Detecting outliers in local ionospheric model for GNSS precise positioning

Tam Dao[1] · Ken Harima[2] · Brett Carter[1] · Julie Currie[1] · Simon McClusky[2] · Rupert Brown[3] · Eldar Rubinov[3] · John Barassi[2] · Suelynn Choy[1]

Abstract

Global Navigation Satellite System fast precise positioning can be achieved with accurate ionospheric corrections computed from an adequate number of GNSS stations in a local region. In low-latitude regions, the presence of electron density gradients over short distances can lead to outliers in the map of ionospheric corrections and decrease its accuracy. In this study, we explored outlier detection in ionospheric correction mapping through statistical residuals during a four-month test in 2021. Our findings indicate that the residuals of the local ionospheric model conform to the Laplace distribution. To determine outliers, we use an empirical rule for the Laplace distribution, setting thresholds at $\mu \pm 3b$, $\mu \pm 3.5b$, and $\mu \pm 5.8b$ for data retention rates of 95%, 97%, and 99.7%, respectively. Here, μ represents the location parameter, which corresponds to the median of data, and b is the scale parameter, calculated as the medium absolute deviation. We found that while removing outliers can improve model accuracy, it may result in unavailable prediction due to a lack of data in a spare network. For example, applying a $\mu \pm 3.5b$ threshold for outlier removal led to approximately 2.5% of recording time having no ionospheric corrections map in low-latitude regions, however, the local model has the potential to improve its mean accuracy by up to 50% for both low and mid-latitudes. Therefore, choosing the appropriate percentile threshold depends on the network configuration and the desired accuracy. Removing erroneous satellite data to improve ionospheric accuracy brings positive impacts on precise positioning.

Introduction

The ionosphere plays a significant role in affecting the propagation of radio signals by causing refraction and delay, thereby limiting the precision of Global Navigation Satellite Systems (GNSS) positioning (Jakowski et al. 2008; Warnant et al. 2009; Grewal et al. 2020). To enable high precision GNSS, it is vital to correct for the effect of the ionosphere. One widely adopted method is the utilization of dual-frequency GNSS measurements. By simultaneously receiving GNSS signals from two different frequencies, it becomes possible to estimate and compensate for the ionospheric delay, leading to improved precision in GNSS positioning and navigation applications. However, for single frequency users, the ionospheric correction is typically derived using ionospheric models and algorithms. These models incorporate data from multiple GNSS stations, either on a global or regional scale, to estimate and predict the ionospheric delay at specific locations and times. Although not as accurate as dual-frequency measurements, these corrections still provide valuable information for mitigating ionospheric effects and enhancing positioning accuracy for single frequency GNSS users. Without ionospheric correction, the ionospheric delay can cause positioning errors of about five meters or more (Subirana et al. 2013).

Global ionospheric models (GIMs) are maps that provide estimates of the ionospheric delays for any given location on

✉ Tam Dao
hanhtamdao@gmail.com

1   SPACE Research Centre, School of Science, Royal Melbourne Institute of Technology (RMIT) University, Melbourne 3000, Australia

2   Geoscience Australia, Canberra, ACT 2609, Australia

3   FrontierSI, Melbourne 3000, Australia

 Springer

the Earth's surface. GIMs are useful for precise positioning applications that require ionospheric correction over a wide area. Generating a GIM involves collecting GNSS data from a global network of stations, which is then used to estimate the total electron content (TEC) of the ionosphere. Subsequently, the TEC values are interpolated to create an electron density map (Yang et al. 2021) which provides estimates of the ionospheric delay (or corrections) for any given location on the Earth's surface.

Regional or local ionospheric models can also be generated to support GNSS precise positioning applications that require ionospheric corrections over a smaller area. These models provide a higher level of spatial resolution compared to the GIMs, thereby improving the accuracy of ionospheric correction for local applications. Local ionospheric models enable fast and accurate precise point positioning (PPP) (Li et al. 2011, 2022; Banville et al. 2014; Dao et al. 2022). The accuracy of ionospheric delay estimates obtained from local models relies on the number and spatial density of GNSS receivers in the network. Regional or local modelling can be achieved through various techniques, including the function fitting methods such as polynomial functions or grid-based approaches (Gao and Liu 2002; Tao and Jan 2016; Banville et al. 2022).

The accuracy of the ionospheric corrections in the low latitude region presents significant variability, primarily due to the highly dynamic nature of the ionosphere. The rapid changes in electron density in a short distance may cause large differences in ionospheric delays through the network (Datta-Barua et al. 2010; Rungraengwajiake et al. 2015; Budtho et al. 2018). This could bring potential errors in the ionospheric correction model (Parkinson and Spilker 1996; Teunissen et al. 2013; Rovira-Garcia et al. 2020). A robust Kalman filter can be applied to detect, identify, and adapt measured errors at individual stations (Zhang et al. 2015; Lotfy et al. 2022). Our study primarily focuses on the detection of outliers within a bilinear model through all stations in the local network.

Various techniques are available for outlier detection. The most common methods involve utilizing statistical parameters such as standard deviation (SD), median absolute deviation (MAD), and interquartile range (IQR) (Yang et al. 2019). The SD threshold technique is well-suited for data that follow a normal distribution, while MAD is more applicable for heavy-tailed distributions such as the Laplace distribution (Lu and Chang 2022). While the empirical rule in the normal distribution is commonly used in practical statistics, the corresponding rule for the Laplace distribution, which is more representative of the ionospheric errors (Marini-Pereira et al. 2020) and Ma et al. (2022), has not been previously discussed. We here present a new method for outlier detection of ionospheric correction in local models utilizing an empirical calculation to establish thresholds

for residuals. Our goal is to enhance the accuracy of ionospheric models by detecting and removing outliers that may affect their reliability and robustness.

Methodology

Observational data

To assess the accuracy of the ionospheric models in low-latitude regions in Australia, we investigated GNSS data of Northern Territory (NT) network. In addition, for comparative analysis, we examined a similar configuration in mid-latitude regions located in Western Australia (WA). Figure 1 shows a map of the GNSS stations installed across Australia, with the two areas investigated in this study defined by the red boxes. The ionospheric delays were computed from GPS data only using the Ginan software.

Ginan software

Ginan is a GNSS analysis centre software developed by Geoscience Australia aimed at providing a range of real-time corrections to enhance GNSS positioning accuracy. It is a critical component of the Australian Government's National Positioning Infrastructure Capability (NPIC). The Ginan software toolkit and service offer a range of correction products that enable PPP for users and industries across Australia. Ginan is regularly updated with new features and capabilities. This is an open-source software which is
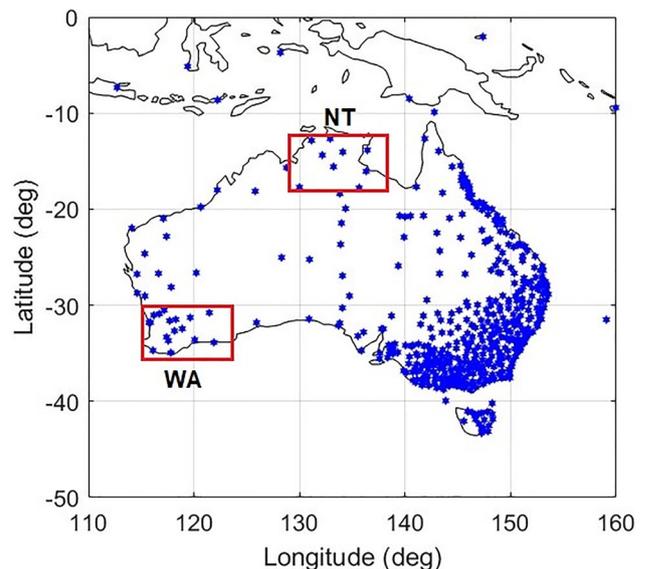

Fig. 1 Map of GNSS receivers (blue stars) over Australia, the red rectangles are locations of two local network investigated in this study

accessible through the Geoscience Australia's GitHub page (Ginan 2023).

We computed the ionospheric delay for each satellite and receiver path using the PPP solution in the Ginan processing software which has been covered in the paper Dao et al. (2022). The stochastic model used to calculate the slant TEC is briefly presented in Table 1. The ionospheric module computes the ionospheric delay (as the slant total electron content—STEC) for each receiver-satellite path for the chosen input stations. Epoch intervals were set at 30 s, and the thin shell ionosphere height was set at 350 km above the Earth's surface. The output ionospheric delay (slant TEC or STEC) was then used for mapping the regional ionospheric model. The unit of the ionospheric delay is TECU, where one TECU approximately corresponds to a delay of 16 cm on the GPS L1 signal frequency.

Ionospheric delay outputs from Ginan

For further analysis, we selected the STEC values with a variance below 0.3 TECU^2, corresponding to an error of 0.5 TECU. Figure 2 (left) shows the STEC and its variance obtained from all stations within the NT network on July 01, 2021. To remove the receiver hardware bias

in each measurement, the between-satellite single difference ionospheric delay (SD STEC) was computed for each epoch. This involves subtracting the measurement from each visible satellite with respect to a chosen reference satellite. The reference satellite is the most commonly visible satellite within the local region, if more than one satellite occurs with the highest common, we select the 1st one in order. Similarly, the single difference of STEC variance was obtained by aggregating the variance from each visible satellite with respect to the reference satellite. The SD STEC corresponds with each satellite-receiver path, and contains the different satellite biases of that satellite and the reference satellite.

Figure 2 (right) displays the SD STEC and its corresponding variance, which aligns with the data shown in the left figure. They are used for creating the map across the local network to facilitate further analysis. It is important to note that the calculation of SD STEC relies on the choice of the reference satellite. Accordingly, as the reference satellite changes, the values of the SD STEC also change, resulting in discontinuities within the data (top left).

The processing steps involved in local ionospheric corrections mapping are depicted in the flowchart in Fig. 3.

Table 1 The stochastic model used to calculate the slant TEC in Ginan

Variable/measurement	Model	Sigma (a-priori)	Process noise
Code measurements	Random	0.3 m/sin^2(el)	
Phase measurements	Random	0.003 m/sin^2(el)	
Station position	Constant	100 m	
Station clock	Random Walk	1000 m	100 m/sec
Ambiguities	Constant	1000 cycles	
Troposphere	Random Walk	0.1 m (stating from models)	10^{-4} m/sec
Troposphere gradient	Random Walk	0.001 m	10^{-6} m/sec
Slant TEC	Random Walk	200 TECU	0.1 TECU/sec

Fig. 2 Ionospheric delays (STEC) and its variance output from the Ginan software (left panels), single difference STEC (SD STEC) and its variance calculated for each satellite at each epoch (right panels) on July 01, 2021
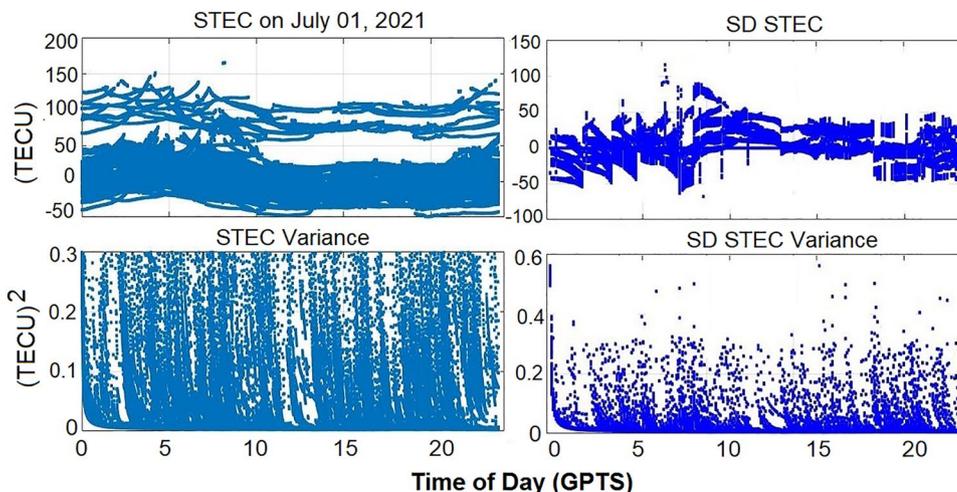
First, the STEC output for all stations was computed by the Ginan software. Subsequently, the SD STEC of each measurement in the network was calculated to eliminate the receiver hardware bias. The SD STEC was then used to construct the local ionospheric corrections map and compute residuals for outlier detection and identification. In this stage, we applied statistical thresholds to identify and remove outliers from the data, resulting in a refined dataset that was used to generate the local ionospheric map, providing estimated corrections for users. The final step involved evaluating the accuracy of the local mapping approach.

Ionospheric correction model using the bilinear regression function.

We retrieve the ionospheric bilinear regression model for a local network as follow:

The distance matrix (H_{SD}) is built from n station coordinates in the network that link to the satellite, with (lat, lon) is the latitude and longitude of each station,

$$H_{SD} = \begin{bmatrix} 1 & lat_{sta1} & lon_{sta1} \\ 1 & lat_{sta2} & lon_{sta2} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & lat_{sta\_n} & lon_{sta\_n} \end{bmatrix} \tag{1}$$

Using least-squares estimation, the linear interpolation coefficients (ionospheric delay, latitude, longitude) for the bilinear interpolation at each epoch are then computed as:

$$\begin{bmatrix} I_0 \\ I_{lat} \\ I_{lon} \end{bmatrix} = \left(H_{SD}^T Q_{SD} H_{SD} \right)^{-1} H_{SD}^T Q_{SD} I_{SD} \tag{2}$$

where the weight matrix (Q_{SD}) is calculated from SD STEC variance:

$$Q_{SD} = diag(var_{SD})^{-1} \tag{3}$$

The ionospheric delay (I_{SD}) from the network is used to build the model:

$$I_{SD} = \begin{bmatrix} I_{SD,sta1} \\ I_{SD,sta2} \\ \dots \\ \dots \\ I_{SD,sta\_n} \end{bmatrix} = \begin{bmatrix} 1 & lat_{sta1} & lon_{sta1} \\ 1 & lat_{sta2} & lon_{sta2} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & lat_{sta\_n} & lon_{sta\_n} \end{bmatrix} \begin{bmatrix} I_0 \\ I_{lat} \\ I_{lon} \end{bmatrix} \tag{4}$$

Using the bilinear regression function mapping, the SD STEC at any location can be predicted as:

$$I_{pre_{SD}} = \begin{bmatrix} 1 & lat_{pre} & lon_{pre} \end{bmatrix} \begin{bmatrix} I_0 \\ I_{lat} \\ I_{lon} \end{bmatrix} \tag{6}$$

The model residual is computed by the difference of SD STEC from model and observed from stations in the network used to build the model:

$$res = I_{pre_{SD}} - I_{obs_{SD}} \tag{5}$$

The model accuracy is computed by the difference of SD STEC from model and observed from the testing stations that not used to build the model:

$$acc = |I_{pre_{SD}} - I_{test_{SD}}| \tag{7}$$

Statistical study of modelling residual for outlier detection

Modelling residuals

The SD STEC obtained from the local NT network is utilised. Figure 4 (left) shows the distribution of the GNSS station network in the NT region, where the red stars represent the stations utilised to construct the ionospheric map. The Darwin (DARW) station, indicated by a blue star, is not included in generation of the map. However, it can be

Fig. 3 Flow chart for the data processing and outlier detection in local mapping

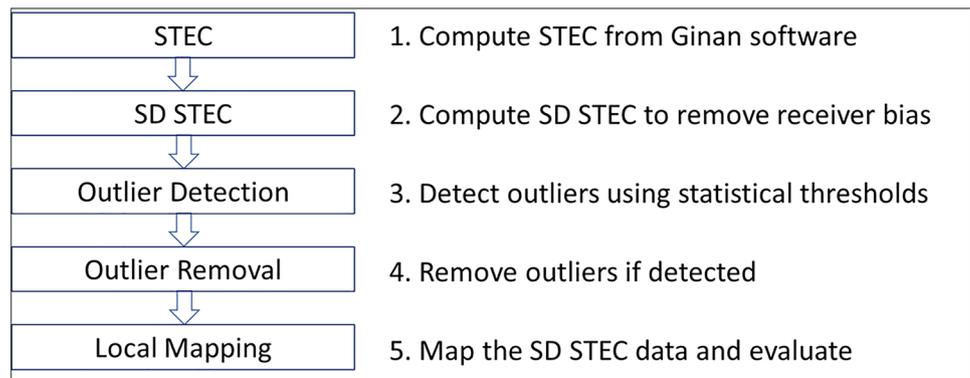| STEC | 1. Compute STEC from Ginan software |
SD STEC	2. Compute SD STEC to remove receiver bias
Outlier Detection	3. Detect outliers using statistical thresholds
Outlier Removal	4. Remove outliers if detected
Local Mapping	5. Map the SD STEC data and evaluate

Fig. 4 Location of the testing station Darwin (DARW) and the remaining stations (red stars) in the local map (left), the illustration of ionospheric mapping for a satellite S_i (right)
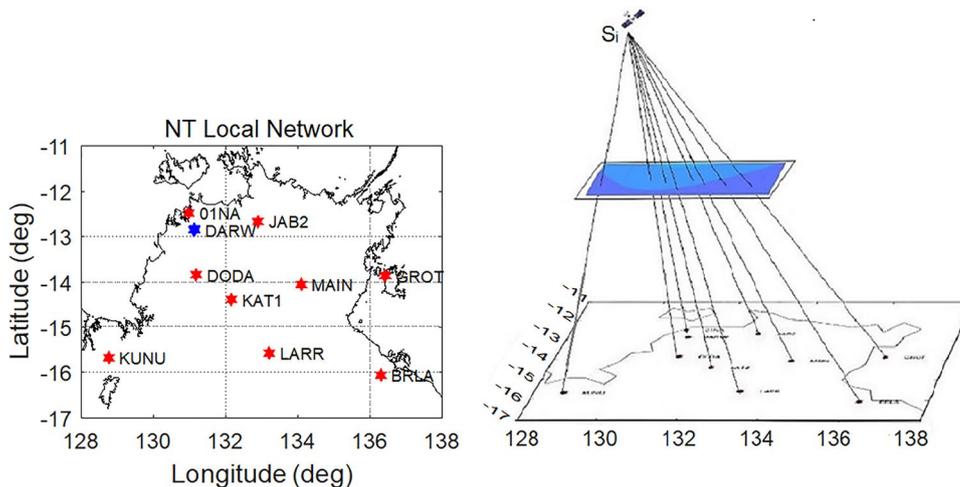
utilized later to evaluate the map's performance. On the other hand, Fig. 4 (right) provides an illustration for a local map of the ith satellite over the NT region. For each satellite, a local map is modelled at every epoch using a single layer ionospheric model mapped at a height of 350 km above the Earth's surface.

In this examination, we constructed a local ionospheric map of a satellite using data from a minimum of five stations automatically. Therefore, in situations where the number of stations receiving data from a satellite falls below five, no ionospheric map is available for that satellite. This results in an "unavailable satellite" or "unavailable local map". If there are no local ionospheric correction maps available during a specific time, there will be no corrections provided to users.

The local model runs over the area for a 5-min interval, enabling faster processing while maintaining appropriate adaptation to the changes in the ionosphere. To detect outliers, we first create the local map from the observational ionospheric delays from the network stations. Then, we calculate the difference between the map and the observed data at all stations within the network. This is the residuals of the model. By analysing the distribution of these residuals, an appropriate threshold for identifying outliers can be determined.

In this study, we collected the residuals data statistically in four months: January, April, July, and October in 2021. These months were characterized by quiet conditions, free from major geomagnetic storms or disturbances. Figure 5 illustrates the distribution of the residuals observed in the local ionospheric model. Notably, the residuals exhibit a Laplace distribution which is represented by the probability density function (red lines). The Laplace distribution is a double exponential distribution, centred in the middle and has fatter tails compared to the normal distribution.
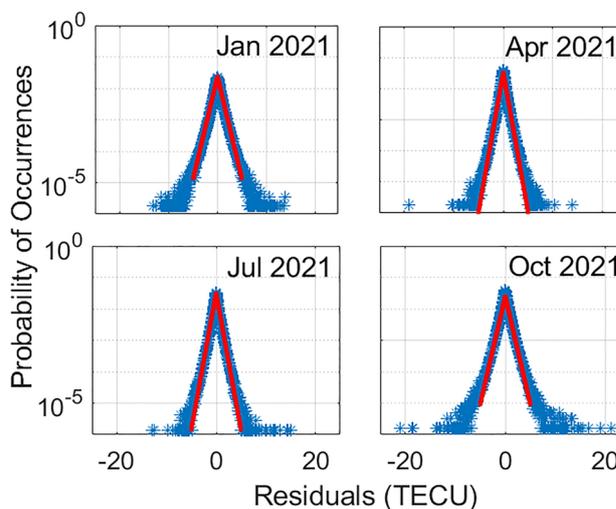
Empirical rule for Laplace distribution.

Fig. 5 Distributions of the ionospheric residuals in logarithm scale processed in four months in 2021. The red lines are the probability density function of Laplace distribution within -5 to 5 TECU of residuals

The empirical rule, also known as the "68–95-99.7" rule or the "three-sigma" rule, is a statistical guideline for a normal distribution. This rule was originally introduced and discovered by Abraham de Moivre in 1733 (Glen 2023). It is predicted that 68% of observations will fall within the first standard deviation ($\mu \pm \sigma$); 95% within the first two standard deviations ($\mu \pm 2\sigma$); and 99.7% within the first three standard deviations ($\mu \pm 3\sigma$).

The Laplace distribution has heavier tails than the normal distribution. Therefore, the empirical rule does not directly apply to the Laplace distribution. We calculate the percentile function of Laplace distribution as follows.

The function of the Laplace distribution is:

$$f(x|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|x - \mu|}{b}\right) \qquad (8)$$

where μ is the location parameter which is the median of data; b is scale parameter, sometimes referred to as the "diversity". In this distribution, b is the medium absolute deviation (MAD), which is calculated as:

$$b = MAD = median\left(|x_i - median(x)|\right) \qquad (9)$$

The percentile function of Laplace distribution is: $\mu + bln(2F)$ for $F \leq \frac{1}{2}$.

Percentile function $2F = 5\%$: the quantile is $bln(0.05) = \mu + 3b$
Percentile function $2F = 3\%$: the quantile is $bln(0.03) = \mu + 3.5b$
Percentile function $2F = 0.3\%$: the quantile is $bln(0.003) = \mu + 5.8b$

Based on this calculation, the thresholds of $\mu \pm 3b$, $\mu \pm 3.5b$, and $\mu \pm 5.8b$ are used to remove outliers and ensure that at least 95%, 97%, or 99.7% of the data falls within the thresholds, respectively. For brevity, we will refer to these thresholds as 3b, 3.5b, and 5.8b.

In practical terms, the ionosphere exhibits daily variations, with electron density levels typically higher during the day and lower at night. To account for these variations, we collected data for each hour throughout the day to ensure comprehensive coverage. Figure 6 shows the distribution of residuals for each hour over 30 days in January 2021. This process was repeated for the remaining three months in 2021, representing the four seasons, in the region. The

scale parameter, represented as b, of residuals varies over time. As shown in Fig. 6, the shapes of the distributions change throughout the day, with varying b values observed in each subfigure. The vertical dash lines in red, cyan, and black represent the corresponding thresholds of 3b, 3.5b and 5.8b, respectively. For further analysis, we selected the 3.5b threshold, for the retention of approximately 97% of the data, to construct the ionospheric models. By selecting this threshold, we attempted to strike a balance between retaining a substantial amount of data for the ionospheric model yet mitigating the impact of outliers on the model accuracy.

We computed the hourly variation of the scale b and median μ for the different seasons (January, April, July, and November) in 2021, shown in Fig. 7. Outlier detection was performed based on the changes of μ and b. Moreover, the statistical results in this method can be used to set up a relative threshold for detecting outliners in NT network during similar low solar activity periods.

Local model assessment

Accuracy and availability of regional model

We applied the statistical data of μ and b for each month to detect outliers. In Fig. 4, the DARW station represented by a blue star was intentionally excluded in the mapping process. This station was reserved for evaluating the accuracy of the local model both before and after removal of outliers. Figure 8 shows the accuracy of the local model on July 01, 2021, comparing the original data with the data

Fig. 6 The distribution of residuals for each hour in January 2021 which shows the changing of distribution sharp over time. The red lines are the probability density function of Laplace distribution for each hour in GPST (GPST: GPS time is based on atomic clocks and has no correction with respect to proper time variations of the Earth's rotation. In 2021, UTs = GPST + 18 s). The time leap between GPST and UTs in 2021 is 18 s. The Local Time in this area is LTs = UTs + 9:30
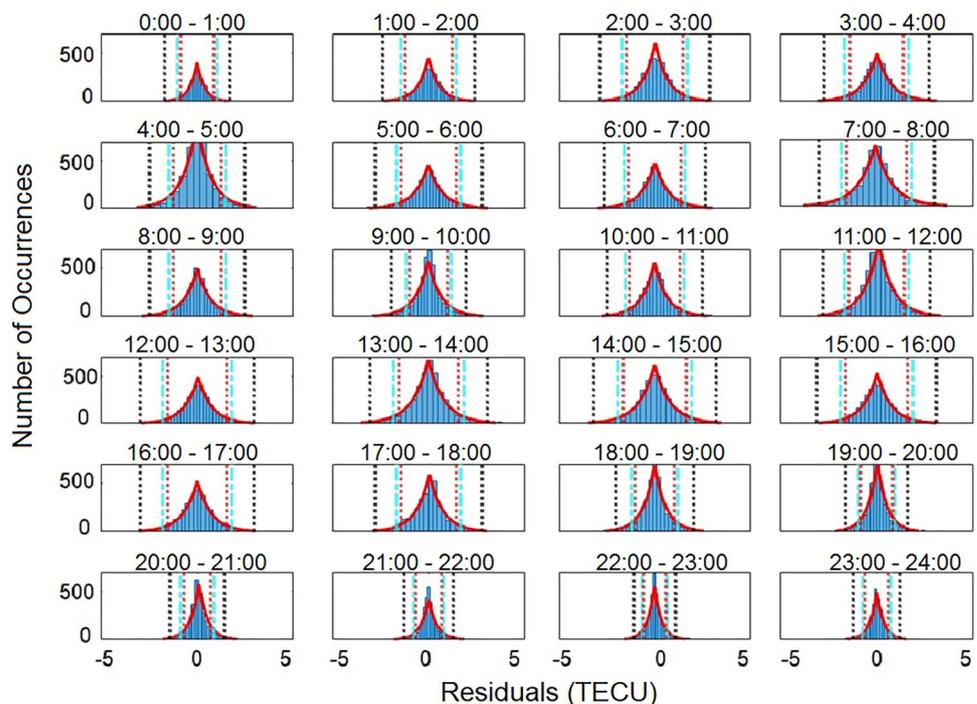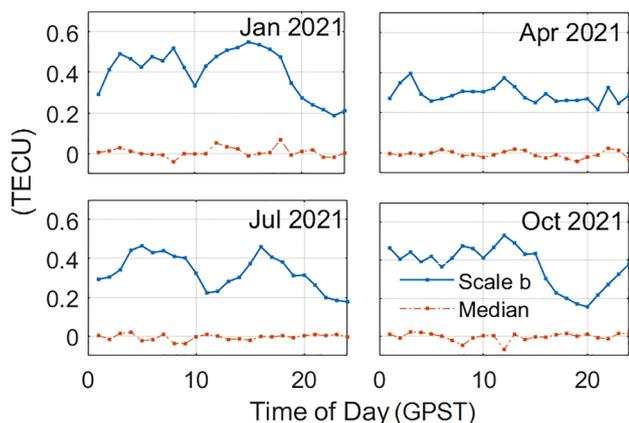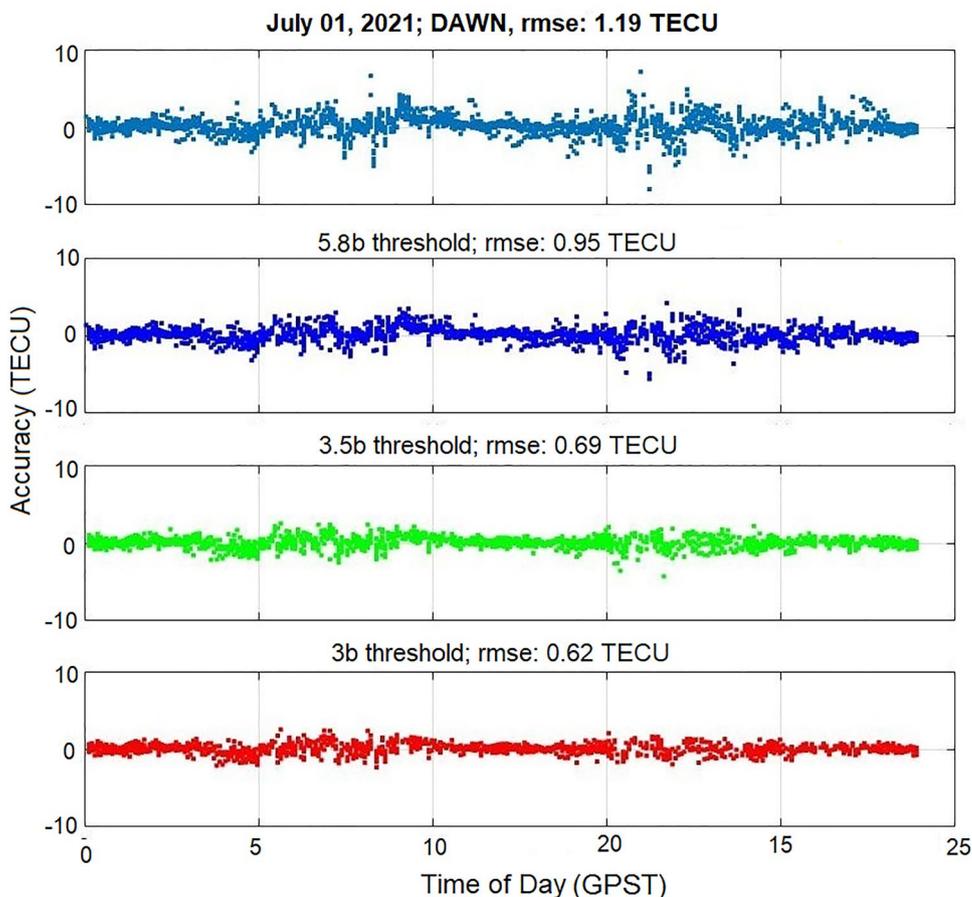
Fig. 7 The 1-h variations of median μ and scale b taken for 30 days of four months (Jan, Apr, Jul, and Nov) in 2021 mapping in NT network

obtained after applying the designated thresholds. Prior to outlier removal, the root mean square error (RMSE) of the original data was 1.19 TECU. However, after applying 5.8b, 3.5b and 3b thresholds to remove outliers, the model error improved to 0.95, 0.69, and 0.62 TECU, respectively. Therefore, selecting the right percentile threshold is based on the network configuration and desired accuracy that we aim to achieve.

Removing outliers using the thresholds of 3b and 3.5b brings an improved accuracy to the model. However, this may also decrease the number of available satellites or available local ionospheric correction maps at each epoch. It is noted that each station in the NT GNSS stations network can receive signals up to 13 GPS satellites at the same time, and one satellite can be chosen as a reference satellite to compute the SD STEC between satellites. Therefore, at any given time, the NT network can create up to 12 local maps for SD STEC using GPS data.

To calculate the number of available satellites at a given epoch, we need to count the number of satellites for which at least five stations in the NT network received the signals. The process is as follow:
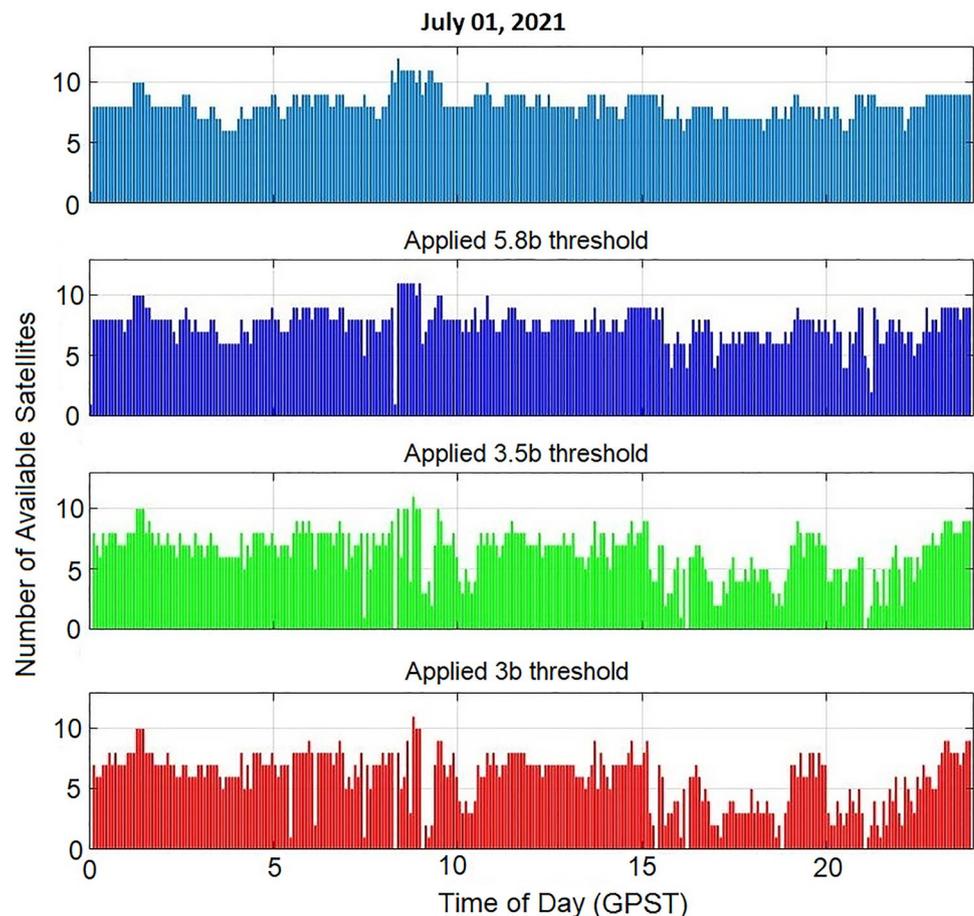
- For each satellite, count the number of stations that have received its signal at the given epoch.
- If the count is at least 5, consider the satellite as an available satellite.
- Repeat the above steps for all satellites in the network.

Fig. 8 Accuracy of ionospheric delays before (first panel) and after applying the thresholds of 5.8b, 3.5b and 3b (three last panels). The accuracy of the testing stations DAWN was measured and showed as rmse in each case

- Count the total number of available satellites at a given epoch.

Figure 9 illustrates the number of satellites that are available for mapping at each epoch during the day for the original data and after applying the 5.8b, 3.5b, and 3b thresholds. The y-axis represents the number of available satellites, ranging from 0 to 12.

As presented, utilizing the percentile function to eliminate outliers can enhance the accuracy of the model. However, this approach may lead to a reduction in available satellites or local ionospheric correction maps, especially in networks with few receivers and in active ionospheric regions like the NT network. Our data indicates that filtering out unreliable or erroneous satellite data can improve positioning accuracy for GNSS measurements. For example, in the NT network on July 1, 2021, applying a 3.5b threshold resulted in a 0.2% reduction in observations, but increased ionospheric correction accuracy by 42–0.69% TECU. Applying a 3b threshold led to a 0.4% reduction, but increased accuracy by 52% to about 0.62 TECU. Therefore, removing a small number of satellite data to enhance ionospheric accuracy has a positive impact on precise positioning.

We calculated the percentage of available ionospheric maps for each epoch of four months and displayed in Fig. 10. The x-axis represents the number of available maps per epoch, and the y-axis displays the corresponding percentage. In the NT network, the ionospheric delays from 1 to 13 satellites per epoch are observed, with the most common value centered around 9 (indicated by blue markers). Initially, the percentage of epochs with only 1–3 available satellites is nearly 0%. However, when a threshold of 3.5b is applied, the number of available satellites decreases. The green marker line illustrates that, after applying the threshold, the maximum number of available satellites is 6–7, with a percentage exceeding 20%. The percentages of epochs with 0–3 available satellites are all below 5%, but higher than the original values. Notably, around 2.5% of epochs exhibit no available satellites, indicating the percentage of cases where no corrections are available using this local network.

Accuracy of the local model in mid and low latitudes

We evaluate accuracy of the local maps by using testing stations within the network, i.e., by removing them from the network prior to performing the mapping, the same as
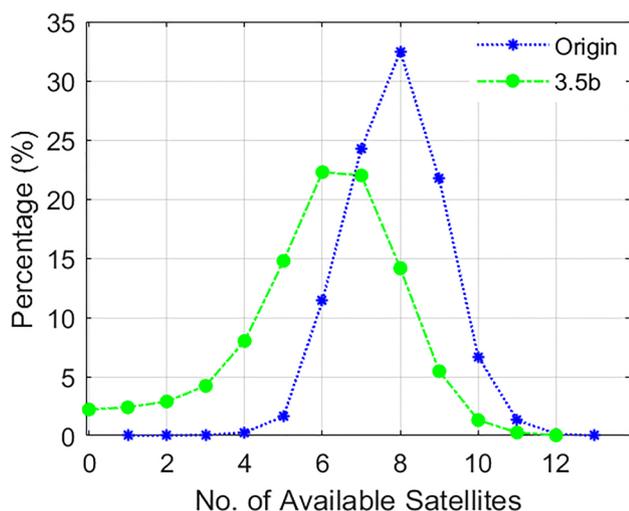
Fig. 9 The number of maps availability at all epochs during the entire day of July 01, 2021. The blue, green, and red colors present the number of available satellites from the network, after applied 5.8b, 3.5b, and 3b thresholds, respectively

Fig. 10 The percentage of available maps at each epoch before and after applying the 3.5b threshold for 4 months in 2021. Applying an outlier detection threshold (green points) decreases the number of available satellites compared with the original data (blue points)

with the Darwin station in the case above. The accuracy is calculated by finding the difference of SD STEC between the prediction from the map and the observation of the testing station. In this study, we used five testing stations located in different areas of the network. We also compared the results over NT with another network in Western Australia (WA). The WA local map has a similar network configuration to NT, but it is in the middle latitudes (30–35° S). Figure 11 displays the maps of both NT and WA networks, with the red stars indicating the receiver locations and color circles indicating the testing stations that were alternatively selected for evaluation.

Figure 12 shows the accuracy of the local model for five testing stations in the NT network after adapting with outlier detection. The results present different accuracy based on the location of the test station. For instance, when the testing station is KUNU, the local map is then created using the remaining stations. Since KUNU has few neighbouring stations, the model's accuracy for this test station is low. On the other hand, GROT is located outside the local map, but two nearby stations contribute to a higher accuracy model compared to KUNU. For stations within the network, the accuracy is better, ranging from 0.5 to 1 TECU in the NT region.

In contrast to the NT region, the local map performs more effectively in the mid-latitude region (WA), as shown by the results in Fig. 13. The accuracy of the model for stations located within and outside of the network shows insignificant differences. In the mid-latitude region, the model shows an accuracy of approximately 0.5 TECU for all test stations.
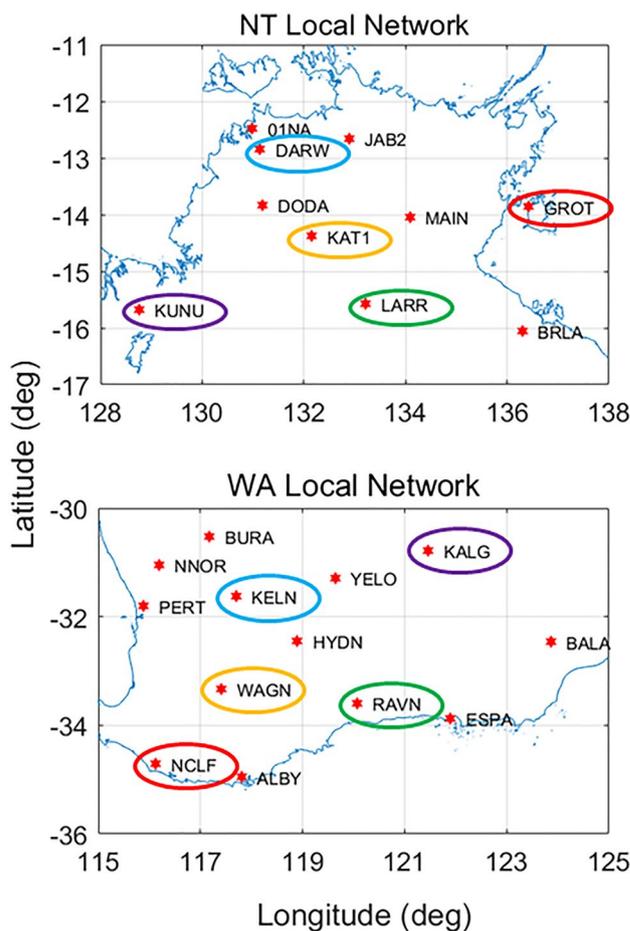

Fig. 11 Maps of testing stations in NT and WA networks. The red stars are receiver coordinates with their names next by. The name within the circle is the testing station used to evaluate local model

To examine model accuracy before and after applying the 3.5b threshold for outlier detection, the mean accuracy of two testing stations DARW (for NT network) and RAVN (for WA network) have been selected for all data processing periods. The accuracy improvement is calculated by:

$$\%improvement = 100\% \frac{\left| Accuracy_{after} - Accuracy_{before} \right|}{Accuracy_{before}}$$

(10)

Figure 14 presents the number of days corresponding to the percentage that improved mean accuracy. As shown in Fig. 14 (left), the accuracy of the NT local model improves by up to 60%. For most days, the improvement is between 20 to 55%. On the other hand, the WA local network shows up to 50% improvement.

The results of this study demonstrate that the presented outlier detection methods can be applied to other latitudes. However, it should be noted that the ionosphere in mid-latitudes tends to be more stable compared to low-latitudes,

Fig. 12 Mean accuracy of five testing stations over NT in different seasons (Jan, Apr, Jul, and Oct 2021). The mean accuracy of the testing station depends on its location which is located inside (DARW, KAT1, and LARR) or outside (KUNU and GROT) the network. Accuracy in Jan 2021 is worse than in other months because of the lower number of GNSS stations in the network
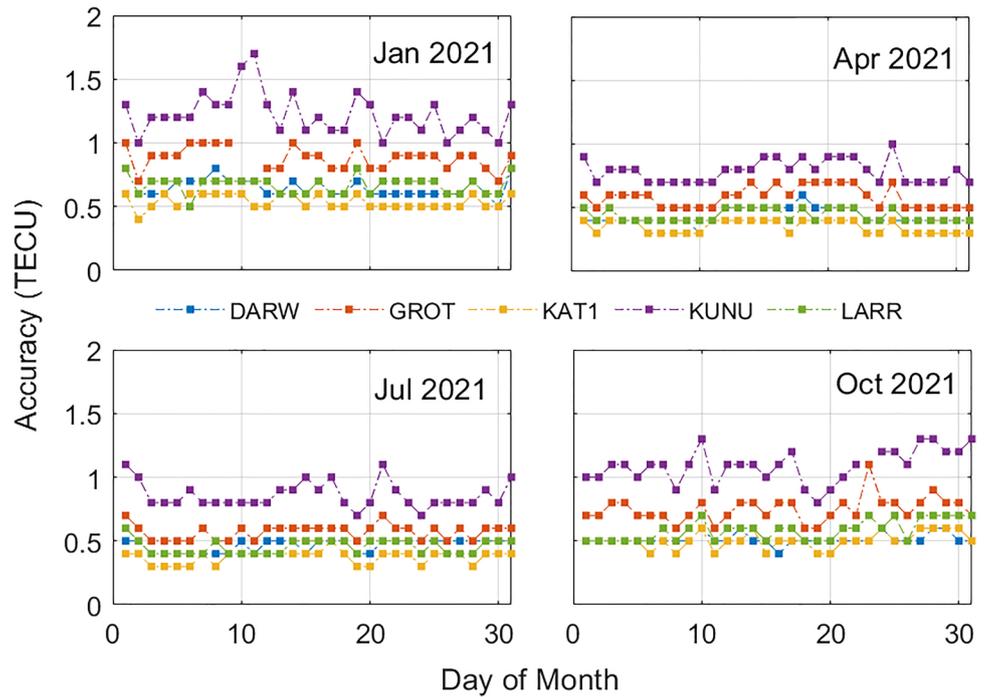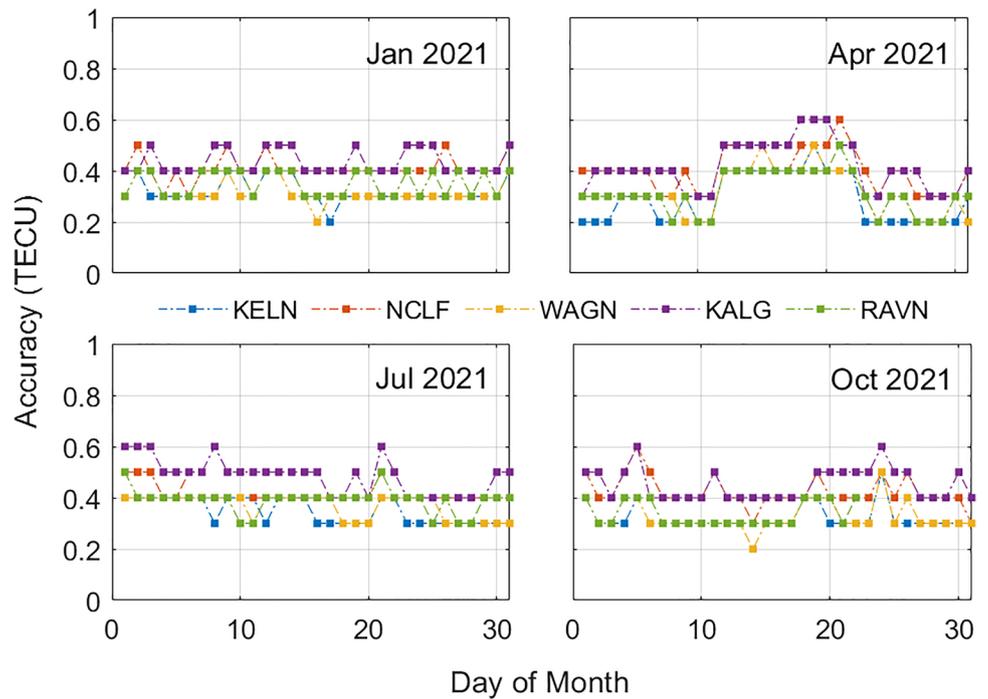

Fig. 13 Mean accuracy of five testing stations over WA (KELN, NCLF, WAGN, KALG, and RAVN) in different seasons (Jan, Apr, Jul, and Oct 2021). The mean accuracy of the model for those stations in this network is within 0.2 to 0.6 TECU

which makes it easier to construct accurate local maps even with a sparse GNSS stations network. In low latitudes, the establishment of a well-distributed GNSS network with an adequate number of receivers becomes crucial for achieving higher accuracy for local ionospheric mapping. For example, in the NT network, the limited number of stations (approximately 9–10) within the region and the absence of additional stations around KUNU contribute to reduced accuracy in that specific location. The low-latitude region is known for its highly variable ionosphere and spatial gradients resulting from the equatorial anomaly, which can significantly impact mapping errors. The findings from the KUNU station, evaluated using the local map over the NT network, demonstrated the importance of having enough data and an appropriate
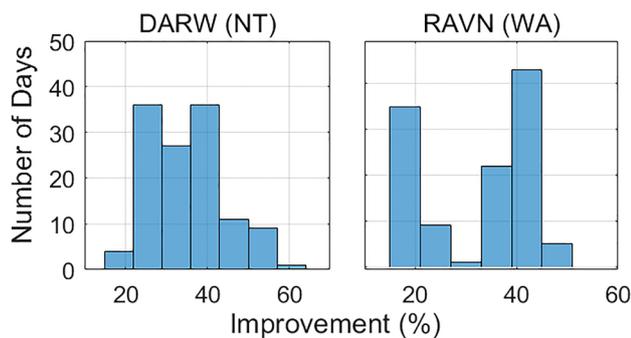
Fig. 14 The distribution of accuracy improving by using 3.5b threshold for DARW station (NT network) and RAVN station (WA network)

network configuration of stations for achieving accurate ionospheric correction mapping in low-latitude regions.

Our data analysis using bilinear ionospheric models revealed that the residuals formed by Laplace distribution. Regional mapping of Marini-Pereira et al. (2020) which used groupings of measurements in time and space, triangulation, linear interpolation, and smoothing also demonstrated that the normal distribution does not accurately represent the model errors. Ma et al. (2022) investigated the errors of both the slant TEC and the derived TEC and found that they follow Laplace distribution rather than Gaussian distribution. Therefore, it is worth to consider data distribution before applying threshold to remove an appropriate percentile of outliers for that distribution.

Conclusion

Detecting outliers in ionospheric local mapping is essential for achieving improved accuracy, especially in low-latitude regions. This study presents a method for detecting outliers in ionospheric corrections, levering statistical residuals derived from the ionospheric local model. New thresholds for outlier detection have been proposed by assigning weight numbers to account for the extended tail of the Laplace distribution. By applying statistical thresholds of $\mu + 5.8b$, $\mu + 3.5b$, and $\mu + 3b$, it becomes possible to detect and eliminate outliers in the model. These thresholds are designed to preserve 99.7%, 97%, or 95% of the values within the respective thresholds. Consequently, modelling outliers can be effectively identified and removed, leading to enhanced accuracy in the ionospheric corrections. Applying modelling thresholds to remove outliers can cause data loss in some circumstances. However, it ultimately leads to a more accurate model. In low-latitude regions, approximately 2.5% of epochs exhibit no available satellites i.e., no correction data can be made using this local network, when the 3.5b threshold is applied, but model accuracy during the testing period justifies this sacrifice. Moreover, the threshold for outlier detection in the ionospheric model is based on

the statistical data of b. Therefore, increasing the amount of data can potentially lead an enhanced accuracy of the model. Future research should focus on exploring different scenarios, such as during periods of geomagnetic disturbance and high solar activity, to evaluate the performance of the outlier detection method in these specific conditions.

In conditions of low solar activity and quiet conditions, of the application of a 3.5b threshold for detecting and removing outliers can improve the accuracy of the model. This improvement is consistent in both low-latitude and mid-latitude regions, with the accuracy improvement of up to 50%. Based on data observation at each epoch using a 3.5b threshold, the mean accuracy of the regional model is approximately 0.5 TECU for mid-latitude regions (WA). In low-latitude regions, such as the NT, the mean accuracy ranges from 0.5 to 2.5 TECU. These results highlight the effectiveness of the 3.5b threshold in reducing modelling outliers and improving the overall accuracy of the local ionospheric model. To detect outliers for future data, regarding the increasing solar activities, updating the statistical location and scale parameters could bring better accuracy in local models.

Declarations

References

Banville S, Collins P, Zhang W, Langley RB (2014) Global and regional ionospheric corrections for faster PPP convergence. J Inst Navig 61:115–124. https://doi.org/10.1002/navi.57

Banville S, Hassen E, Walker M, Bond J (2022) Wide-area grid-based slant ionospheric delay corrections for precise point positioning. Remote Sens 14:1073. https://doi.org/10.3390/rs14051073

Budtho J, Supnithi P, Saito S (2018) Analysis of quiet-time vertical ionospheric delay gradients around Suvarnabhumi Airport. Thailand Radio Sci 53:1067–1074. https://doi.org/10.1029/2018RS006606

Dao T, Harima K, Carter B, Currie J, McClusky S, Brown R, Rubinov E, Choy S (2022) Regional ionospheric corrections for high accuracy GNSS positioning. Remote Sens 14:2463. https://doi.org/10.3390/rs14102463

Datta-Barua S, Lee J, Pullen S, Luo M, Ene A, Qiu D, Zhang G, Enge P (2010) Ionospheric threat parameterization for local area global-positioning-system-based aircraft landing systems. J Aircr 47:1141–1151. https://doi.org/10.2514/1.46719

Gao Y, Liu ZZ (2002) Precise ionosphere modeling using regional GPS network data. J Glob Position Syst 1:18–24

Ginan (2023) Geoscience Australia, Canberra, https://doi.org/10.26186/146649

Glen S (2020) Empirical rule (68–95–99.7) & empirical research, From StatisticsHowTo.com: elementary statistics for the rest of us! https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/empirical-rule/

Grewal MS, Andrews AP, Bartone CG (2020) Global navigation satellite systems, Inertial Navigation, and Integration, 4th Edition, ISBN: 978–1–119–54783–9. https://doi.org/10.1002/9781119547860

Jakowski N, Mayer C, Wilken V, Hoque M (2008) Ionospheric impact on GNSS signals, book: Fisica de la Tierra, Publicaciones Universidad Complutense de Madrid

Li X, Zhang X, Ge M (2011) Regional reference network augmented precise point positioning for instantaneous ambiguity resolution. J Geodesy 85:151–158. https://doi.org/10.1007/s00190-010-0424-0

Li W, Li Z, Wang N, Liu A, Zhou K, Yuan H, Krankowski A (2022) A satellite-based method for modeling ionospheric slant TEC from GNSS observations: algorithm and validation. GPS Solut 26:14. https://doi.org/10.1007/s10291-021-01191-2

Lotfy A, Abdelfatah M, El-Fiky G (2022) Improving the performance of GNSS precise point positioning by developed robust adaptive Kalman filter. Egypt J Remote Sens Space Sci 25–04:919–928. https://doi.org/10.1016/j.ejrs.2022.09.005

Lu K-P, Chang S-T (2022) Robust switching regressions using the laplace distribution. Mathematics 10:4722. https://doi.org/10.3390/math10244722

Ma G, Fan J, Wan Q, Li J (2022) Error characteristics of GNSS derived TEC. Atmosphere 13:237. https://doi.org/10.3390/atmos13020237

Marini-Pereira L, Lourenço LF, Sousasantos J, Moraes AO, Pullen S (2020) Regional ionospheric delay mapping for low-latitude environments. Radio Sci 55(12):1–6. https://doi.org/10.1029/2020RS007158

Rovira-Garcia A, Ibáñez-Segura D, Orús-Perez R, Juan JM, Sanz J, González-Casado G (2020) Assessing the quality of ionospheric models through GNSS positioning error: methodology and results. GPS Solut 24:4. https://doi.org/10.1007/s10291-019-0918-z

Rungraengwajiake S, Supnithi P, Saito S, Siansawasdi N, Saekow A (2015) Ionospheric delay gradient monitoring for GBAS by GPS stations near Suvarnabhumi airport. Thailand Radio Sci 50:1076–1085. https://doi.org/10.1002/2015RS005738

Spilker JJ Jr, Axelrad P, Parkinson BW, Enge P (Eds.) (1996) Global positioning system: theory and applications (Vol 1). American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/5.9781600866388.0000.0000

Subirana JS, Zornoza JMJ, Hernández-Pajares H (2013) GNSS data processing. Vol I: Fundamentals and Algorithms. ESA TM-23/1, May 2013. European Space Agency Communication. ISBN: 978-92-9221-886-7. https://gssc.esa.int/navipedia/GNSS_Book/ESA_GNSS-Book_TM-23_Vol_I.pdf

Tao A-L, Jan S-S (2016) Wide-area ionospheric delay model for GNSS users in middle- and low-magnetic-latitude regions. GPS Solut 20:9–21. https://doi.org/10.1007/s10291-014-0435-z

Teunissen PJG, de Bakker PF (2013) Single-receiver single-channel multi-frequency GNSS integrity: outliers, slips, and ionospheric disturbances. J Geod 87:161–177. https://doi.org/10.1007/s00190-012-0588-x

Warnant R, Foelsche U, Aquino M, Bidaine B, Gherm V, Hoque MM, Kutiev I, Lejeune S, Luntama JP, Spits J, Strangeways HJ (2009) Mitigation of ionospheric effects on GNSS. Ann Geophys 52(3–4):373–390. https://doi.org/10.4401/ag-4585

Yang H, Monte-Moreno E, Hernández-Pajares M, Roma-Dollase D (2021) Real-time interpolation of global ionospheric maps by means of sparse representation. J Geodesy 95:71. https://doi.org/10.1007/s00190-021-01525-5

Yang J, Rahardja S, Fränti P (2019) Outlier detection: how to threshold outlier scores? AIIPCC '19 In: Proceedings of the international conference on artificial intelligence, information processing and cloud computing, Vol 37: pp 1–6. https://doi.org/10.1145/3371425.3371427

Zhang Q, Zhao L, Zhao L, Zhou J (2015) An improved robust adaptive kalman filter for GNSS precise point positioning. IEEE Sens J 18(10):4176–4186. https://doi.org/10.1109/JSEN.2018.2820097

Tam Dao is a research fellow at RMIT University. She completed her PhD at Nagoya University in 2017. Tam conducts research in space sciences and space applications. Her studies related to ionospheric physics, space weather forecasts, and ionospheric models applied to precise positioning and navigation. She specializes in data analysis of ground-based observations, satellite data, and simulations.

Ken Harima received a bachelor's degree in electronic engineering from Universidad Simon Bolivar, Caracas, Venezuela, in 2002, and his Ph.D. degree from the University of Tokyo, Tokyo, Japan, in 2012. He is currently developing GNSS analysis software for Geoscience Australia, Canberra, Australia. His research interests include GNSS receivers and GNSS augmentation systems.

Brett Carter is currently an Associate Professor at RMIT University. He obtained his PhD at La Trobe University in 2011 on the topic of ionospheric plasma waves observed using ground-based HF radar. He has expertise in ionospheric physics, space weather impacts on GPS and space situational awareness.

Julie Currie is a lecturer at RMIT University. She obtained her PhD in 2017 from the University of Newcastle and was a Vice-Chancellor Postdoctoral Fellow at RMIT University after that. Her research is focused on understanding how space weather affects the dynamics of near-Earth space and the impacts of this on ground-based and space-based communication, surveillance, and positioning technologies.

Simon McClusky is currently a GNSS Scientist at Geoscience Australia. He received a PhD in surveying from the School of Surveying at the University of New South Wales (UNSW) in 1994. He has twenty years of experience in the fields of geodetic and geophysical research. His expertise in geodetic data analysis, geophysical data modelling and scientific software development.

Rupert Brown is a Product Manager and System Architect accustomed to complex scientific and engineering environments. He is currently working on the product definition for global navigation satellite system-derived precise point positioning software and analysis products. Rupert has a BSc (Hons), MBA, MIET, Eur Ing.

Eldar Rubinov is a Positioning and Geodesy Technical Lead at FrontierSI, a not-for-profit research organisation in Australia. He has completed a PhD in stochastic modelling for high-precision GNSS at the University of Melbourne and has over 20 years of experience working in all aspects of PNT. Eldar is the current chair of the FIG Working Group 5.4 on GNSS.

John Barassi is currently working on the Ginan Positioning Analysis Centre Software delivery. He graduated from Monash University and majored in Electrical and Computer Systems Engineering with a study in Artificial Neural Networks for Control Systems. Over a career with many highlights, he has focused on enhancing capability for Telecommunications, Energy and Utility companies, including Networks, Provisioning, Media Streaming and Vehicle tracking systems.

Suelynn Choy is a professor of Satellite Navigation at RMIT University. Her research interests are concentrated in the field of precise satellite positioning and navigation, geodesy, and the use of geospatial and satellite technologies for disaster management and atmospheric remote sensing. Suelynn has contributed to industry collaborations and advised national and international committees on satellite navigation and positioning.