \tag{4.5}$$

which is with respect to the predicted state vector $\boldsymbol{x}(k/k-1)$ through the time update from t_{k-1} to t_k, and

$$\boldsymbol{H}^T(k)=\frac{\partial \boldsymbol{H}(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}(k)}=\begin{pmatrix} \frac{\partial \boldsymbol{H}_1(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_1(k)} & \frac{\partial \boldsymbol{H}_1(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_2(k)} & \cdots & \frac{\partial \boldsymbol{H}_1(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_n(k)} \\ \frac{\partial \boldsymbol{H}_2(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_1(k)} & \frac{\partial \boldsymbol{H}_2(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_2(k)} & \cdots & \frac{\partial \boldsymbol{H}_2(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_n(k)} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \boldsymbol{H}_h(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_1(k)} & \frac{\partial \boldsymbol{H}_h(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_2(k)} & \cdots & \frac{\partial \boldsymbol{H}_h(\boldsymbol{x}(k),k)}{\partial \boldsymbol{x}_n(k)} \end{pmatrix} \tag{4.6}$$

which is with respect to the estimated state vector $\boldsymbol{x}(k)$ through the measurement update before the constraints are applied at t_k.

The following gives the algorithm in the form of Extended Kalman filter by referring to Sections 3.4 and 3.5:

1) THE MEASUREMENT MODEL

The predicted state vector exclusive of the effect from the process noise vector:

$$\boldsymbol{v}_x(k)=\boldsymbol{x}_h(k)-\boldsymbol{B}(k,k-1)\boldsymbol{w}_h(k-1)-\boldsymbol{l}_x(k) \tag{3.25}$$

$$\boldsymbol{l}_x(k)=\boldsymbol{A}(\boldsymbol{x}_h(k-1),k,k-1) \text{ (vs. (3.26))} \tag{4.7}$$

$$\boldsymbol{D}_{l_x l_x}(k)=\boldsymbol{A}(k,k-1)\boldsymbol{D}_{xx}(k-1)\boldsymbol{A}^T(k,k-1) \tag{3.27}$$

The process noise vector as a group of the pseudo-measurements: the same as (3.28), (3.29) and (3.30).

A group of the measurements from the measurement vector $\boldsymbol{z}(k)$ at t_k:

$$\boldsymbol{v}_{l_z}(k)=\boldsymbol{C}(\boldsymbol{x}_h(k),k)-\boldsymbol{l}_z(k) \quad \text{(vs. (3.17))} \tag{4.8}$$

$$\boldsymbol{l}_z(k)=\boldsymbol{z}(k) \tag{3.18}$$

$$\boldsymbol{D}_{l_z l_z}(k)=\boldsymbol{R}(k) \tag{3.19}$$

A group of the constraints on the states:

$$\boldsymbol{H}(\boldsymbol{x}_h(k),k)-\boldsymbol{h}=\boldsymbol{o} \quad \text{(vs. (3.1))} \tag{4.9}$$

2) THE SOLUTION

The state vector, the process noise vector and the Lagrange multiplier vector:

$$\begin{aligned}\boldsymbol{x}_h(k/k)&=\boldsymbol{x}(k/k)\\&-\boldsymbol{D}_{xx}(k/k)\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}]\end{aligned} \tag{4.10}$$

$$\begin{aligned}\boldsymbol{w}_h(k-1)&=\boldsymbol{l}_w(k)\\&+\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\&-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{D}_{xx}(k/k)\\&\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}]\end{aligned} \tag{4.11}$$

$$\boldsymbol{k}_h(k)=\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}] \tag{4.12}$$

wherein

$$\boldsymbol{x}(k/k)=\boldsymbol{x}(k/k-1)+\boldsymbol{K}(k)\boldsymbol{d}(k) \tag{4.13}$$

$$\boldsymbol{x}(k/k-1)=\boldsymbol{A}(\boldsymbol{x}_h(k-1),k,k-1)+\boldsymbol{B}(k)\boldsymbol{w}_0(k) \tag{4.14}$$

$$\boldsymbol{d}(k)=\boldsymbol{z}(k)-\boldsymbol{A}(\boldsymbol{x}_h(k-1),k,k-1)-\boldsymbol{B}(k)\boldsymbol{w}_0(k) \tag{4.15}$$

The var-covariance matrices of the state vector and the process noise vector: the same as (3.49), (3.50) and (3.51).

3) THE MEASUREMENT RESIDUALS

The residual vectors:

$$\begin{aligned}\boldsymbol{v}_{l_x}^h(k)&=\boldsymbol{D}_{l_x l_x}(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\&-\boldsymbol{D}_{l_x l_x}(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\\&\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}]\end{aligned} \tag{4.16}$$

$$\begin{aligned}\boldsymbol{v}_{l_w}^h(k)&=\boldsymbol{Q}(k)\boldsymbol{B}^T(k)\boldsymbol{D}_{xx}^{-1}(k/k-1)\boldsymbol{K}(k)\boldsymbol{d}(k)\\&-\boldsymbol{Q}(k)\boldsymbol{B}^T(k)[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C}(k)]^T\\&\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}]\end{aligned} \tag{4.17}$$

$$\begin{aligned}\boldsymbol{v}_z^h(k)&=[\boldsymbol{C}(k)\boldsymbol{K}(k)-\boldsymbol{I}]\boldsymbol{d}(k)-\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k)\\&\cdot\boldsymbol{H}(k)\boldsymbol{N}_{hh}^{-1}(k)[\boldsymbol{H}(\boldsymbol{x}(k/k),k)-\boldsymbol{h}]\end{aligned} \tag{4.18}$$

The variance matrices of the residual vectors: are the same as (3.60) – (3.65).

4) **THE REDUNDANCY CONTRIBUTION**:

The same as in Section 3.8.

For the convenience of practical implementation and better understanding of the proposed framework, an algorithmic flow is suggested in Fig. 4.1.

5. CONCLUDING REMARKS

This manuscript exhibited flexible algorithmic formulation for Kalman filtering with equality constraints on the system states, and practically developed an analytic framework for comprehensive error analysis accordingly. Specifically, this manuscript has:

(a) Developed a unique formula set as an innovative framework on the base of the three independent error sources that influence the system state estimate (Section 3.1-3.6);

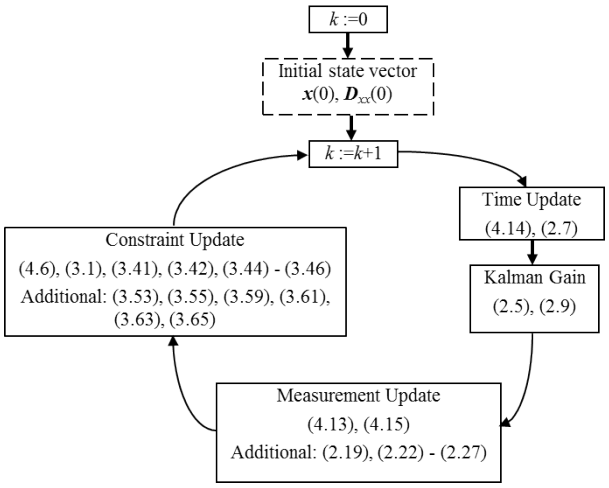


Fig. 4.1 an algorithmic flow of EKF with Constraints

(b) Specifically introduced the equation for the residual vector of the process noise vector , as well as their covariance matrices (Section 3.6);

(c) Made the reliability analysis feasible through parametrically introducing the redundancy contribution for the predicted state vector, process noise vector, and measurement vector, and the individual redundant indexes for the elements in the process noise and measurement vectors under the assumption of diagonal $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ (Section 3.7); and

(d) Pointed out its essential potentials how further algorithmic extension may be accomplished from the proposed formulation (Sections 3.9).

This work took an important step towards a standardized generic approach to performing Kalman filtering with equality constraints that enables comprehensive and rigorous error analysis, which is particularly important for high accuracy applications, for instance, the centimeter level kinematic positioning and navigation using GNSS and/or multisensor-integrated systems in the modern direct-georeferencing technology, autonomous vehicle driving, and other robotic applications etc., wherever it is important to examine the sources of any deviations in the estimated system states. The issue of comprehensive error analysis in Kalman filtering has been addressed previously [Wang, 1997; Caspary and Wang, 1998; Wang, et al, 2021; etc.], but not yet in the context of a Kalman filter with equality constraints. It is the authors' hope that comprehensive error analysis becomes a necessary part of the estimation process in the constrained Kalman filtering as a result of this work.

Acknowledgements:

The authors would like to sincerely thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support. We would also like to thank the editor-in-charge and anonymous reviewers for their valuable comments and suggestions.

References

Alouani, A. T. and Blair, W. D. (1991): Use of a kinematic constraint in tracking constant speed, maneuvering targets, *Proceedings of the 30th IEEE Conference on Decision and Control*, doi.org/10.1109/cdc.1991.261780.

Caspary, Wilhelm and Wang, Jianguo (1998): *Redundanzanteile und Varianzkomponentenim Kalman Filter*, Vol.123, No.4, 1998, pp.121-128.

Gopaul, Nilesh; Wang, Jianguo and and Jiming Guo (2010): Improving of GPS Observation Quality Assessment through Posteriori Variance-covariance Component Estimation, Proceedings of *CPGPS 2010 Navigation and Location Services: Emerging Industry and International Exchanges*, Scientific Research Publishing Inc., Shanghai, 2010.

Hasberg, C.; Hensel, S. and Stiller, C. (2012): Simultaneous localization and mapping for path-constrained motion. *IEEE Transactions on Intelligent Transportation Systems*, *13*(2), 541–552. doi.org/10.1109/tits.2011.2177522.

Khabbazi, M.-R. and Esfanjani, R. M. (2014): Constrained two-stage Kalman filter for Target Tracking, *4th International Conference on Computer and Knowledge Engineering, 2014*, doi.org/10.1109/iccke.2014.6993432.

Mikhail, Edward (1970): Parameter Constraints in Least Squares, Photogrammetric Engineering, 1970, pp. 1277 - 1291.

Pizzinga, Adrian (2012): Restricted Klaman Filtering – Theory, Methods, and Application, Springer Science+Business Media New York 2012.

Qian, Kun; Wang, Jianguo and Baoxin Hu (2016): A posteriori estimation of stochastic models for multi-sensor integrated inertial kinematic positioning and navigation on basis of variance component estimation, Journal of GPS, 2016, 14:5.

Qian, Kun (2017): Generic Multisensor Integration Strategy and Innovative Error Analysis for Integrated Navigation, PhD Dissertation, York University, Canada, 2017.

Qian, Kun; Wang, Jianguo and Baoxin Hu (2015): Novel Integration Strategy for GNSS-aided Inertial Integrated Navigation, No. 2, Vol. 69, Geomatica, pp. 217-230.

Rao, C. Radhakrishna and Toutenburg, Helge (1999): Linear Models: Least Squares and Alternatives, 2nd edition, Springer, ISBN 0-387-98848-3, 1999.

Simon, D. (2010): Kalman filtering with state constraints: A survey of linear and nonlinear algorithms, *IET Control Theory & Applications*, *4*(8), 1303–1318.

Simon, D. and Tien Li Chia (2002): Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems*, *38*(1), 128–136. https://doi.org/10.1109/7.993234.

Teixeira, B., Chandrasekar, J., et al (2008) : Gain-constrained Kalman filtering for linear and nonlinear systems, *IEEE Transactions on Signal Processing*, *56*(9), 4113–4123, doi.org/10.1109/tsp.2008.926101.

Wang, Jianguo; Qiu, Weining; Yao, Yinbin and Wu, Yun (2019): Error Theory and Foundation of Surveying Adjustment, English Edition, Wuhan University Press, 2019.

Wang, Jianguo (1997): *Filtermethoden zur fehler-toleranten kinematischen Positionsbestimmung*, Schritenreihe Studiengang Vermessungswesen, Federal Arm-Forced University Munich, Germany, No. 52, Neubiberg, 1997.

Wang, Jianguo (2008): Test Statistics in Kalman Filtering. Vol. 7, No. 1, Journal of GPS, pp. 81-90.

Wang, Jianguo (2009): Reliability Analysis in Kalman Filtering, Journal of GPS, Vol. 8, No. 1, pp.101-111.

Wang, Jianguo; Gopaul, Nilesh and Scherzinger, Bruno (2009): Simplified Algorithms of Variance Compo-nent Estimation for Static and Kinematic GPS Single Point Positioning**,** J. of GPS,Vol. 8, No. 1:43-51.

Wang, Jianguo; Gopaul, Nilesh and Jiming Guo (2010): Adaptive Kalman Filter based on Posteriori Variance-covariance Component Estimation, Proceedings of *CPGPS 2010 Navigation and Location Services: Emerging Industry and International Exchanges*, Scientific Research Publishing Inc., Shanghai, 2010.

Wang, Jianguo; Kun Qian and Baoxin Hu (2015): An Unconventional Full Tightly-Coupled Multi-Sensor Integration for Kinematic Positioning and Navigation, CSNC 2015 Proceedings, Volume III, Vol. 7, No.1, 2008, pp. 81~90.

Wang, Jianguo; Aaron Boda and Baoxin Hu (2021): Comprehensive Error Analysis beyond System Innovations in Kalman Filtering, Chapter 3, Learning Control, Elservier, pages 59-92.

Yang, C.; Bakich, M. and Blasch, E. (2005): Nonlinear constrained tracking of targets on roads, *2005 7th International Conference on Information Fusion*, doi.org/10.1109/icif.2005.1591860.

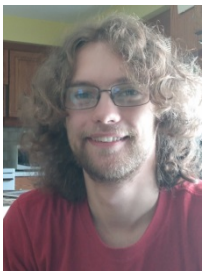
Appendix: Proof of (3.24) and (3.25)

$$
\begin{aligned}
&\{\boldsymbol{D}^{-1}_{l_x^n l_x^n}(k)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\}^{-1}\\
&=\{\boldsymbol{D}^{-1}_{xx}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\}^{-1}\\
&=\boldsymbol{D}_{xx}(k/k-1)-\boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\{\boldsymbol{R}(k)+\boldsymbol{C}(k)\boldsymbol{D}_{l_x l_x}(k)\boldsymbol{C}^T(k)\}^{-1}\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k-1)\\
&=\boldsymbol{D}_{xx}(k/k-1)-\boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\boldsymbol{D}^{-1}_{dd}(k)\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k-1)\\
&=[\boldsymbol{E}-\boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k)\boldsymbol{D}^{-1}_{dd}(k)\boldsymbol{C}(k)]\boldsymbol{D}_{xx}(k/k-1)\\
&=[\boldsymbol{E}-\boldsymbol{K}(k)\boldsymbol{C}(k)]\boldsymbol{D}_{xx}(k/k-1)\\
&=\boldsymbol{D}_{xx}(k/k)
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{D}^{-1}_{xx}(k)\boldsymbol{x}(k)=\boldsymbol{D}^{-1}_{xx}(k)[(\boldsymbol{E}-\boldsymbol{K}(k)\boldsymbol{C}(k))\boldsymbol{x}(k/k-1)+\boldsymbol{K}(k)\boldsymbol{z}(k)]\\
&=[\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)+\boldsymbol{D}^{-1}_{xx}(k/k-1)][(\boldsymbol{E}-\boldsymbol{K}(k)\boldsymbol{C}(k))\boldsymbol{x}(k/k-1)+\boldsymbol{K}(k)\boldsymbol{z}(k)]\\
&=\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)-\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{K}(k)\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)+\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{x}(k/k-1)\\
&\quad-\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{K}(k)\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{C}(k)\boldsymbol{K}(k)\boldsymbol{z}(k)+\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{K}(k)\boldsymbol{z}(k)\\
&=\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\{\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)\\
&\quad+[(\boldsymbol{D}_{dd}(k)-\boldsymbol{C}(k)\boldsymbol{D}_{xx}(k/k-1)\boldsymbol{C}^T(k))\boldsymbol{D}^{-1}_{dd}(k)+\boldsymbol{C}(k)\boldsymbol{K}(k)]\boldsymbol{d}(k)\}\\
&=\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\{\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)+\boldsymbol{z}(k)-\boldsymbol{C}(k)\boldsymbol{x}(k/k-1)\}\\
&=\boldsymbol{D}^{-1}_{xx}(k/k-1)\boldsymbol{x}(k/k-1)+\boldsymbol{C}^T(k)\boldsymbol{R}^{-1}(k)\boldsymbol{z}(k)
\end{aligned}
$$

Authors

Dr.-Ing. Jianguo Wang is a professor in Geomatics Science and Engineering, Department of Earth and Space Science and Engineering, York University, Canada. He holds a Dr.-Ing. in Surveying Engineering from the University of German Federal Armed Forces Munich (Universität der Bundeswher München), a MEng. and a BEng. in Surveying Engineering from Wuhan Technical University of Surveying and Mapping, China. His current research is focused on novel multisensor integration strategy for kinematic positioning and navigation and advanced estimation algorithms.



Benjamin Brunson is currently a Doctorate student at the Department of Earth and Space Science and Engineering of York University. His research is focused on multisensor integrated kinematic positioning and navigation, especially on the usage of IMU Arrays and development of advanced filtering algorithms. Prior to his PhD studies, he also received his B. Eng. In Geomatics engineering in 2017 and MSc. focus on phase unwrapping for SAR interferometry in Earth and Space Science in 2019 from the same institution.

Dr. Baoxin Hu received PhD degree in Remote Sensing and GIS at Boston University, USA in 1998. She is currently a Professor in Geomatics Science and Engineering, Department of Earth and Space Science and Engineering, York University, Canada. Her research interests include vegetation characterization from multi-source remotely sensed data, 3D scene reconstruction, information fusion, and deep learning.

Journal of Global Positioning Systems (*2022*)
Vol. 18, No. 1: 70-83
DOI: 10.5081/jgps.18.1.70

Bio-inspired Map Construction based on Brain Navigation Mechanism for Indoor Robots

Yixuan Long, Hao Wang, Fang Ye, Yibing Li, Qian Sun

College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001

✉ **Corresponding author**: Qian Sun, qsun@hrbeu.edu.cn

Abstract: Mapping is critical for an autonomous robot performing tasks in an unknown environment, which provides the environment information for task planning. Inspired by the presence of cells in the mammals' brain that help mammals rapidly cognize the surroundings, considering visual ambiguity that may be happened indoors, an orientation-independent boundary cell model based on the boundary vector cells in the brain is proposed to tackle the obstacle information in the environment, and it is fused into a metric-topological map to represent the structural information which increases the functionality of the map. The simulation results show that the expression of boundaries or obstacles in the environment can be obtained through the firing rate of boundary cells, which enhances the information content of the map. Meanwhile, the algorithm can build a consistent representation of the environment with sensor noise and achieves a root mean square error of 11.42cm in a 16m×17m indoor environment, effectively calibrating the sensor drift error, and ensuring the accuracy of the map.

Keywords: bio-inspired; mapping; metric-topological map; indoor robots

1 Introduction

In recent years, robots have been widely used in a variety of industries to execute tasks autonomously. Simultaneous localization and mapping (SLAM) is a critical technology for robots working in an unknown environment, which allows the robot to map the environment while running and then utilize the built map as the prior information for navigation. Thus, constructing a suitable map is significant considering the intention of the tasks and the working environment. As humans and many animals have the nature to explore and navigate in new environments, researchers are inspired to imitate this mechanism in robotics to enhance their mapping and cognitive ability.

The spatial cognition mechanism in the mammal's brain has been researched by neuroscientists for decades. Early in 1971, O'Keefe and Dostrovsky found that some cells in the rodent's hippocampus are active only when the rat visit a specific location, which serves as an internal cognitive map, named place cells [1]. Based on this discovery, more spatial cells are founded. In the entorhinal cortex, a kind of neuron is observed activating at multiple specific locations which form a hexagonal grid in space [2]. These grid cells integrate the rat's self-motion information and provide the path integration function of brain navigation, while head direction cells respond to the movement of the rat's head facing, improving an animal's ability to solve spatial problems [3]. With the help of the directional information, the boundary vector cells which produce a high firing rate when the rat reaches the boundary of the test environment can generate direction-independent activity [4]. Through the coordinated activity of neural circuits between these cells, the positioning system in the brain is formed.

Current research on bio-inspired map construction methods can be divided into two groups: one group uses the place cells to code the whole environment which means that each site corresponds to a specific neuron. Arleo proposed a computational model based on head direction cells and place cells to achieve navigation in a two-dimensional small-scale environment [5]. Tian used competitive Hebbian learning to select grid cell subsets to generate place cell population activities, thereby obtaining a robot exploration map [6]. Instead of using the place cells only, Zhou proposed a new place cell representation under multi-information perception generated by a weighted fusion of visual-related and grid cells [7]. Tang established a cognitive map based on the functioning of the hippocampus and the entorhinal cortex [8]. Although these works have high biology fidelity, the complexity and required resources could increase greatly when working in larger environments, while the other group builds the map using a hybrid map model based on the episodic memory mechanism, which can be utilized in a large area. Milford proposed a metric-topological experience map that records the activity of cells in a representative place as an experience so that the map is more extensible [9]–[11]. In this system, because visual information is the only source of allothetic cues, many improvements to the visual system have been done to increase the reliability of the map when facing visual ambiguity. The FAB-MAP is introduced to deal with the changes over time [12]. In [13], an appearance-based frequency-tuned model is proposed, which converts visual input into saliency maps to reduce the influence of light. To avoid perceptual aliasing, a dynamic growing self-organizing map based on direction and feature parameters is introduced and gains a good performance in the office environment [14]. There are many other sources of information used by animals during navigation, such as Zeng proposed that local view information and motion information were introduced into HD-by-velocity cells and GirD-by-velocity cells, and the two types of cells were connected and the relocation was achieved according to attractor dynamics [15]. Based on the self-centered confidence map, Gupta generates the current positioning through multi-scale confidence superposition [16]. In addition, literature [17]-[18] points out that distance information extracted from tactile sense and audio can also be used for localization. However, there are a few types of research on using structural information. In BatSLAM, visual templates are replaced by sonar fingerprints [19]. But in their experiment, using only one type of source still could suffer from ambiguity. For example, the lidar-based method can hardly distinguish repetitive environments with similar structures [20]. In ViTa-SLAM, the local tactile information is saved as a histogram template and a slope distribution array in the node [21], but it is hard to extract the global structure from the map. In this paper, a novel cognitive map is proposed to increase the feature redundancy and functionality in the indoor visual ambiguous environment. An ego-allocentric modulation method is proposed to transform the egocentric lidar information into the allocentric boundary information so that the lidar information can be fused into the map in a more biological way to provide an absolute reference for the localization process.

The rest of this paper is organized as follows. The computational model of navigation cells and the problem formulation are both introduced in Section 2. The whole mapping system including the ego-allocentric modulation and the cognitive map-building process is described in Section 3 and Section 4. The experimental results are demonstrated in Section 5. Finally, the conclusion is presented in Section 6.

2 Preliminaries

2.1 Pose cells model

Since the activity of place cells and head direction cells indicate the rodent's pose in the real world, a pose cell network is proposed to represent the belief of the agent's current location and orientation [22]. To simulate the activity of the

above two cells, the pose cell network uses a three-dimensional continuous attractor neural network (CANN) which is a recurrent dynamical network that is presented to model the space-related cells in neuroscience[23]. The CANN consists of a group of attractors and the connections between them, while the connections' weights are fixed and can be both excitatory and inhibitory. The activation of each attractor is continuous ranging from zero to one, which indicates the firing rate of the pose cell as well as the belief of the robot's pose in this paper, and it is operated by the self-motion cues, and the activation changes from other connected navigation cells.

The three dimensions of the pose cell network represent the plane coordinates (*x*, *y*) and the robot head facing θ, respectively. Due to the attractor dynamics of CANN, the activity in the pose cell network will converge to a single cluster, which is named the activity packet or energy packet. Thus, the attractor dynamics of the pose cell network can be represented as

$$\Delta P_{x',y',\theta'} = \sum_{i=0}^{s_{xy}} \sum_{j=0}^{s_{xy}} \sum_{k=0}^{s_{th}} P_{i,j,k} \varepsilon_{a,b,c} - \varphi + I_{ext} \tag{1}$$

where $P_{i,j,k}$ and $\Delta P_{x',y',\theta'}$ denote the activity level and the activation change of the corresponding cell, while the subscripts indicate the cell's location in the network. s_{xy} and s_{th} represent the side length of the CANN. $\varepsilon_{a,b,c}$ is the fixed connection weight matrix between pose cells which activates neighboring neurons and suppresses the distant cells, while constant φ serves as the global inhibition in the network. Because the pose cell also connects with other navigation cells, I_{ext} is introduced to represent the external excitatory input. As the internal connection weight is related to cells' relative distance, the matrix is given by

$$\begin{aligned} \varepsilon_{a,b,c} = & \exp\left(-\frac{(a^2+b^2)}{k_p^{exc}}\right)\exp\left(-\frac{c^2}{k_d^{exc}}\right) \\ & -\exp\left(-\frac{(a^2+b^2)}{k_p^{inh}}\right)\exp\left(-\frac{c^2}{k_d^{inh}}\right) \end{aligned} \tag{2}$$

where k_p and k_d are the variance constants, The inhibitory variance k^{inh} is larger than the excitatory variance k^{exc}, forming the Mexican hat function. And the indices *a*, *b* and *c* are the distances between cells in different dimensions.

The activity packet in the network can be shifted by self-motion cues during path integration which is analog to the dead reckoning in a robotic system. If the self-motion signal only has the translation velocity, the activity packet would move within the *x*-*y* plane; If the angular velocity is not zero, the packet would shift along the θ axis. The magnitude of shift depends on the translation velocity *v* and angular velocity ω, which is calculated as follows:

$$\delta x_o = \lfloor k_x v \cos\theta \rfloor, \delta x_f = k_x v \cos\theta - \delta x_o \tag{3}$$

$$\delta y_o = \lfloor k_y v \sin\theta \rfloor, \delta y_f = k_y v \sin\theta - \delta y_o \tag{4}$$

$$\delta\theta_o = \lfloor k_\theta \omega \rfloor, \delta\theta_f = k_\theta \omega - \delta\theta_o \tag{5}$$

Here, $\lfloor x \rfloor$ is a floor function. k_x, k_y, and k_θ are the path integral constants. Thus, δx_o and δx_f represent the integral and fractional part of the shift size along x-axes, respectively. The path integration process in the pose cell network can be expressed as follow:

$$\Delta P_{x',y',\theta'} = \sum_{i=\delta x_o}^{\delta x_o+1} \sum_{j=\delta y_o}^{\delta y_o+1} \sum_{k=\delta\theta_o}^{\delta\theta_o+1} \alpha_{i,j,k} P_{(x'+i)(y'+j)(\theta'+k)} \tag{6}$$

$$\alpha_{i,j,k} = g(\delta x_f, i-\delta x_o)\, g(\delta y_f, i-\delta y_o)\, g(\delta\theta_f, i-\delta\theta_o) \tag{7}$$

$$g(a,b) = \begin{cases} 1-a, & b=0 \\ a, & b=1 \end{cases} \tag{8}$$

$\alpha_{i,j,k}$ is a residual component that is spread over a $2\times2\times2$ cube to quantify the effect of the grid.

2.2 Boundary vector cells model

The firing rate of boundary vector cells is independent of the head direction, which means the receptive field of the cell is a specific distance and allocentric direction (such as south or east) in the world frame. Thus, the computational model of the i-th boundary vector cell's firing rate is as follows :

$$\delta f_i = g(r,\theta)\delta\theta_i \tag{9}$$

$$g_i(r,\theta) \propto \frac{\exp[-(r-d_i)^2/2\sigma_{rad}^2(d_i)]}{\sqrt{2\pi\sigma_{rad}^2(d_i)}} \times \frac{\exp[-(\theta-\phi_i)^2/2\sigma_{ang}^2]}{\sqrt{2\pi\sigma_{ang}^2}} \tag{10}$$

Here, (d_i,ϕ_i) is the receptive field of the cell, while (r,θ) denotes the location of boundary section in the polar coordinate, supposing the rodent is in the origin of the coordinate system. σ_{rad} and σ_{ang} are constant variances of distance and direction respectively, and δ is the impulse function. Fig.1 shows the relationship between the boundary vector cell's receptive field and the firing rate.

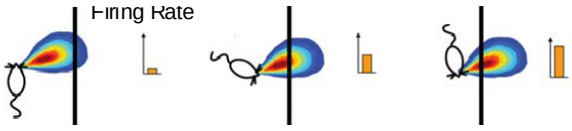


Fig. 1 The boundary vector cell model[4]

2.3 Local view cells model

The local view cells model is an array of cells that record the visual information during experiments. The cell's firing rate is related to the vision system which converts the visual input into a visual template and compares it with existing templates. The comparison results, including the absolute difference between input and database as well as the shift angle, activate the local view cell according to the index of templates.

2.4 Problem formulation

In mammals' brains, the activity patterns of those navigation cells can generate episodic memory and are served as a cognitive map. In this paper, we refer to the experience map that is used in RatSLAM[9] to construct a metric-topological map that can be utilized for robots. The map can be described as a tuple:

$$\mathrm{EM} = \langle \boldsymbol{E}, \boldsymbol{L} \rangle \tag{2}$$

where $\boldsymbol{E} = \{e_1,\ldots,e_n\}$ is a set of vertices and $\boldsymbol{L} = \{l_1,\ldots,l_n\}$ is a set of edges which denote the relative positional relationship between linked vertices. Each vertex e_i is an experience during running, which can then be defined as:

$$e_i = \{P^i, V^i, B^i, \mathbf{p}^i\} \tag{3}$$

where P^i, V^i, B^i are the state of pose cells, local view cells, and the boundary vector cells, respectively. $\mathbf{p}^i$ is the location of the experience node in the world coordinate system.

3 System Framework

To enhance the functionality and enrich the information of the map, this paper proposes a framework to combine the visual and lidar cues into the experience map. As shown in Fig.2, the framework consists of three threads for tackling different types of inputs. During the mapping process, the wheel odometry and onboard sensors record the idiothetic and environmental information and send them to the corresponding thread. Since the visual cues are transferred as images, the region of interest which is assigned manually is clipped

from the image and then downsize to calculate similarities and the offsets of head direction with existing templates using Sum of Absolute Differences (SAD). A local view cell is activated when there is a difference lower than the threshold, which means the robot sees a familiar appearance, while a new local view cell can be generated if the current appearance has never been seen. Meanwhile, the pose cell network is driven by the movement of the agent by path integration which has been described in section 2. As the lidar information is generated in a self-centered coordinate system, we propose an egocentric boundary cell model which is analog to the parietal window boundary coding (PWb) cells[25] found in the mammal's brain to tackle it. The egocentric boundary cells' firing rate is then transformed into the allocentric pattern through the ego-allocentric modulation utilizing the orientation information in the pose cell network.

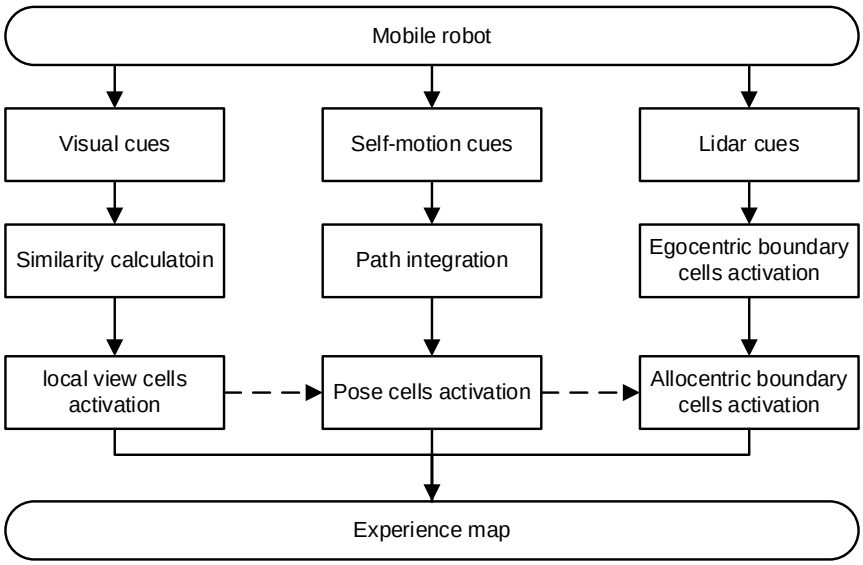


Fig. 2 The framework of the proposed mapping algorithm (The dotted line indicates the transfer of orientation information)

4 Methodology

4.1 Egocentric boundary cells model and ego-allocentric modulation

During operation, the data received from lidar at time t is an array of distances to obstacles $\boldsymbol{l}_t = \{l_{t1}, \ldots, l_{tn}\}$ while each l_{ti} corresponding to a fixed angle which takes the agent's orientation as the positive of the x-axis. Therefore, the structural information provided by LiDAR is self-centered. The egocentric boundary cells model is constructed to process it and every cell has its receptive field (ρ, θ) which responds to the boundary section within it. According to the firing pattern of boundary cells, the computational model of a cell's firing rate is described as follows [24]:

$$r_e^k = \frac{1}{\rho}\exp\left(-\left(\frac{\theta_k - \theta_b}{\sigma_s}\right)^2\right)\exp\left(-\left(\frac{\rho_k - \rho_b}{\sigma_\rho}\right)^2\right) \quad (4)$$

where (ρ_b, θ_b) is the coordinate of a boundary section in the robot coordinate system, while (ρ_k, θ_k) is the receptive field of the k-th egocentric boundary cell [25]. The firing rate of a cell is proportional to the distance and orientation of the obstacle.

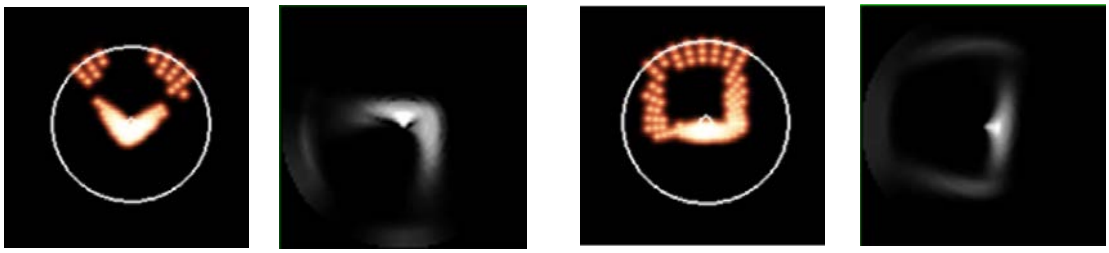
Inspired by the gain-field circuit in the retrosplenial cortex [26], head direction cells provide a transformation between egocentric and allocentric representation, which is similar to the

transformation matrix used in traditional SLAM. Instead of directly providing the representation in the world coordinate system, the information is converted into the allocentric frame. When the robot is in the mapping process, the odometer information is integrated by the pose cells. Thus, the location of the centroid of the activity package in the theta-axis denotes the integral result of all historical orientations, which can be considered as the current robot's head direction. To simulate the function of the gain-field circuit, we propose a model to realize the ego-allocentric modulation with the help of the head direction information in the pose cell network θ_{HD}. The firing rate of the allocentric boundary cell is as follows:

$$r_a^j = \sum_{i=0}^{N_{oc}} r_e^i \delta\left(\theta_i + \theta_{HD} - \theta_j\right) \tag{5}$$

where θ_i and θ_j are the angular receptive field of the i-th egocentric and the j-th allocentric cells, respectively. $\delta(x)$ is the impulse function.

Fig.3 shows the results of ego-allocentric modulation. According to the firing rate of head-oriented cells, the conversion from egocentric to allocentric firing rate is realized.



egocentric allocentric egocentric allocentric

(a) (b)

Fig. 3 Results of ego-allocentric modulation

(a) The robot is in a corner and faces northwest. (b) The robot moves towards the north.

4.2 Experience map building

In the mapping process, every vertex of the map contains the spatial information that the robot experienced at this location. Therefore, to save resources, an experienced node should be created when the perceived information has been largely changed, compared to the current node. In this paper, we use an experience similarity score S to evaluate whether a node needs to be generated. Since the appearance in some places, like a long corridors, it's inadequate to estimate the change using only visual cues. Meanwhile, when there is no node in the system that can sufficiently describe the current experience, it means either the robot is distant from the existed nodes or the appearance is changed. Thus, the states of pose cells and local view cells are applied to determine the score. The experience similarity score is calculated as follows:

$$S = \mu_V S_V + \mu_P S_P \tag{6}$$

where S_V and S_P denotes the similarity between the current states of pose cells and local view cells and the previous activity patterns preserved in the current experience node while μ_V and μ_P are the weight of corresponding cells. Considering the distance influence, the measurement of similarity of pose cells' state is defined as follows:

$$S_P = \begin{cases} 0, & d \geq 2 \\ 2 - d, & d < 2 \end{cases} \tag{7}$$

$$d = \frac{\sqrt{\left(x_{pc}^{'} - x_i^{'}\right)^2 + \left(y_{pc}^{'} - y_i^{'}\right)^2 + \left(\theta_{pc}^{'} - \theta_i^{'}\right)^2}}{r_c} \tag{8}$$

where are the coordinates of the centroid of the

largest activity packet in the pose cell network, $\left(x_i^{'}, y_i^{'}, \theta_i^{'}\right)$ is the pose cell's location which is connected with the currently activated experience node e_i, and r_c is the distance constant.

When the system receives external visual cues, the corresponding local view cells are activated, and the scoring metric of the corresponding local view cells is

$$S_V = \begin{cases} 0, & V^i \notin \boldsymbol{V}^{curr} \\ 1, & V^i \in \boldsymbol{V}^{curr} \end{cases} \tag{9}$$

where $\boldsymbol{V}^{curr}$ is a set of currently activated local view cells, and V^i is the cell that is linked to the current experience node. Thus, if the experience similarity score is less than the threshold $S_{\max}$, a new experience e_j will be created and linked to the previously activated experience node.

$$\boldsymbol{e}_j = \left\{ \mathbf{p}^i + \Delta\mathbf{p}^{ij}, P^{curr}, V_{max}^{curr}, \boldsymbol{B}^{curr} \right\} \tag{10}$$

where P^{curr} is the index of the cell with the largest activity level in the pose cell population, and V_{max}^{curr} is index of the most active local view cell. Meanwhile, $\boldsymbol{B}^{curr}$ denotes the population activity pattern of allocentric boundary cells, which indicates that the firing rates of all cells are recorded. The firing rates are originally stored in a matrix, but to demonstrate the environmental structure intuitively, we transfer the data format as an image. In this population snapshot, cells are arranged in the polar coordinates based on their receptive fields. $\Delta\mathbf{p}^{ij}$ is a vector to describe the movement between the newly created and the previous experience nodes, which also means the length of the edge l_{ij} in the experience map.

$$\Delta\mathbf{p}^{ij} = \mathbf{p}^j - \mathbf{p}^i = \begin{pmatrix} x_j \\ y_j \\ \theta_j \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \\ \theta_i \end{pmatrix} = \begin{pmatrix} \Delta x_{ij} \\ \Delta y_{ij} \\ \Delta\theta_{ij} \end{pmatrix} \tag{11}$$

If the robot revisits a place along an exit path, information in the edge is updated by averaging the odometry data:

$$\Delta\mathbf{p}_{new}^{ij} = A \cdot \Delta\mathbf{p}_{old}^{ij} - B \cdot \Delta\mathbf{p}_{curr}^{ij} \tag{12}$$

$$\boldsymbol{A} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & \Delta s\cos\Delta\theta & -\Delta s\sin\Delta\theta \\ 0 & \Delta s\sin\Delta\theta & \Delta s\cos\Delta\theta \end{pmatrix} \boldsymbol{B} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{22}$$

$$\Delta\theta = \frac{1}{2}\left[\tan^{-1}\left(\frac{\Delta y_{curr}^{ij}}{\Delta x_{curr}^{ij}} \right) - \tan^{-1}\left(\frac{\Delta y_{old}^{ij}}{\Delta x_{old}^{ij}} \right) \right] \tag{13}$$

$$\Delta s = \left(d_{curr}^{ij} + d_{old}^{ij} \right) \Big/ \left(2 d_{old}^{ij} \right) \tag{14}$$

4.3 Map correction

In the experience map, each experience node not only receives links from other experiences but also sends out links to one or more nodes. Therefore, the correction of the pose of each node needs to integrate the information of all the connected experiences. Thus, the implementation of correction is as follows

$$\Delta\mathbf{p}^i = \alpha \left[\sum_{j=1}^{N_f} \left(\mathbf{p}^j - \mathbf{p}^i - \Delta\mathbf{p}^{ij} \right) + \sum_{k=1}^{N_t} \left(\mathbf{p}^k - \mathbf{p}^i - \Delta\mathbf{p}^{ki} \right) \right] \tag{15}$$

where α is the correction rate constant, N_f is the number of links from experience $\boldsymbol{e}_i$ to other experiences, and N_t is the number of links from other experiences to experience $\boldsymbol{e}_i$. The correction of the map is iteratively applied during the operation, making the position of nodes in the map gradually approach a distribution that minimizes the average error of the trajectory. As a result, links between nodes also need to be updated to align with the change in orientation

$$\Delta \mathbf{p}_{new}^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Delta\theta_i & -\sin\Delta\theta_i \\ 0 & \sin\Delta\theta_i & \cos\Delta\theta_i \end{pmatrix} \cdot \Delta \mathbf{p}_{curr}^{ij} \quad (16)$$

5 Experiments

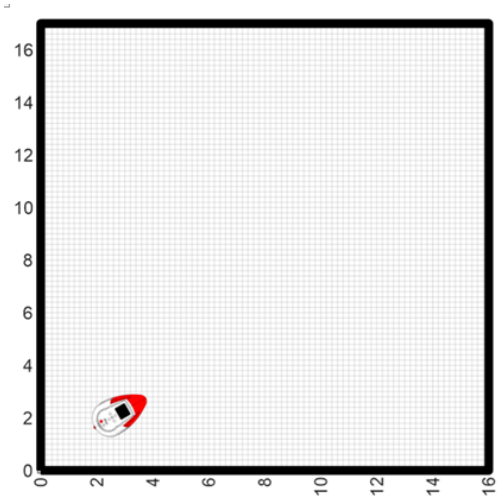
5.1 Experimental settings and criteria

The dataset was generated using MATLAB, simulating a static two-dimensional indoor environment of 16m×17m. The output data of the camera, LiDAR, and odometry sensors are simulated. The frequency of all sensors is 10Hz. Camera data cannot be obtained directly in the simulated two-dimensional environment. In this paper, the environment is divided into 0.2m×0.2m grids, and the robot obtains the output of different visual templates in different grids. The motion of the robot follows the mode of walking against the wall, and its velocity follows the distribution: $N(1,0.05)$ (m/s). When the robot runs in a straight line along the wall, the velocity follows the distribution: $N(1,0.01)$ (rad/s). When the robot reach the corner, the velocity follows the distribution: $0.0982N(1,0.01)$ (rad/s). The LiDAR scanning rate is 400 points per turn, the angle increment is fixed to 0.9°, and the additive noise on distance measurement follows the distribution:

$$noise_{\text{laser}} \sim \begin{cases} N(0,0.005), d \le 1.5 \\ N(0,0.015), d > 1.5 \end{cases} \quad (27)$$

During the operation, the robot moves along the same path, and the sensor data is released almost at the same time in the experiment, without considering the asynchronous problem.

The generated test environment and data format are shown in Fig.4.




```
path:       mapping.bag
version:    2.0
duration:   3:37s (217s)
start:      Jun 25 2021 20:34:16.50 (1645792456.50)
end:        Jun 25 2021 20:37:53.78 (1645792673.78)
size:       15.2 MB
messages:   8692
compression: none [20/20 chunks]
types:      nav_msgs/Odometry        [cd5e73d190d741a2f92e81eda573aca7]
            ratslam_ros/EgoCues      [595417e279b603926045805e4631ff91]
            ratslam_ros/ViewTemplates [b4b1d645c31606cb3d41fe361c493814]
            rosgraph_msgs/Log        [acffd30cd6b6de30f120938c17c593fb]
            std_msgs/Int32           [da5909fbe378aeaf85e547e830cc1bb7]
topics:     /egocues     2171 msgs   : ratslam_ros/EgoCues
            /odom        2171 msgs   : nav_msgs/Odometry
            /rosout         4 msgs   : rosgraph_msgs/Log
            /rosout_agg     4 msgs   : rosgraph_msgs/Log
            /seq_num     2171 msgs   : std_msgs/Int32
            /vtid        2171 msgs   : ratslam_ros/ViewTemplates
```

Fig. 4 The test environment: (left) the visualization of the experimental environment; (right) the information of the dataset (the topics of egocues, odom, and vtid submitter to the LiDAR, odometry and camera outputs)

Since the coordinates of the built map usually is not consistent with the given ground truth coordinate system, a transformation matrix $\boldsymbol{S}$ is implied to align them. Therefore, the absolute trajectory error (ATE) for the i-th frame is defined as follows:

$$\boldsymbol{F}_i = \boldsymbol{Q}_i^{-1}\boldsymbol{S}\boldsymbol{P}_i \quad (28)$$

where $\boldsymbol{P}_i$ represents the pose of the robot of the

i-th frame estimated by the algorithm while $\boldsymbol{Q}_i$ represents the ground truth pose of the robot. However, when evaluating the performance of an algorithm, the root means square error (RMSE) is usually used to deal with ATE:

$$\mathrm{RMSE}\left(\boldsymbol{F}_{1:n},\Delta\right)=\left(\frac{1}{m}\sum_{i=1}^{m}\left\|\mathrm{trans}\left(\boldsymbol{F}_i\right)\right\|^2\right)^{\frac{1}{2}} \tag{29}$$

where Δ represents the time interval and trans($\boldsymbol{F}_i$) represents the translational part of the absolute trajectory error.

5.2 Results and discussion

Fig.5 shows the comparison of the experience map built by the proposed method, the result of path integration in the pose cell network, and the dead reckoning trajectory based on the raw odometry information. The dead reckoning is suffered from sensor noise which leads to a large drift. Thus, it cannot form a consistent trajectory, and the inconsistency will affect robot localization and task performance. Since the path integration of self-motion information is conducted in pose cells which receive the calibration cues from local view cells, it can restore the real trajectory, but there are sharp changes in some positions, as shown in Fig.5 A, when a closed loop is detected. Among them, the experience map achieves the best performance in consistency, which effectively correct the drift caused by sensor noise. As the rotational movement leads to a fast change in neurons' states, more experience nodes are created when the angular velocity increases. In Fig.5 B, the density of the experience node in the up-left corner is much higher than in other places.

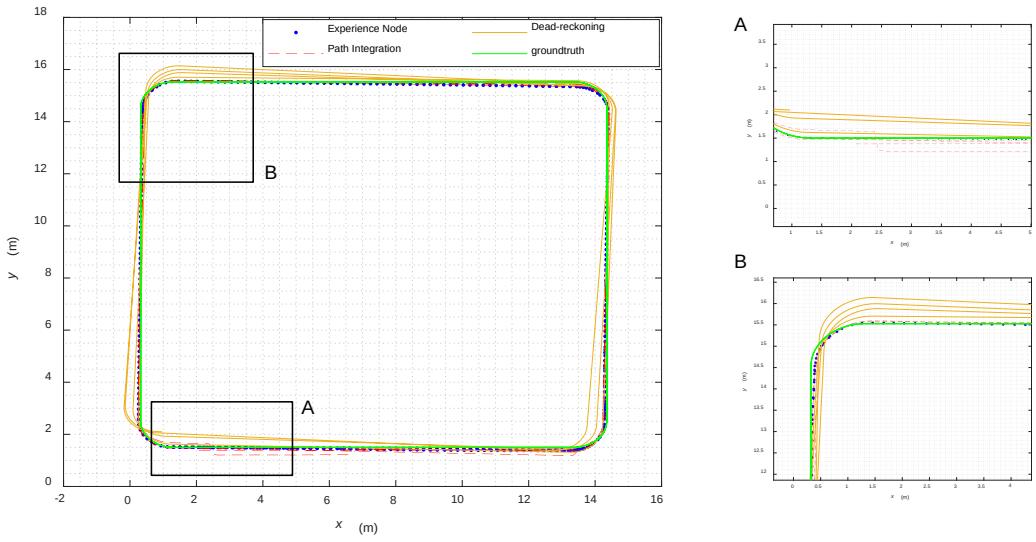


Fig. 5 Comparison of trajectory

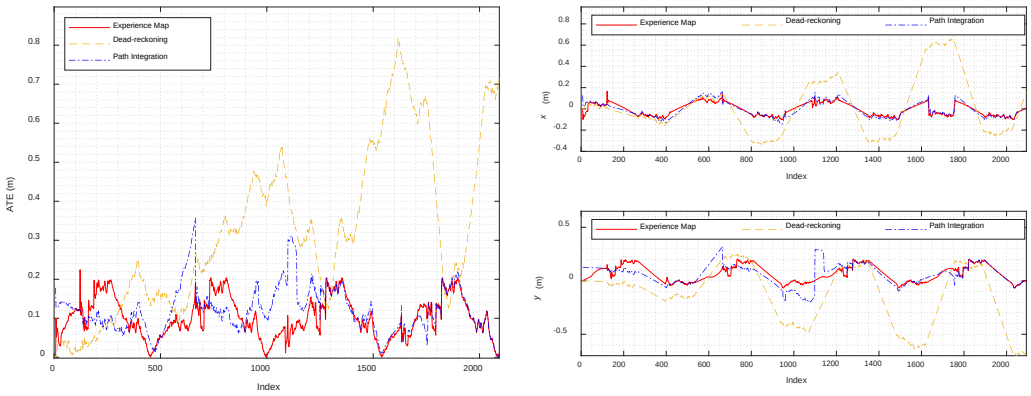


Fig. 6 Comparison of absolute trajectory error; The index corresponds to the frame number of the camera. Since the frequency is fixed, the index is linearly related to the running time

Fig.6 shows the comparison of the absolute trajectory errors of the three algorithms to construct the map. In the left column, as the cumulative displacement of time increases, the dead reckoning error also accumulates. While the path integration result has the largest error at first, with the robot's repetitive visits, the performance is close to the experience map which realizes the loop-closure and calibrates the trajectory. Therefore, the error of the proposed method shows a periodic change and converges to a size of about 0.1m. Meanwhile, in the right column, the errors of the trajectories are calculated from the x and y directions, respectively. The errors of dead reckoning share the same pattern as before, while the errors of the path integration and the experience map fluctuate around a value. Additionally, in the y-axis direction, the path integration result shows two large jumps in the interval of (600, 800) and (1000, 2000), which are due to the loop correction illustrated in Fig.5 B, while the amplitude of the experience map is relatively stable.

Table 1 shows the statistical data of absolute trajectory error of experience map, path integration, dead reckoning, and ground truth. It can be seen that the statistical results of the proposed method are better than the others in all three indicators. The span of the data set used in the manuscript is about 4 minutes. If the method runs for a longer period, in the case of repeating the same trajectory, the odometer drift error, will generate a new experience node. When the existing experience node is matched again, the displacement error position introduced by the new experience node can be alleviated by the graph relaxation algorithm. This is due to loop-closure testing will be detected from the current experience node movement is triggered after a certain distance, by adjusting the parameters match experience node density and test environment after repeated several times track experience node has been very intense, under the condition of not introducing additional noise, the algorithm will no longer cause a new experience, so repeating the same motion can guarantee the same trajectory. Therefore, the proposed method will still have a good ability for map construction for a long time.

Table 1 Comparison of absolute trajectory error

	RMSE(m)	Mean(m)	Median(m)
Dead-reckoning	0.4067	0.3372	0.3073
Path integration	0.1340	0.1201	0.1201
Experience map	0.1142	0.1015	0.0935

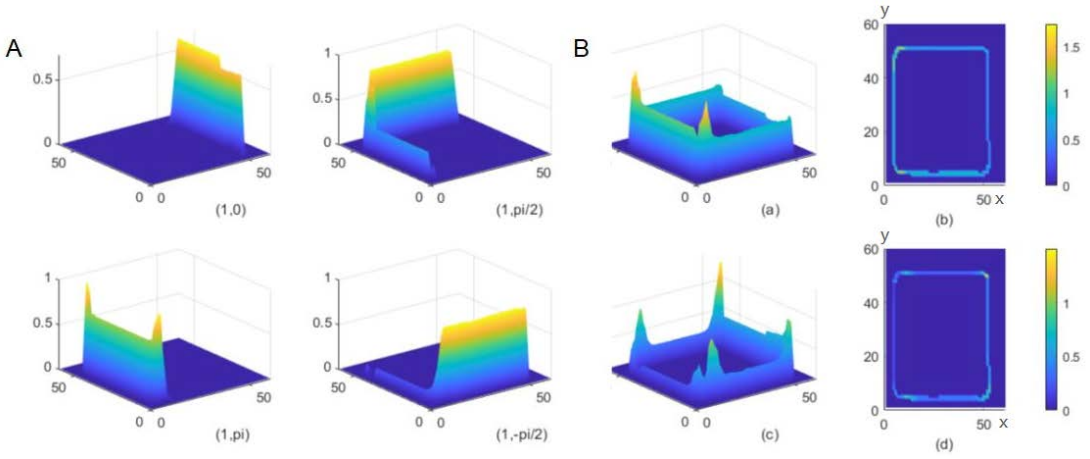


Fig. 7 The firing rate map of allocentric boundary cells in the mapping process

Fig.7 shows the firing rate map of allocentric boundary cells in the environment during the entire robot operation process. In Fig. 7, the x-y plane is corresponding to the scaled coordinate system of the experience map and the z-axis is the firing rate of the cell with

$$x_{firingmap} = x_{experience} * 0.3 \,(\mathrm{m}) \tag{30}$$

$$y_{firingmap} = y_{experience} * 0.3 + 5 \,(\mathrm{m}) \tag{31}$$

Its left part shows the firing rate maps of cells with different receptive fields and the right column is the summation of cells' firing rates. The upper one combines four cells whose receptive fields are: $(1,0)$, $(1,\pi/2)$, $(1,\pi)$, and $(1,-\pi/2)$, while its lower part demonstrates the sum of cells with their receptive fields of $(4,0)$、$(4,\pi/2)$、$(4,\pi)$ and $(4,-\pi/2)$. The centers of the receptive fields of the corresponding cells shown in Fig.7 A are (1, 0), (1, π/2), (1, π) and (1, -π/2), respectively. For a single orientation-independent boundary cell, it is activated only when there is a barrier in the receptive field. Because the robot is set to move along the wall, only one cell in this figure is activated in most cases. And since the shape of the receptive field is a Gaussian distribution, the firing rate is possible to be nonzero in the corner when the robot is close to the wall in this test. In Fig.7B, the upper part shows the population firing rate map by superposing the four figures shown in Fig.7A, while the lower part is similar to it but the radial receptive fields of the cell are changed to be 4. Considering that the distance between the trajectory of the robot and the boundary is ranging from 0.2m to 2m, the cells with a radial receptive field of 1m will obtain higher firing rates, so spatial relationships between detected obstacles and the robot could be indicated from the activation of cells. Thus, according to these two population firing rate maps, we can see the probability of the barrier's existence at 1 or 4 meters away from the robot's current location. In the experiment, multiple boundary cells with different receptive fields are used, so a more accurate understanding and description of the environment structure could be obtained, which enhances the information on the map.

6 Conclusions

In this paper, aiming at the insufficient features in the existing map when working in a visually ambiguous environment, an allocentric boundary cells model is proposed in this paper, which is integrated into the experience map. Allocentric boundary cells generate the orientation-independent structural information through the head direction modulation, and their population activities are fused into the experience map which improves the

functionality of the map. Simulation shows that the proposed model can effectively reflect the boundary structure in the environment, and indicates the obstacle information on the map. At the same time, the algorithm can build a consistent representation of the environment with sensor noise and achieves a root mean square error of 11.42 cm in an indoor environment, which effectively corrects the drift and ensures the accuracy of the map. In the future, it is considered to build the connection between pose cells and every single boundary cell, and explore a sparser expression that makes it possible to spontaneously recover the firing rate pattern in the boundary cells population when activating the experience node, without storing a large number of connection weights. Meanwhile, it is intended to apply the experience map to path planning, so that the robot can judge the traversability of the path according to the boundary cell's state.

7 References

[1] J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat," Brain Research, vol. 34, no. 1, pp. 171–175, Nov. 1971.

[2] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," Nature, vol. 436, no. 7052, pp. 801–806, Aug. 2005.

[3] J. Taube, R. Muller, and J. Ranck, "Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations," J. Neurosci., vol. 10, no. 2, pp. 436–447, Feb. 1990.

[4] C. Lever, S. Burton, A. Jeewajee, J. O'Keefe, and N. Burgess, "Boundary Vector Cells in the Subiculum of the Hippocampal Formation," Journal of Neuroscience, vol. 29, no. 31, pp. 9771–9777, Aug. 2009.

[5] A. Arleo and W. Gerstner, "Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity," Biological cybernetics, vol. 83, no. 3, pp. 287–299, 2000.

[6] M. Yuan, B. Tian, V. A. Shim, H. Tang, and H. Li, "An Entorhinal-Hippocampal Model for Simultaneous Cognitive Map Building," p. 7.

[7] Y. Zhou and D. Wu, Spatial Representation and Location Estimation Model Based on Place Cell, Journal of Shanghai Jiao Tong University, no. 04 vo 52, pp. 488–494, 2018.

[8] H. Tang, W. Huang, A. Narayanamoorthy, and R. Yan, "Cognitive memory and mapping in a brain-like system for robotic navigation," Neural Networks, vol. 87, pp. 27–37, 2017.

[9] D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford, "OpenRatSLAM: an open source brain-based SLAM system," Autonomous Robots, vol. 34, no. 3, pp. 149–176, 2013.

[10] M. Milford, A. Jacobson, Z. Chen, and G. Wyeth, "RatSLAM: Using Models of Rodent Hippocampus for Robot Navigation and Beyond," in Robotics Research, Isrr, 2016, vol. 114, pp. 467–485.

[11] M. J. Milford and G. F. Wyeth, "Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System," Ieee Transactions on Robotics, vol. 24, no. 5, pp. 1038–1053, 2008.

[12] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, May 2010, pp. 3507–3512.

[13] S. Yu, J. Wu, H. Xu, R. Sun, and L. Sun, "Robustness Improvement of Visual Templates Matching Based on Frequency-Tuned Model in RatSLAM," Front. Neurorobot., vol. 14, p. 568091, Sep. 2020.

[14] T. Xu, Y. Ling, and M. Chen, A bio-inspired algorithm integrated with DGSOM neural network. CAAI Transactions on Intelligent Systems, vol. 12, no. 03, pp. 405–412, 2017.

[15] Zeng T, Si B. Cognitive Mapping Based on Conjunctive Representations of Space and Movement[J]. Frontiers in Neurorobotics, 2017, 11: 61.

[16] Gupta S, Davidson J, Levine S, et al. Cognitive

mapping and planning for visual navigation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2616-2625.

[17] Salman, M., & Pearson, M.J. (2018). Whisker-RatSLAM Applied to 6D Object Identification and Spatial Localisation. *Living Machines*.

[18] Chen, M., & Hu, W. (2019). Research on BatSLAM Algorithm for UAV Based on Audio Perceptual Hash Closed-Loop Detection. *Int. J. Pattern Recognit. Artif. Intell., 33*, 1959002:1-1959002:19.

[19] J. Steckel and H. Peremans, “BatSLAM: Simultaneous Localization and Mapping Using Biomimetic Sonar,” Plos One, vol. 8, no. 1, p. e54076, 2013.

[20]Z. Wu, Y. Yue, M. Wen, J. Zhang, J. Yi and D. Wang, "Infrastructure-Free Hierarchical Mobile Robot Global Localization in Repetitive Environments," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021.

[21] O. Struckmeier, K. Tiwari, M. Salman, M. J. Pearson, and V. Kyrki, “ViTa-SLAM: A Bio-inspired Visuo-Tactile SLAM for Navigation while Interacting with Aliased Environments,” in 2019 IEEE International Conference on Cyborg and Bionic Systems (CBS), 2019, pp. 97–103.

[22] M. J. Milford, G. F. Wyeth, and D. Prasser, “RatSLAM: A hippocampal model for simultaneous localization and mapping,” in 2004 Ieee International Conference on Robotics and Automation, Vols 1- 5, Proceedings, 2004, pp. 403–408.

[23] A. Samsonovich and B. Mcnaughton, “Path Integration and Cognitive Mapping in a Continuous Attractor Neural Network Model,” The Journal of neuroscience : the official journal of the Society for Neuroscience, vol. 17, pp. 5900–20, Sep. 1997.

[24] Bicanski A, Burgess N. A neural-level model of spatial memory and imagery[J]. eLife, 2018, 7: e33752.

[25] A. Bicanski and N. Burgess, “A neural-level model of spatial memory and imagery,” eLife, vol. 7, p. e33752, Sep. 2018.

[26] A. Bicanski and N. Burgess, “Environmental Anchoring of Head Direction in a Computational Model of Retrosplenial Cortex,” J Neurosci, vol. 36, no. 46, pp. 11601–11618, Nov. 2016.

Authors

Yixuan Long has obtained her Master's Degree in College of Information and Communication Engineering, Harbin Engineering University. Her research interest is Simultaneous Localization and Mapping.

Hao Wang is pursuing his Master's Degree in College of Information and Communication Engineering, Harbin Engineering University. His research interest is Simultaneous Localization and Mapping.

LI Yibing received his Ph.D. Degree from Harbin Engineering University in 2003. He is a professor in the School of Information Engineering, Harbin Engineering University. His research interests include communication signal processing, radio navigation and positioning.

YE Fang received her Ph.D. Degree from Harbin Engineering University in 2006. She is currently working in the College of Information and Communication Engineering, Harbin Engineering University. Her research interests include cognitive confrontation and intelligent decision-making.

Qian Sun received his Ph.D. Degree from the School of Automation, Harbin Engineering University in 2016, and is currently a tutor of Master's Degree at the School of Information and Communication Engineering, Harbin Engineering University. His research interest is focused on collaborative localization.

Journal of Global Positioning Systems (*2022*)

Vol. 18, No.12: 84-85

High precision positioning for pedestrian navigation in dense urban environments

Luo Huan, Email: neu_lh@163.com

Supervisors: **Dr. Wu Chen**

University: **Hong Kong Polytechnic University**

Defense Date: **08.04.2022**

Abstract

Due to reflections or blockages of GNSS satellite signals by buildings and infrastructures, urban positioning with GNSS is a great challenge. Traditional receiver autonomous integrity monitoring (RAIM) based methods are insufficient to obtain positioning solutions with high accuracy in urban canyons where the majority of satellite signals may be contaminated by multipath interference and non-line-of-sight (NLOS) reception. This thesis will focus on the positioning performance improvement for pedestrians using lowcost devices in urban canyons. With the help of three-dimension (3D) city models, GNSS positioning performance can be improved by predicting visibility or path delays of satellite signals. Shadow matching is a 3D-mapping-aided (3DMA) approach utilizing SNR of satellite signals, which is available in position (NMEA format data) and measurement (raw GNSS measurement data) domains with a wide range of applications. However, the performance of shadow matching will be degraded when it fails to distinguish the grids of neighboring streets, or when it is affected by dynamic interference or 3D model errors. A new weighting method, grid weight smoothing and clustering (GWSC) method, is proposed to improve the performance of grid identification, and experiments in Hong Kong streets showed that the newly proposed method improved the cross-street accuracy of shadow matching from 19.4m to 2.1m with a large improvement rate (IR) of 89.2%, significantly outperforming the weighted average (WA)-based method of 15.3m accuracy, which had an IR of 21.1%. NLOS correction-based approach is another 3DMA method to improve overall positioning performance by simulating reflected paths of satellite signals in the measurement domain. Diffraction, which is also a type of NLOS, is not considered in the conventional reflection model. In this study, we apply the radio signal diffraction models to develop an improved NLOS correction-based approach using a more comprehensive reflection model considering more types of NLOS, along with the GWSC method in weighting. Through experiments, the estimated delays were consistent with the received errors, where over 95% of signals showed estimated errors below 15m. Moreover, the improved approach achieved accuracy of 7.4-15m in static tests, and 11.9m in kinematic tests, compared with up to 86.4m by the conventional GNSS method in typical Hong Kong streets. The proposed method showed a significant IR between 62.7% and 89.7% of positioning accuracy in all experiments in urban canyons. The computation load with city 3D models is very high as it needs to consider different satellite constellations at different time. Moreover, many cities may not have public 3D models available. In this study, a novel approach, named multi-epoch offset searching (MEOS), which does not need 3D city models, is proposed to mitigate multipath effects. With the implementation of measurement smoothing and the GWSC method, the new approach can provide high-precision positioning solutions for pedestrians in urban canyons. It is showed that the

approach achieved accuracy of within 9m and 15m in several static and kinematic tests, respectively, compared to the poor accuracy, up to 57.7m and 27.5m, from raw GNSS outputs from conventional low-cost GNSS devices. The proposed method has significant IRs up to 88.3% in static tests, while its IR reached 47.6% in kinematic tests. To make the positioning system more stable and robust, multiple techniques are integrated with sensors existing in the smartphones. The integration of GNSS-based approaches and pedestrian dead reckoning (PDR) technology improves the positioning availability and further reduces the positioning errors. Further integrating Bluetooth-low-energy (BLE) into the system makes the positioning system more flexible and effective. Owing to the proposal of BLE-based heading estimation and improvement of step detection, this integration system achieved a high accuracy of within 5m in outdoor and seamless areas.

Keywords: GNSS, PDR, Pedestrian Navigation

Journal of Global Positioning Systems (*2022*)
Vol. 18, No.12: 86-87

Research on the Underwater Vehicle Navigation Based on Bayesian Filter

Huimin Liu, Email: upcliuhm@foxmail.com
Supervisor: Prof. **Zhenjie Wang**
University: China University of Petroleum (East China)
Defense date: June, 2020

Abstract

The underwater multi-sensor integrated navigation technology provides guarantees for the long-term and large-scale execution of underwater vehicle diving missions. In multi-source navigation information fusion, observational models and navigation sensor noise are spatially and temporally complex. It is of great theoretical interest and practical value for constructing accurate functional and stochastic models. In this paper, we focus on the fusion of multi-source navigation information for underwater vehicles and work on high-precision long-baseline acoustic system filter design, nonlinear filtering, colored noise filtering, integrated navigation fault-tolerant design, and multi-vehicle coordinated navigation. The work and results of this study are as follows.

The paper is divided into seven chapters and is structured as follows:

Chapter 1 introduces the research status of common commercial underwater navigation sensors at home and abroad, the development status of filtering theory under the framework of Bayesian filtering and the research status of underwater integrated navigation filtering at home and abroad, introduces the research content and technical route of the paper, and gives the chapter division of the paper.

Chapter 2 presents the filtering principle for underwater multi-sensor integrated navigation. Definitions and transformation relations have been studied for commonly used underwater integrated navigation systems, Strapdown inertial navigation systems, Doppler logs, acoustic USBL systems, LBL systems, and pressure sensor and measurement error models. In this paper, we introduce the underwater navigation sensor noise analysis method and its mathematical foundations in the framework of Bayesian filtering.

In Chapter 3, sound velocity estimation and sound velocity error correction methods in LBL acoustic localization are investigated. Considering the delayed nature of acoustic propagation, we study the acoustic delay of LBL systems and analyze the PDOP problem in acoustic localization. The underwater carrier navigation filter algorithm is designed based on the LBL/INS loose-binding and tight-binding modes, and the simulation experiments are designed to validate it.

Section 4 presents the application of the Bayesian filtering algorithm to nonlinear systems

corresponding to the specific operational context of underwater integrated navigation. In this paper, we introduce SINS fast alignment problem under large misalignment angle, USBL/DVL integrated navigation under depth constraint, CKF based compact combination algorithm for SINS/USBL, and localization problem for nonlinear ranging equations when acoustic ranging is short, which are validated by simulations and experiments.

In Section 5, the effect of systematic errors on the integrated navigation filter is analyzed and colored noise treatments including augmented state methods, measurement difference methods, and time series analysis methods are introduced. A hierarchical adaptive information filter is proposed to separate and estimate the system noise parameters. Combining simulation experiments of USBL/DVL navigation and localization in environments with unknown ocean currents with actual measurement experiments, the proposed modified adaptive filtering method was analyzed and validated to be effective.

In Chapter 6, fault detection and identification methods for underwater integrated navigation are studied, and the SINS/USBL/DVL/PG sensor refinement model is developed based on Bayesian filtering theory and observed features of underwater navigation sensors. Based on the interactive multi-model filtering principle, we propose an adaptive federated interactive filtering design scheme for deep-sea towing systems.

Chapter 7 presents the principles and methods of collaborative navigation filtering for underwater multi-vehicle applications, including decentralized and centralized modes. Based on the Bayesian filtering principle, the delay value is estimated using the maximum A-posteriori. Motivated by the existence of bidirectional ranging or mutual localization observations in multi-carrier formation networks, and considering the possibility of anomalous pilots, we propose a centralized filtering algorithm for underwater multi-carrier formation based on a weight-selective filtering model. Simulation experiments are designed to validate the results.

The summary and outlook sections mainly summarize the main research content of this paper, point out the limitations of the current research, plan the next research work, and give an outlook for future research work.

Key words: underwater vehicle, Bayesian filter, integrated navigation, nonlinear filtering, colored noise, fault tolerant filter, collaborative navigation

Journal of Global Positioning Systems (*2022*)

Vol. 18, No.12: 88-89

Research on LiDAR/INS/ODO/GNSS vehicle integrated navigation algorithm based on graph optimization

Le Chang, Email: changlesgg@whu.edu.cn

Supervisors: Prof. **Xiaoji Niu**, Assoc. Prof. **Tisheng Zhang**

Institution: Integrated & Intelligent Navigation (i2Nav) Group, GNSS Research Center, Wuhan University

Graduation Date: Dec 2021

Abstract

With the rapid development of digital earth and smart city, the demand for localization-based services is becoming urgent. However, continuous, accurate, and reliable positioning navigation in complex environments is a common key technical issue that need to be solved. While GNSS positioning deteriorates or even fails in urban canyons; the positioning error of low-cost INS quickly diverges over time; LiDAR has poor positioning availability when environmental features are insufficient. In order to improve the positioning and navigation service capabilities, China plan to build a more ubiquitous, more integrated, and smarter national comprehensive PNT (Positioning, Navigation, and Timing) system in 2035. And the multi-sensor information fusion is one of the key components of the comprehensive PNT.

Since the insufficient vertical resolution of the low-beam LiDAR causes the degradation of LiDAR odometry in some environments, we propose a feature point-based probability map matching method, which combines the advantages of matching by feature point with a probability map. The process extracts the ground feature points and the non-ground feature points by segmentation. A probability map with different resolutions will be constructed to deal with those features, respectively, with a higher vertical resolution for the ground feature and a higher horizontal resolution for the non-ground part. Scan matching by a probability map constructed by feature points minimizes the dependence on the line and surface features in the environment. It has been compared with the well-known open-source LiDAR odometry, i.e. Cartographer and LeGO-LOAM. Evaluations were carried out in different feature scenes. In the areas with rich line and surface features, the positioning accuracy of the proposed method is better than Cartographer, primarily the positioning result on the elevation and horizontal attitude. In areas lacking line and surface features or with the ramped ground, the positioning error of LeGO-LOAM is larger than the proposed method, and it even crashed in some challenging scenarios.

Since incorrect exterior parameters of sensors cause inaccurate navigation results, a method for IMU/ODO and LiDAR/IMU calibration by utilizing ODO and IMU pre-integration are proposed:

(1) A method of IMU/ODO calibration by utilizing ODO pre-integration is proposed. It is based on the graph optimization theory and not sensitive to the

initial value of the exterior parameters. There is no need to estimate the state of the vehicle. It has the capability of lever arm estimation. The simulation and field tests show that the error of the lever arm is less than 5 cm, and the error of mounting angles is less than 0.1°. Since the ODO measurement value by the calibration result is equivalent to centimeter-level mileage increment, the calibration method can meet the requirements of centimeter-level positioning.

(2) A method of IMU pre-integration is proposed to eliminate the motion distortion of LiDAR, which not relies on GNSS information, and avoid the double alignment process in common calibration methods. In addition, the integral of IMU mitigates the impact of IMU noise. Simulation and real tests show that the lever-arm estimation error is about 1 cm, and the mounting angle estimation error is about 0.1°.

In order to improve the accuracy and usability of vehicle positioning in complex environments, a multi-source fusion algorithm based on graph optimization is designed and implemented. Based on the LiDAR odometry, this algorithm adopts IMU/ODO pre-integration constraints to enhance the positioning stability in environments with insufficient features. And the global pose assistance by a pre-built probability map matching ensures positioning accuracy when GNSS fails. The marginalization and the sliding window are employed to remove the historical parameters while keeping the efficiency of the optimizer. A fully functional LiDAR/INS/ODO/GNSS vehicle navigation software - *LIOGNS*, has been developed for algorithm validation. The performance of three navigation modes were tested and analyzed through multiple datasets, including positioning by a pre-built map, GNSS/INS/LiDAR fusion, and GNSS/INS/ODO/LiDAR fusion.

The mode of pre-built map matching of *LIOGNS* is tested in the international competition (JD Digital Technology Global Explorer Competition, ‘Autonomous Driving Map Optimization and Sensor Fusion’ track, 2018). The multi-source data fusion algorithm was utilized to optimize the point cloud mapping and the positioning mode with a pre-built map. **It achieved 5cm position accuracy and 0.1° attitude accuracy, and won the championship of the global finals**. The GNSS/INS/LiDAR-SLAM integrated positioning performance of *LIOGNS* was tested and compared with Cartographer and LIO-SAM, through simulated GNSS interruption and real frequent GNSS occlusion in campus environment. Since the Cartographer was designed suitable for low dynamic vehicles, its roll, pitch, and elevation errors are relatively large. Limited by the IMU noise modeling, the horizontal attitude and elevation error of LIO-SAM are equivalent to Cartographer, and the horizontal position accuracy is better than Cartographer. **The position and attitude accuracy of the proposed *LIOGNS* is superior than LIO-SAM and Cartographer, especially in the environment where the GNSS signal is weak or blocked.** The accuracy and usability of the odometer assistance were tested based on simulated GNSS interruptions and in a tunnel that lack of environmental features. Insufficient features in the tunnel lead to significant positioning drift along the longitudinal direction. However, **the odometer limits such drift effectively and significantly enhancing the reliability when both GNSS and LiDAR fail**.

In summary, this thesis proposed a LiDAR/INS/ODO/GNSS vehicle integrated navigation algorithm based on graph optimization, and completed the core algorithm design and software implementation. Through the datasets of open-sky area with simulated GNSS interruption, GNSS frequent failure environment and tunnel scenes, the proposed algorithms were thoroughly tested and analyzed. The scheme can meet the requirements of continuous, accurate and reliable positioning and navigation in complex environments for autonomous driving and mobile robot applications.

Key words: LiDAR SLAM, GNSS/INS, Wheel Odometer, Integrated Navigation, Preintegration, Graph Optimization, Exterior Parameter Calibration

Journal of Global Positioning Systems (*2022*)

Vol. 18, No.12: 90-91

Attitude estimation methods using low-cost GNSS and MEMS MARG sensors and their integration

Wei Ding, Email: wei.ding1@ucalgary.ca

Supervisors: Prof. Yang Gao

Institution: University of Calgary

Graduation Date: August 15, 2022

Abstract

For low-cost magnetic, angular rate, and gravity (MARG) sensors based on the microelectromechanical system (MEMS) technology, the sensor errors and measurement noises are significantly large. Attitude errors by integrating gyro data accumulate rapidly. When the vehicle is quasi-static, the roll and pitch angles can be determined by accelerometer measurements which use the local gravity as the reference. The magnetometer is resorted to generate heading information by measuring the geomagnetic field. However, the accelerometer and magnetometer measurements can be deteriorated by the vehicle maneuver and ambient artificial magnetic disturbances, respectively.

Thereby a quaternion-based error state Kalman filter (ESKF) is developed to fuse the MEMS MARG sensor measurements for accuracy improved attitude estimation. The error state vector constitutes attitude error and gyro bias variation. the gyro-measured angular rates are used to continuously propagate the vehicle's three-dimensional attitude quaternion in its sampling rate, whilst accelerometer and magnetometer measurements are employed for the state correction. Disturbances such as external accelerations and magnetic anomalies are excluded, and the measurement noise covariance matrix is adaptively adjusted according to the innovations.

Global navigation satellite system (GNSS) based attitude estimation shows time-independent error characteristics. The pitch and heading angles can be determined using a single GNSS antenna based on the time differenced carrier phase (TDCP) observations or derived from a moving baseline formed between two firmly mounted GNSS antennas. The major challenges of the former include cycle slips, carrier phase discontinuity, and slow vehicular velocity which should be excluded from attitude estimation. Whereas the integer ambiguity resolution is indispensable for the latter, the baseline length constrained least-squares ambiguity decorrelation adjustment (C-LAMBDA) method can be applied.

The GNSS/MARG sensors integrated attitude estimation methods are investigated to exploit the complementary merits of the high precision of MARG sensor during the short period and the performance stability of GNSS over the long term. The ESKF developed for the MARG sensor is extended to utilize the GNSS-derived heading and pitch angles for additional measurement updates. The solution

continuity is guaranteed by the MARG sensor alone during the periods when the GNSS-derived attitude angles are unavailable.

Key words: Attitude estimation; MARG sensor; GNSS; C-LAMBDA; Data fusion; Error state Kalman filter

Volume 18, No. 1, 2022

Journal of Global Positioning Systems

Published by

International Association of Chinese Professionals in Global Positioning Systems (CPGPS)

ISSN 1446-3156 (Print version)

ISSN 1446-3164 (CD-ROM Version)