Satellite Navigation
https://satellite-navigation.springeropen.com/

ORIGINAL ARTICLE

Open Access

Amplitude scintillation detection with geodetic GNSS receivers leveraging machine learning decision tree

Wang Li[1], Yiping Jiang[1*], Hongyuan Ji[1] and Wenqiang Wei[2]

Abstract

The amplitude scintillation detection is typically achieved by using the scintillation index generated by dedicated and costly ionospheric scintillation monitoring receivers (ISMRs). Considering the large volume of common Global Navigation Satellite System (GNSS) receivers, this paper presents a strategy to accurately identify the ionospheric amplitude scintillation events utilizing the measurements collected with geodetic GNSS receivers. The proposed detection method relies on a pre-trained machine learning decision tree algorithm, leveraging the scintillation index computed from the carrier-to-noise data and elevation angles collected at 1-Hz. The experimental results using real data demonstrate a 99% accuracy in scintillation detection can be achieved. By combining advanced machine learning techniques with geodetic GNSS receivers, this approach is feasible to effectively detect ionospheric scintillation using non-scintillation GNSS receivers.

Keywords Ionosphere, Amplitude scintillation, Machine learning

Introduction

The ionosphere is a region of the Earth's upper atmosphere, from an altitude of about 50 km to about 1000 km (Enge, 1994). It contains charged particles, such as electrons and ions, that can affect the propagation of radio waves. The ionospheric scintillation poses a threat to Global Navigation Satellite Systems (GNSS) users by causing rapid amplitude and random phase variations of the GNSS signals (Pi et al., 1997). Under scintillation events, the GNSS receivers are more vulnerable to cycle slips and loss of lock, leading to a degraded navigation performance (Kintner et al., 2007; Seo et al., 2011). Scintillation events frequently occur in the equatorial, auroral, and polar regions with more frequent and intense scintillation in the equatorial region (within ±5° around the magnetic equator) (Jiao & Morton, 2015). One of the major causes of scintillation in the equatorial region is the equatorial plasma bubble (EPB) occurring after local sunset. The EPB is characterized by the large-scale depletion of F-region electron densities induced by the Rayleigh-Taylor instability (Ott, 1978). These irregularities in plasma density result in localized regions of depletion electron density, forming bubble-like structures. In addition to the EPB, sporadic-E (Es) can also cause scintillation events because of their strong vertical electron density gradients (Seif et al., 2017). Modeling and prediction of ionospheric scintillation is hard since it involves many variable factors such as wave interactions, local electric field, and interplanetary magnetic field activities (Yeh & Liu, 1982).

Detecting and monitoring scintillation is crucial for space-based applications such as GNSS. Accurate and early detection allows for the development of algorithms and techniques to mitigate its impact on navigation accuracy

*Correspondence:
Yiping Jiang
yiping.jiang@polyu.edu.hk
[1] Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, SAR
[2] The School of Electronics and Communication Engineering, Sun Yat-sen University, Xingang Xi Road, Guangzhou, People's Republic of China

(Lee et al., 2017). Precise and timely detection is pivotal for the development of algorithms and techniques to mitigate the adverse effects of scintillation on navigation accuracy. The identification of scintillation empowers GNSS users to proactively anticipate potential signal disruptions, enabling the implementation of strategies to either maintain signal lock or expedite recovery following signal losses (Vila-Valls et al., 2018). The detection of ionospheric scintillation is equally critical in assessing its potential impact on navigation systems integral to these applications. This knowledge enables users to adopt precautionary measures and anticipate challenges. In the instances where accuracy is paramount, users may opt for alternative navigation methods or implement safeguards during the periods of strong scintillation. Furthermore, continuous monitoring of scintillation events contributes to the refinement of models and predictions related to ionospheric disturbances. This ongoing effort enhances our comprehension of Earth's upper atmosphere, ultimately fortifying the reliability and safety of space-based applications (Spogli et al., 2016).

The S_4 index, generated by dedicated ionospheric scintillation monitoring receivers (ISMR), is a well-known measurement for amplitude scintillation and a valuable indicator of amplitude scintillation occurrences. The S_4 index is derived from the detrended signal intensities of GNSS signals computed based on the 100-Hz in-phase and quadrature channel correlator outputs of the ISMR. Based on S_4, signal intensity, in-phase, and quadrature correlation outputs, several methods have been proposed to detect the amplitude scintillation events (Adewale et al., 2012; Taylor et al., 2012; Abadi et al., 2014, Curran et al., 2014, Jiao et al., 2016; 2017, Linty et al., 2019; Favenza et al., 2017). The simplest approach, known as the hard detection method, involves comparing S_4 to a pre-defined threshold ($\mathcal{T}_{S_4}$) (Adewale et al., 2012; Taylor et al., 2012). The ionospheric scintillation is present if S_4 exceeds $\mathcal{T}_{S_4}$. However, due to the rapid variability of S_4, using a single threshold may lead to frequent status changes between scintillation and non-scintillation. Moreover, large S_4 values resulting from low elevation satellites could be falsely identified as scintillation. To reduce the false alarms caused by multipath and other propagation errors, a semi-hard method was proposed (Abadi et al., 2014; Curran et al., 2014). This method incorporates the additional conditions defined on elevation angle (θ_{el}) and carrier to noise (C/N_0) to exclude the measurements that are too noisy. Scintillation is considered as present only if the following conditions are met,

$$S_4 > \mathcal{T}_{S_4} \wedge \theta_{el} > \mathcal{T}_{\theta_{el}} \wedge C/N_0 > \mathcal{T}_{C/N_0} \tag{1}$$

where $\mathcal{T}_{\theta_{el}}$ and $\mathcal{T}_{C/N_0}$ are thresholds for θ_{el} and C/N_0, which are typically set to 30° and 37 dBHz, respectively (Abadi et al., 2014; Curran et al., 2014; Kuruva et al., 2024). However, the semi-hard method might discard important measurements and result in significant risk of missed detection of scintillation events. Manual visual inspection is regarded as the most accurate and reliable method of detecting scintillation events. While manual visual inspection is deemed the most accurate method, relying on the scrutiny of S_4, C/N_0, and extensive experience with scintillation characteristic, it is not automated, susceptible to human errors, and time-consuming. The techniques based on supervised machine learning algorithms such as support vector machine (SVM) (Jiao et al., 2016; 2017), decision tree (Linty et al., 2019; Favenza et al., 2017), and eXtreme Gradient Boosting (XGBoost) (Lin et al., 2021) have shown promising results which resembles manual visual inspection in detecting scintillation events. These machine learning algorithms are trained using a substantial amount of real scintillation data labeled through human visual inspection. In Jiao et al. (2017), the SVM machine learning algorithm was employed with the power spectrum density (PSD) function of signal intensity as input features. This approach resulted in an accuracy ranging from 91 to 96%. Linty et al. (2019) utilized a decision tree machine learning algorithm with averaged 50 Hz in-phase and quadrature correlation outputs as input features. This method achieved an accuracy of 98%. In addition, the detection technique based on semi-supervised machine learning algorithm is proposed to reduce the time of manual labelling (Franzese et al., 2020). However, the major limitation of these methods is their reliance on I and Q data generated by dedicated ISMRs, which are not commonly installed in regional or global GNSS networks. Given the abundance of common geodetic GNSS receivers, there is a growing need for scintillation event detection methods based on these receivers, offering broader applicability.

An alternative index, denoted as S_{4c} and resembling the traditional S_4, has been introduced, which is computed based on the Carrier-to-Noise Density (C/N_0) measurements obtained with common geodetic receivers (Luo et al., 2020). The S_{4c} shows a high correlation with S_4. However, compared with ISMRs which employ resilient tracking loops, low-phase noise oscillators, and stable clocks with advanced signal processing techniques, geodetic receivers are more susceptible to noise and multipath interference (Imam et al., 2023). Using S_{4c} derived from geodetic receivers for scintillation detection might suffer from a significant risk of missed detection and false alarms. Therefore, it is necessary to reduce the impact of noise and multipath on the scintillation detection while retaining valuable data. Recognizing the periodic nature of multipath effects, which differs from

the irregularity of scintillation in a fixed receiver location, we conducted a detailed analysis of multipath patterns. Leveraging this unique feature, we devised a strategy to substantially reduce multipath effects in the detection algorithm by subtracting the S_{4c} values observed under normal conditions. The S_{4c}, along with C/N_0 and the elevation angle (θ_{el}) are used as features for the machine learning decision tree algorithm to achieve automatic amplitude scintillation detection. The main contribution of this study is proposing an automatic method for the accurate detection of amplitude scintillation using geodetic GNSS receivers.

The paper is organized as follows. Section Methods describes the methodology for the amplitude scintillation detection. Section Data collection presents the data source used in this study. Section Results gives detailed results and analysis. Section Conclusions draws and discusses the future work.

Methods

To achieve amplitude scintillation detection with geodetic receivers, it is imperative to initially compute the amplitude scintillation index based on the data collected with common GNSS receivers. Subsequently, we employ a multipath mitigation technique leveraging the known multipath pattern of the fixed receiver to address the distortions caused by multipath in the obtained amplitude scintillation index. Finally, the multipath-mitigated amplitude scintillation index, supplemented by C/N_0 measurements, elevation angle, and azimuth derived from ephemeris calculations, constitutes the comprehensive input for the pre-trained machine learning (ML) algorithm. This ML algorithm is then utilized to classify or detect the presence of amplitude scintillation.

Amplitude scintillation index derivation

The S_4 index serves as a well-established indicator of amplitude scintillation. This index quantifies the strength of variations in the amplitude of the received signal (Van Dierendonck et al., 1993; Vilà-Valls et al., 2020), expressed as:

$$S_4 = \sqrt{\frac{\langle \mathrm{SI}^2 \rangle - \langle \mathrm{SI} \rangle^2}{\langle \mathrm{SI} \rangle^2}} \tag{2}$$

where $\langle\rangle$ represents the time average operator, and SI denotes the signal intensity. SI is typically detrended by normalizing it to a low-passed version of raw signal intensity ($\mathrm{SI}_{\mathrm{raw}}$). The detrended signal intensity, denoted as $\mathrm{SI}_{\mathrm{det}}$, is computed using the narrow band power (NBP) and wide band power (WBP):

$$\mathrm{SI}_{\mathrm{raw}} = \left(\sum_{i=1}^{M} I_i\right)^2 + \left(\sum_{i=1}^{M} Q_i\right)^2 - \sum_{i=1}^{M} \left(I_i^2 + Q_i^2\right) \tag{3}$$

where I_i and Q_i represent the 1-KHz in-phase and quadrature-phase prompt correlator samples obtained from ISMR.

Since the common geodetic GNSS receivers cannot generate the 1-KHz I_i and Q_i data, an alternative method for computing the $\mathrm{SI}_{\mathrm{det}}$ is proposed based on Carrier-to-Noise Density Ratio (C/N_0) measurements (Luo et al. 2020),

$$\mathrm{SI}_{\mathrm{det}} = \frac{S/N_0(k)}{\langle \sum_{i=1}^{n} S/N_0(k-i) \rangle} (k > n) \tag{4}$$

where S/N_0 represents the signal-to-noise density ratio, and n denotes the total number of data points over a 60-second span. S/N_0 is expressed as,

$$S/N_0 = 10^{0.1(C/N_0)} \tag{5}$$

where C/N_0 corresponds to the carrier-to-noise ratio values derived from the Receiver Independent Exchange Format (RINEX) file provided by common GNSS receivers. The S_{4c} can be obtained by substituting (4) into (2).

Both S_4 and S_{4c} measure variations in detrended signal intensity. However, the major difference between S_4 and S_{4c} is the input parameter. For S_4 index, the high-rate in-phase and quadrature-phase provided by the ISMRs are used as input parameters. For S_{4c} index, the input parameter is C/N_0 acquired from common GNSS receivers. Although the in-phase and quadrature-phase components are correlated with C/N_0, they cannot be acquired from RINEX files (Motella et al., 2008).

The upper panels of Fig. 1a and b show S_{4c} derived by a common geodetic receiver (HKOH in this case) and S_4 derived by the ISMR near HKOH. While the magnitude of S_4 and S_{4c} may differ, and both indices exhibit a notable sharp increase, exceeding 0.4 at approximately 21:00 local time (LT). This surge is attributed to the influence of amplitude scintillation. Furthermore, it is observed that the magnitudes of both S_4 and S_{4c} tend to escalate when the elevation angle is low. However, the C/N_0 of common geodetic receivers changes more rapid changes than that of ISMR, particularly during amplitude scintillation. This observation underscores the superior robustness of ISMRs in the presence of ionospheric anomalies compared to common geodetic receivers. It also elucidates the rationale behind utilizing the C/N_0 measurements from common geodetic receivers to to compute S_{4c} for reflecting amplitude scintillation. However, owing to the absence of advanced

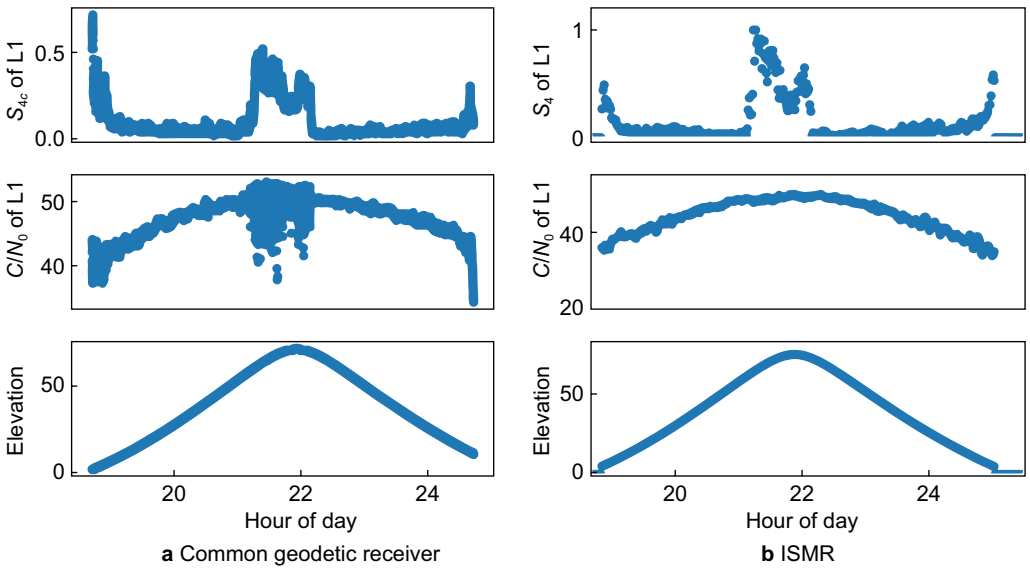

Fig. 1 Comparison between geodetic receiver and ISMR of GPS PRN 24 on September 14, 2014

signal processing techniques, the time series of S_{4c} and C/N_0 from common geodetic receivers exhibit more pronounced noise than those from ISMRs. The fluctuations in S_{4c}, attributed to multipath and noise, may potentially obscure the true amplitude scintillation patterns.

Multipath effect mitigation

Due to the influence of multipath, distinguishing whether the increase in S_{4c} is attributable to multipath or amplitude scintillation poses a significant challenge. This challenge is particularly pronounced when the elevation angle is low since the rise in S_{4c} might mask the presence of scintillation.

To address this challenge, a conventional approach involves implementing an elevation mask, typically set at 30°. However, this method results in the exclusion of a substantial amount of valuable data, rendering scintillation detection impractical for the satellites with elevation angles below 30°. To retain a more significant portion of useful data, an alternative approach employs a smaller mask angle of 5°.

To further mitigate the impact of multipath and thermal noise, two strategies are implemented. Firstly, the scintillation index (S_{4c}) is averaged over the observation period using a short observation window. The resulting smoothed S_{4c} ($\hat{S}4c$), obtained through averaging with a 60-second window size (Fig. 2), exhibits reduced noise levels and diminished multipath effects, enhancing its utility for scintillation detection. However, it's crucial to note that the increased values of $\hat{S}4c$ (exceeding 0.2 around 19:00 LT in Fig. 2), attributed to multipath at

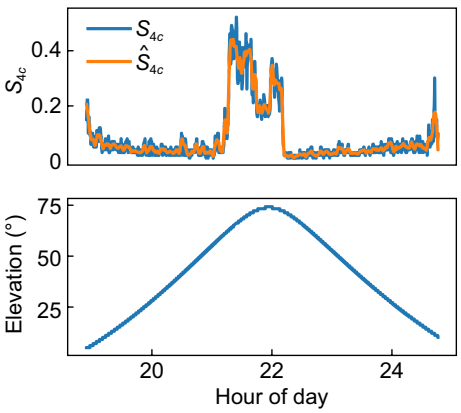

Fig. 2 Illustration of S_{4c} averaging with a window size of 60 s for PRN 24 on September 14, 2014

low elevations during initial satellite tracking, may be mistakenly interpreted as scintillation events. Careful consideration of such instances is essential for accurate scintillation detection.

The second strategy is proposed to further mitigate the impact of multipath, particularly for the satellites with low elevation angles. Figure 3 illustrates the $\hat{S}4c$ and corresponding θ_{el} (elevation angle) for station HKOH with respect to PRN 24 from September 14 to 20, 2014. All these curves show a U shape as $\hat{S}4c$ increase sharply at the beginning and the ending of the time period. These increases are corresponding to the satellite signals with low elevation angles. As the geometric relationship between the GPS constellation and a stationary station repeats every sidereal day, the θ_{el}

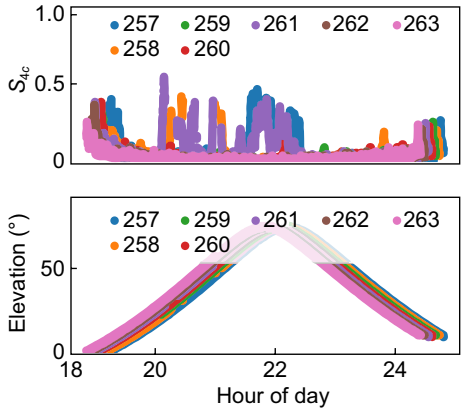

Fig. 3 $\hat{S}_{4c}$ values for PRN 24 observed by HKOH station from September 14 to September 20, 2014

curves remain consistent with a slight time shift. The repeat period for all GPS satellites is one sidereal day which is 23 h 55 m 55 s (86155 s) (Choi et al., 2004). The fluctuation in $\hat{S}4c$ induced by multipath effects repeats every sidereal day, as evident in the $\hat{S}4c$ curves at the beginning and end of the selected period when θ_{el} is small. This periodic pattern can be harnessed to alleviate the multipath effect. To represent the $\hat{S}4c$ fluctuation caused by multipath, the time series of $\hat{S}4c$ from a non-scintillation day (e.g., September 20) is selected. Subsequently, this $\hat{S}4c$ is employed to mitigate the multipath effect by subtracting it from the $\hat{S}4c$ observed on the target day.

Figure 4 compares the $\hat{S}_{4c}$ and $\hat{S}_{4c}$ curves with multipath reduced. The time series of $\hat{S}_{4c}$ on September 20 is used to correct the multipath effect for the target day of September 14. After correction, the large $\hat{S}_{4c}$ in tails due to multipath of low elevation is largely reduced. Therefore, the corrected $\hat{S}_{4c}$ is used for the scintillation detection and expected to outperform using $\hat{S}_{4c}$.

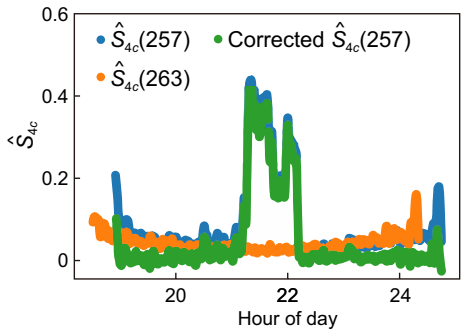

Fig. 4 The $\hat{S}_{4c}$ curves of HKOH with reduced multipath effect on September 14 for PRN 24 (Day of year 257 to 263)

Machine learning

Machine learning encompasses a diverse array of algorithms designed to construct models based on given datasets and facilitate predictions. These algorithms fall into three main categories: supervised learning, unsupervised learning, and semi-supervised learning. The distinction among these categories lies in the availability of labeled, unlabeled, or a combination of labeled and unlabeled datasets. Given that the primary objective of this study is the detection, i.e., classification of scintillation events, the chosen approach involves employing a supervised learning algorithm. The terms "classification" and "detection" have the same meaning in this study. A visual representation of the supervised machine learning process is depicted in the flow diagram presented in Fig. 5.

As illustrated in Fig. 5, the dataset comprising observations and ephemeris information is collected. Utilizing this data, features such as θ_{el}, $\hat{S4c}$, and C/N_0 are computed, providing descriptive characteristics of the observed domain. The dataset is then labeled by manual annotation which is used for training the machine learning algorithm and validating its performance.

Detection employs the decision tree machine learning algorithm, a widely utilized and versatile tool in the realm of supervised machine learning. Renowned for its simplicity, interpretability, and capability to address a broad spectrum of problems, the decision tree algorithm is a robust choice. Compared with SVM, XGBoost, decision tree is more computationally efficient. In addition, the decision tree algorithm was demonstrated to be effective in detecting amplitude scintillation in

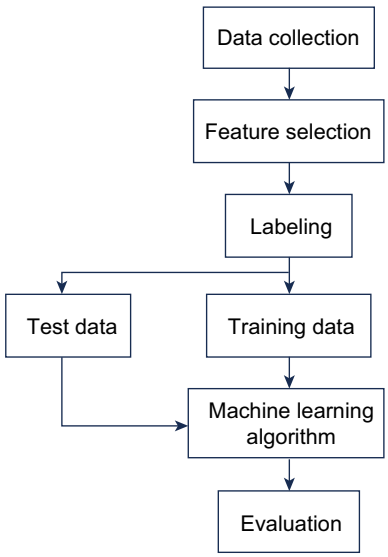

Fig. 5 Flow diagram of the supervised machine learning process targeting a binary classification task

previous study (Linty et al., 2019). Operating on a tree structure, it predicts the value of a target variable by deducing simple decision rules from the features present in the dataset. Particularly effective for classification problems, decision trees excel in handling complex relationships within the data.

The decision tree algorithm operates through a recursive partitioning of the dataset to construct a tree-like structure, which is then employed for decision-making and predicting class labels for new, unseen data. The process initiates at the root node, encompassing the entire dataset. Different features are evaluated to identify the optimal split, defined as the one maximizing information gain or minimizing impurity (e.g., Gini impurity, entropy) in the resulting subsets.

Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled. For a node t, the Gini impurity ($G(t)$) is calculated as,

$$G(t) = 1 - \sum_{i=1}^{c} (p_i)^2 \tag{6}$$

where c is the number of classes. In the scintillation detection, where two classes are considered ($c = 2$), p_i represents the proportion of samples in class i at node t.

Information gain assesses a feature's effectiveness in reducing uncertainty about the dataset's classes after a split on feature A in node t:

$$IG(t, A) = G(t) - \sum_{v \in \text{values(A)}} \frac{N_v}{N_t} G(v) \tag{7}$$

where values(A) is the set of possible values for feature A. N_v is the number of samples in node v after the split. N_t is the total number of samples in node t. $G(v)$ represents the Gini impurity of node v.

The algorithm iterates this process for each subset, generating child nodes and further splits. This recursive procedure continues until a stopping criterion is met, such as reaching a maximum tree depth or having a minimum number of samples in a leaf node. Once the tree is constructed, each leaf node corresponds to a class label or a probability distribution over the class label. When a new data point is inputted, the tree traverses from the root to a leaf node based on the feature values of the data point, and the associated class label is assigned to the input data.

The selection of features from the measurements provided by the GNSS receiver, namely the observation and ephemeris data, is a critical step in training the decision tree model. However, choosing these features is not a straightforward task, as the performance and generality of the machine learning algorithm are contingent on this selection. To aid this decision, a statistical tool known as the correlation matrix is employed to highlight the relationships between each pair of features. The Pearson correlation coefficient ($|\rho(X, Y)|$) between features X and Y is used. The correlation coefficient ranges from -1 to 1. A strong correlation is evident when $|\rho(X, Y)|$ exceeds 0.68, a moderate correlation if $|\rho(X, Y)|$ falls in the range 0.36–0.67, a weak correlation if $|\rho(X, Y)|$ is smaller than 0.35, and no correlation when $|\rho(X, Y)|$ equals 0. Figure 6 shows the correlation matrix between the manual ground truth and $\hat{S}_{4c}$, C/N_0, θ_{el}, azimuth angle (θ_{az}), and PRN number.

As depicted in Fig. 6, manual annotation exhibits a strong correlation with $\hat{S}4c$, a moderate correlation with θ_{el} and C/N_0, and a weak correlation with PRN number and θ_{az}. Consequently, a feature set comprising $\hat{S}4c, \theta_{el}, C/N_0$ is selected, encompassing the observables with the highest correlation to manual annotation.

Data collection

The GNSS data used in this study were collected from Hong Kong Satellite Positioning Reference Station Network (SatRef) which comprises 18 GNSS stations strategically located in pre-surveyed positions. Initially established in 2001 with 6 stations, the network was expanded in 2014 to encompass 18 stations evenly distributing across the region. The geomagnetic latitude of these stations are between 12.65° North and 13.49° North, which are near closely to the equatorial anomaly. The scintillation in this area is known to be stronger and more frequent than near the magnetic equator (Kintner et al., 2004). The C/N_0 data with a 1-Hz interval in the RINEX file collected from SatRef stations of year 2014 are used.

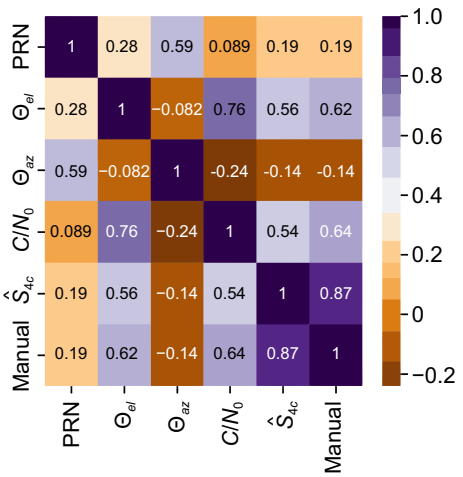

Fig. 6 Correlation matrix between observables of the signal

The dataset with assigned class labels is used to train the decision tree algorithm. Two class labels are assigned in this study by manual annotation: 0 for non-scintillation events and 1 for scintillation events. Three cases (weak, medium, and strong scintillation cases) illustrating the dataset labeling process is exemplified in Fig. 7. Typically, the $S_{4c} < 0.2$ indicates no scintillation, $0.2 \leq S_{4c} < 0.4$ means weak scintillation, $0.5 \leq S_{4c} < 0.6$ is considered as medium, and $S_{4c} > 0.6$ means strong scintillation (Kai et al., 2017).

A detailed list of the training data segments collected from SatRef in year 2014 is provided in Table 1. The total length of all the training data is approximately 60.5 h, consisting of 153,180 points, maintaining a ratio of scintillation signals to non-scintillation signals at approximately 1:2.1. A 30% hold-out validation is configured to assess the decision tree algorithm's performance. This implies that 70% of the training data is randomly selected for training the decision tree algorithm, while the remaining 30% is reserved for validating the trained algorithm.

Table 1 Data used for training and validation

Station	PRN	Date	Start and end (LT)	Duration
HKPC	1	01/03/2014	21:17:34–23:40:53	2 h 23 min 19 s
HKCL	27	01/03/2014	16:53:34–21:04:30	4 h 13 min 56 s
HKNP	27	02/03/2014	18:12:26–22:08:30	3 h 56 min 4 s
HKMW	1	05/03/2014	20:56:36–23:31:58	2 h 35 min 22 s
HKSS	24	14/09/2014	19:42:20–24:09:40	4 h 27 min 20 s
HKWS	12	14/09/2014	21:24:20–23:26:00	2 h 1 min 40 s
HKPC	25	14/09/2014	22:02:00–23:59:20	1 h 57 min 40 s
HKKS	29	14/09/2014	17:19:20–20:39:20	3 h 20 min 0 s
HKNP	24	15/09/2014	18:57:46–22:41:19	3 h 42 min 40 s
HKSL	25	16/09/2014	00:43:49–02:37:40	1 h 53 min 51 s
HKSS	29	16/09/2014	19:47:10–22:09:20	2 h 22 min 10 s
HKKT	12	17/09/2014	21:07:57–21:42:50	0 h 34 min 53 s
HKLT	29	18/09/2014	17:35:20–21:13:06	3 h 38 min 40 s
HKPC	29	20/09/2014	20:50:37–21:45:15	0 h 54 min 38 s
HKNP	25	05/11/2014	19:24:20–20:47:40	1 h 23 min 20 s
HKMW	25	10/11/2014	20:22:40–23:30:09	3 h 7 min 29 s

Results

Validation

In this section, we evaluate the detection capabilities of the proposed method which leverages the decision tree machine learning algorithm for a two-class classification task (0 for non-scintillation events and 1 for scintillation events). To gauge the effectiveness of the machine learning classification algorithm, four metrics are employed with their meanings explained as follows:

1. Accuracy: Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of instances. Accuracy is calculated by

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{8}$$

2. Precision: Precision measures the accuracy of positive predictions. It is the ratio of correctly predicted positive instances to the total predicted positive instances.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{9}$$

3. Recall (Sensitivity): Recall measures the ability of the model to capture all the positive instances. It is the ratio of correctly predicted positive instances to the total actual positive instances.

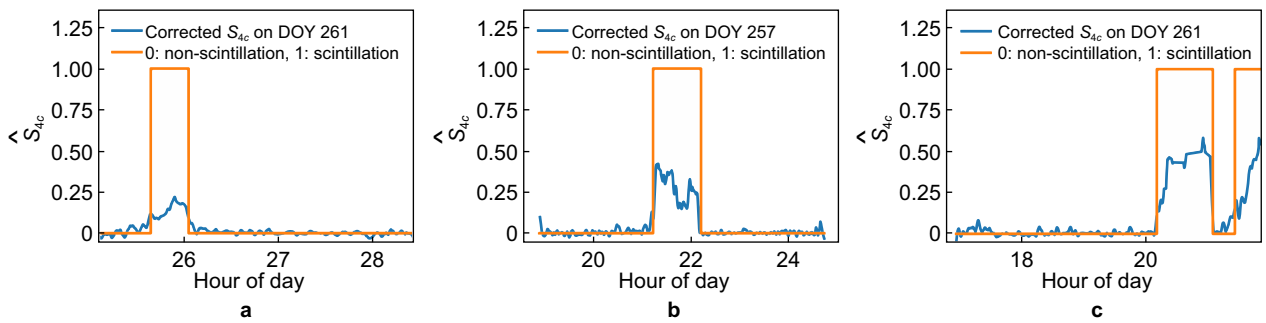

Fig. 7 Labeling results based on manual inspection. **a** weak scintillation case of PRN 29 on September 18. **b** medium scintillation case of PRN 24 on September 14. **c** strong scintillation case of PRN 25 on September 18

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (10)$$

4. F-score: The F-score is the harmonic mean of precision and recall. It provides a balance between precision and recall.

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

The detection performance of the decision tree, hard, and semi-hard methods, considering various feature sets, is detailed in Table 2. As previously discussed, the hard and semi-hard methods represent conventional approaches to scintillation detection, relying on single and multiple thresholds, respectively. Notably, the thresholds set for S_{4c}, C/N_0, and elevation angle are 0.2, 37 dBHz, and 30°, respectively. A comparative analysis in terms of accuracy, precision, recall, and F-score is provided with a particular emphasis on accuracy and F-score due to their holistic assessment of algorithmic performance and consideration of data distribution. The results underscore the superior performance of the decision tree method over the hard and semi-hard methods, exhibiting high-accuracy detection and classification. The hard method, despite its simplicity, demonstrates lower detection accuracy and F-score. The semi-hard method shows improved performance by imposing restrictions on C/N_0 and elevation angle features. However, this approach overlooks valuable information from satellite signals with the elevation angles below the threshold. Furthermore, the performance of the semi-hard methods relies on predefined thresholds for C/N_0, and elevation angle, which vary with surrounding environment of location. In contrast, the machine learning algorithm retains potentially valuable information and is location-independent. Substituting the feature S_{4c} with $\hat{S}_{4c}$ results in reduced multipath effects, leading to the enhancements in accuracy and F-score across all methods. Notably, employing $\hat{S}_{4c}$ in the decision tree method yields an approximate 7% improvement in detection performance. Moreover, the incorporation of features such as θ_{el} and C/N_0 further augments detection performance, achieving a remarkable 99.9% detection accuracy. These findings underscore the efficacy of the trained decision tree algorithm, utilizing $\hat{S}_{4c}$, C/N_0, and θ_{el} as input features, in capturing the intricate dynamics of scintillation and facilitating accurate detections. The decision tree algorithm has also been applied to fields such as predicting seismo-ionospheric anomalies (Akhoondzadeh, 2016) and forecasting total electron content (Han et al., 2022).

Table 2 Detection performance with different feature sets

Method	$\mathcal{L}$	Accuracy	Precision	Recall	F-score
Hard	S_{4c}	79.92%	89.77%	54.14%	51.96%
	$\hat{S}_{4c}$	81.80%	89.52%	58.56%	59.43%
Semi-hard	$S_{4c}, C/N_0, \theta_{el}$	90.28%	89.73%	90.27%	89.97%
	$\hat{S}_{4c}, C/N_0, \theta_{el}$	91.11%	93.52%	88.94%	90.32%
Decision tree	S_{4c}	90.45%	89.94%	90.36%	90.13%
	$\hat{S}_{4c}$	97.52%	97.54%	97.31%	97.42%
	$\hat{S}_{4c}, C/N_0$	99.13%	99.24%	98.96%	99.10%
	$\hat{S}_{4c}, C/N_0, \theta_{el}$	99.91%	99.92%	99.88%	99.90%

Linty et al. (2019) previously employed the decision tree algorithm for scintillation detection. However, our proposed methodology differs in terms of the features utilized within the decision tree. While Linty et al. (2019) relied on measurements from ISMR, our approach incorporates measurements from common geodetic receivers. In their study, Linty et al. (2019) achieved a detection accuracy of 96.7% using the decision tree algorithm, with features including S_4, C/N_0, and θ_{el}. In comparison, our proposed methodology exhibits a better performance, achieving a detection accuracy of 99.9% with $\hat{S}_{4c}$, C/N_0, and θ_{el} as input features. Additionally, Linty et al. (2019) demonstrated that a detection accuracy of 99.7% can be achieved considering signal-based features I and Q, along with averaged I and Q values. Despite this, our approach yields comparable and slightly better results. This indicates the efficacy of our approach over the method described in Linty et al. (2019).

While the machine learning decision tree approach demonstrates high detection accuracy, there is a risk of overfitting as the depth increases due to an increased number of data splits, reducing the number of data points per feature and invoking the curse of dimensionality. In other words, as the tree depth increases, accuracy on the training dataset may continue to improve, but accuracy on the test dataset may lower. Additionally, algorithm complexity increases with tree depth. To mitigate overfitting and reduce algorithm complexity, an optimal tree depth must be determined. Figure 8 displays the accuracy, recall, precision, and F-score on the test dataset with varying tree depths. The plot indicates that increasing the tree depth initially improves performance on the test dataset until a depth of 8 levels. Beyond this point, the algorithm tends to overfit the training dataset, resulting in worse performance on the holdout dataset. Therefore, a tree depth of 8 is chosen for this case.

Figure figtree illustrates the accuracy, recall, precision, and F-score achieved by running the decision tree algorithm using varying numbers of input points for training. It is essential that the training dataset encompasses

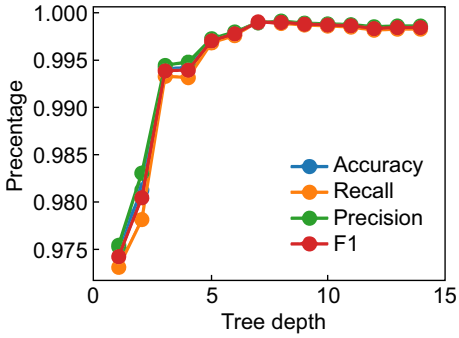

Fig. 8 Detection performance on train and test dataset for different tree depths

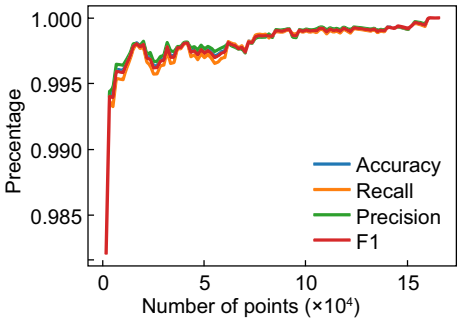

Fig. 9 Detection performance of the decision tree algorithm versus the number of points used in the training set

enough points to represent diverse levels of scintillation events, including weak, medium, and strong occurrences. Moreover, the size of the training dataset must be sufficiently large to ensure satisfactory detection performance. As depicted in Fig. 9, a minimum training dataset of 100,000 points is recommended to attain an accuracy and F-score exceeding 99.7%. This ensures robust performance across different levels of scintillation events and underscores the importance of adequate data coverage in training for reliable detection outcomes.

Test on novel data

Some test results on the novel data collected on other days using the trained decision tree algorithm are presented in Fig. 10. Notably, these data were not involved in the training process. Each subplot displays the S_{4c} and $\hat{S}4c$ index values alongside the predicted classes for all blocks. In comparison, $\hat{S}4c$ is found to be more suitable than S_{4c} in indicating scintillation. Moreover, valuable data is preserved for the satellite signals with a small elevation angle, as evidenced by the end of the time period shown in Fig. 10a and the beginning of the time period shown

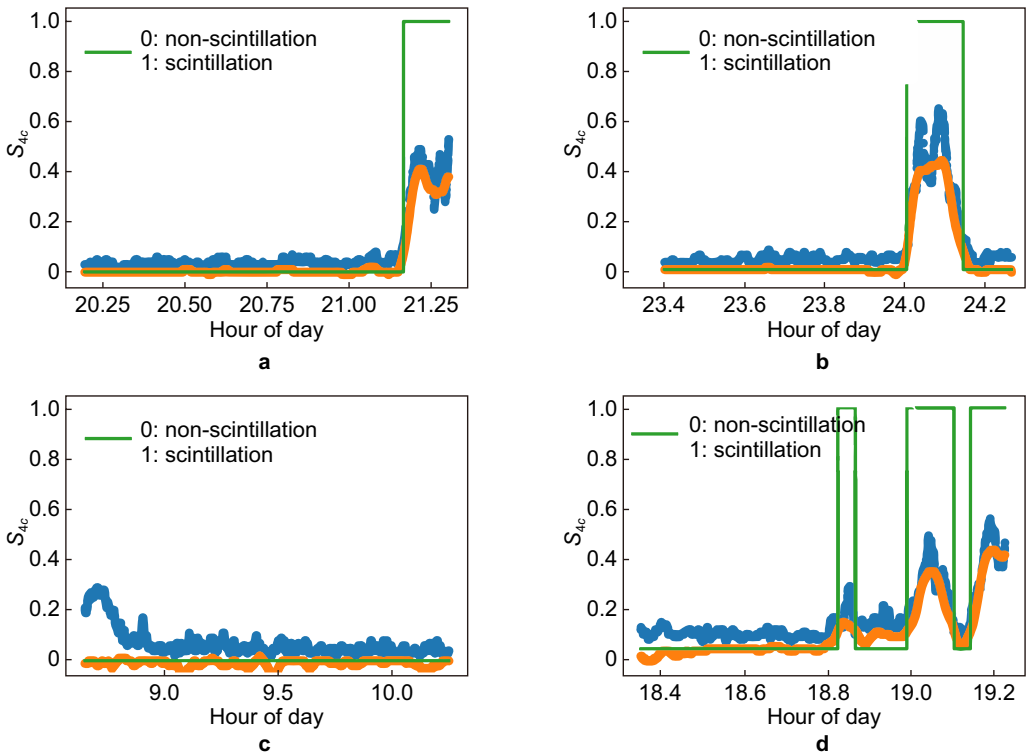

Fig. 10 Decision tree detection results of HKOH on novel data. The blue and orange curves denote the S_{4c} and $\hat{S}_{4c}$, respectively. **a** September 21, PRN 29. **b** September 21, PRN 21. **c** September 21, PRN 20 **d** September 21, PRN 24

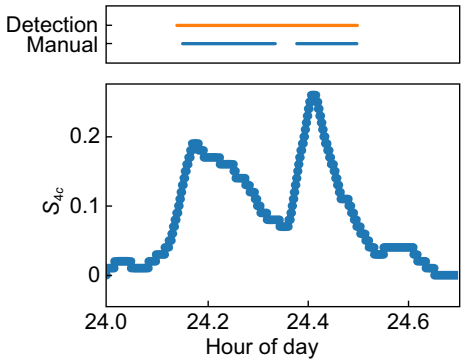

Fig. 11 Comparison of decision tree detection results and manual annotation for PRN 27, November 11

in Fig. 10c. The machine learning approach addresses issues associated with predefined thresholds in hard and semi-hard rules, thereby reducing missed detection rates and enhancing overall accuracy. Traditional approaches, reliant on fixed thresholds, might inaccurately exclude the points as non-scintillation when $\hat{S}_{4c}$ values decrease below the threshold. In contrast, the machine learning algorithm demonstrates an understanding of the presence of scintillation events, encompassing the transient time before and after the strong phase. For instance, in Fig. 10b and d, the machine learning approach accurately classifies the rising and falling edges of the weak, medium, and strong scintillation events. Overall, through visual inspection, the pre-trained decision tree algorithm demonstrates the ability to capture scintillation events and make correct classifications.

In Fig. 11, a certain number of the points not accurately detected by the decision tree algorithm is depicted. Some false positives occur at the onset of the scintillation, suggesting an early detection capability of the decision tree algorithm. Note that this does not necessarily indicate a precise prediction of the scintillation event, as instances of later detection are also observed. The discrepancy between decision tree detection and manual annotation of the scintillation onset may be attributed to human errors during manual annotation, influencing the quality of the training dataset used for the decision tree algorithm. In addition to false positives at the beginning, occurrences of false positives also manifest in the middle of the scintillation period. Here, the decision tree algorithm treats the entire period between 24:09:00 and 24:30:00 as a single scintillation event, contrary to the manual annotation which identifies it as two distinct scintillation events. Although the S_{4c} values initially drop below the threshold commonly associated with scintillation, they subsequently rise after a brief interval, suggesting that the later scintillation event may actually be a continuation of the preceding prolonged event. In general, these false positives are not serious failures since they might be caused by the carelessness in the visual manual inspection or ambiguous situations.

Conclusion

This paper introduces an alternative methodology for the detection of amplitude scintillation utilizing common geodetic GNSS receivers. The detection process utilizes a machine learning decision tree algorithm, capable of learning from historical pre-classified data and making informed decisions on new data. The input to the detection algorithm comprises $\hat{S}_{4c}$ with multipath effects reduced, along with satellite elevation angle and C/N_0 information. Extensive scintillation data including strong, medium, and weak scintillation events are collected to facilitate the training and testing of the decision tree detector. The results demonstrates the superior performance of this detector, surpassing state-of-the-art techniques in terms of accuracy, precision, recall, and F-score. Moreover, tests on novel data confirm its efficacy, reaching levels comparable to manual human-driven annotation. Taking advantage of the widely applied geodetic GNSS receivers, this method has a great potential for ionospheric research and space weather monitoring.

Acknowledgements

The work described in this paper was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 25202520; 15214523) and the National Natural Science Foundation of China (Grant No. 42004029). We appreciate Prof. Zhizhao Liu at the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University for helpful guidance and the provision of ISMR data.

Author contributions

Wang Li conceived of the presented idea. Wang Li, Wenqiang Wei, and Hongyuan Ji performed the research. Yiping Jiang supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding

The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 25202520; 15214523) and the National Natural Science Foundation of China (Grant No. 42004029).

Availability of data and materials

The GNSS dataset analyzed in this study are available in the Hong Kong Geodetic Survey Services website (https://www.geodetic.gov.hk/en/index.htm).

Declarations

Competing interests

The authors declare no conflict of interest.

Received: 12 January 2024 Accepted: 12 April 2024
Published online: 03 June 2024

References

Abadi, P., Saito, S., & Srigutomo, W. (2014). Low-latitude scintillation occurrences around the equatorial anomaly crest over Indonesia. *Annales Geophysicae, 32*, 7–17.

Adewale, A. O., Oyeyemi, E. O., Adeloye, A. B., Mitchell, C. N., Rose, J. A. R., & Cilliers, P. J. (2012). A study of L-band scintillations and total electron content at an equatorial station, Lagos Nigeria. *Radio Science, 47*(2), 1–12. https://doi.org/10.1029/2011rs004846

Akhoondzadeh, M. (2016). Decision tree, bagging and random forest methods detect TEC seismo-ionospheric anomalies around the time of the Chile, (Mw=8.8) earthquake of 27 February 2010. *Advances in Space Research, 57*(12), 2464–2469. https://doi.org/10.1016/j.asr.2016.03.035

Choi, K., Bilich, A., Larson, K. M., & Axelrad, P. (2004). Modified sidereal filtering: Implications for high-rate GPS positioning. *Geophysical Research Letters*. https://doi.org/10.1029/2004gl021621

Curran, J.T., Bavaro, M., Morrison, A., & Fortuny, J. (2014). Developing a multi-frequency for GNSS-based scintillation monitoring receiver. In: Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014), Tampa, Florida, pp. 1142–1152. https://www.ion.org/publications/abstract.cfm?articleID=12266

Enge, P. K. (1994). The global positioning system: Signals, measurements, and performance. *International Journal of Wireless Information Networks, 1*, 83–105. https://doi.org/10.1007/BF02106512

Favenza, A., Farasin, A., Linty, N., & Dovis, F. (2017). A machine learning approach to GNSS scintillation detection: Automatic soft inspection of the events. In: *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017)* (pp. 4103–4111). Portland, Oregon: Institute of Navigation. https://doi.org/10.33012/2017.15351

Franzese, G., Linty, N., & Dovis, F. (2020). Semi-supervised GNSS scintillations detection based on deepinfomax. *Applied Sciences, 10*(1), 381. https://doi.org/10.3390/app10010381

Han, Y., Wang, L., Fu, W., Zhou, H., Li, T., & Chen, R. (2022). Machine learning-based short-term GPS TEC forecasting during high solar activity and magnetic storm periods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 115–126. https://doi.org/10.1109/jstars.2021.3132049

Imam, R., Alfonsi, L., Spogli, L., Cesaroni, C., & Dovis, F. (2023). On estimating the phase scintillation index using TEC provided by ISM and IGS professional GNSS receivers and machine learning. *Advances in Space Research*. https://doi.org/10.1016/j.asr.2023.07.039

Jiao, Y., Hall, J., & Morton, Y. J. (2016). Performance evaluations of an equatorial GPS amplitude scintillation detector using a machine learning algorithm. In: Proceedings of the 29th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2016), pp. 195–199. Institute of Navigation, Portland, Oregon. https://doi.org/10.33012/2016.14554

Jiao, Y., Hall, J. J., & Morton, Y. T. (2017). Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm. *IEEE Transactions on Aerospace and Electronic Systems, 53*(1), 405–418. https://doi.org/10.1109/taes.2017.2650758

Jiao, Y., & Morton, Y. T. (2015). Comparison of the effect of high-latitude and equatorial ionospheric scintillation on GPS signals during the maximum of solar cycle 24. *Radio Science, 50*(9), 886–903. https://doi.org/10.1002/2015rs005719

Kai, G., Yan, Z., Yang, L., Jinling, W., Chunx, Z., & Yanbo, Z. (2017). Study of ionospheric scintillation characteristics in Australia with GNSS during 2011–2015. *Advances in Space Research, 59*(12), 2909–2922.

Kintner, P. M., Ledvina, B. M., & De Paula, E. R. (2007). GPS and ionospheric scintillations. *Space Weather, 5*(9), 1–23. https://doi.org/10.1029/2006sw000260

Kintner, P. M., Ledvina, B. M., De Paula, E. R., & Kantor, I. J. (2004). Size, shape, orientation, speed, and duration of GPS equatorial anomaly scintillations. *Radio Science, 39*(2), 1–23. https://doi.org/10.1029/2003rs002878

Kuruva, L., Avula, M. R., & Achanta, D. S. (2024). Detection of GNSS ionospheric scintillations in multiple directions over a low latitude station. *Journal of Applied Geodesy*. https://doi.org/10.1515/jag-2023-0076

Lee, J., Morton, Y. T. J., Lee, J., Moon, H.-S., & Seo, J. (2017). Monitoring and mitigation of ionospheric anomalies for GNSS-based safety critical systems: A review of up-to-date signal processing techniques. *IEEE Signal Processing Magazine, 34*(5), 96–110. https://doi.org/10.1109/MSP.2017.2716406

Linty, N., Farasin, A., Favenza, A., & Dovis, F. (2019). Detection of GNSS ionospheric scintillations based on machine learning decision tree. *IEEE Transactions on Aerospace and Electronic Systems, 55*(1), 303–317. https://doi.org/10.1109/taes.2018.2850385

Lin, M., Zhu, X., Hua, T., Tang, X., Tu, G., & Chen, X. (2021). Detection of ionospheric scintillation based on XGBoost model improved by SMOTE-ENN technique. *Remote Sensing, 13*(13), 2577. https://doi.org/10.3390/rs13132577

Luo, X., Gu, S., Lou, Y., Cai, L., & Liu, Z. (2020). Amplitude scintillation index derived from C/N0 measurements released by common geodetic GNSS receivers operating at 1 Hz. *Journal of Geodesy*. https://doi.org/10.1007/s00190-020-01359-7

Motella, B., Pini, M., & Dovis, F. (2008). Investigation on the effect of strong out-of-band signals on global navigation satellite systems receivers. *GPS Solutions, 12*(2), 77–86. https://doi.org/10.1007/s10291-007-0085-5

Ott, E. (1978). Theory of Rayleigh-Taylor bubbles in the equatorial ionosphere. *Journal of Geophysical Research: Space Physics, 83*(A5), 2066–2070. https://doi.org/10.1029/JA083iA05p02066

Pi, X., Mannucci, A. J., Lindqwister, U. J., & Ho, C. M. (1997). Monitoring of global ionospheric irregularities using the worldwide GPS network. *Geophysical Research Letters, 24*(18), 2283–2286. https://doi.org/10.1029/97gl02273

Seif, A., Liu, J., Mannucci, A. J., Carter, B. A., Norman, R., Caton, R. G., & Tsunoda, R. T. (2017). A study of daytime L-Band scintillation in association with sporadic e along the magnetic dip equator. *Radio Science, 52*(12), 1570–1577. https://doi.org/10.1002/2017rs006393

Seo, J., Walter, T., & Enge, P. (2011). Correlation of GPS signal fades due to ionospheric scintillation for aviation applications. *Advances in Space Research, 47*(10), 1777–1788. https://doi.org/10.1016/j.asr.2010.07.014

Spogli, L., Cesaroni, C., Di Mauro, D., Pezzopane, M., Alfonsi, L., Musicó, E., Povero, G., Pini, M., Dovis, F., Romero, R., Linty, N., Abadi, P., Nuraeni, F., Husin, A., Le Huy, M., Lan, T. T., La, T. V., Pillat, V. G., & Floury, N. (2016). Formation of ionospheric irregularities over southeast Asia during the 2015 St. Patrick's day storm. *Journal of Geophysical Research: Space Physics, 121*(12), 12211–12233. https://doi.org/10.1002/2016ja023222

Taylor, S., Morton, Y., Jiao, Y., & Triplett, J. (2012). An improved ionosphere scintillation event detection and automatic trigger for a GNSS data collection system. In: Proceedings of the 2012 International Technical Meeting of The Institute of Navigation, Newport Beach, pp. 1563–1569. https://www.ion.org/publications/abstract.cfm?articleID=10034

Van Dierendonck, A. J., Klobuchar, J., & Hua, Q. (1993). Ionospheric scintillation monitoring using commercial single frequency C/A code receivers. In: Proceedings of the 6th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1993), Salt Lake City, pp. 1333–1342. https://www.ion.org/publications/abstract.cfm?articleID=4318

Vila-Valls, J., Closas, P., Fernandez-Prades, C., & Curran, J. T. (2018). On the mitigation of ionospheric scintillation in advanced GNSS receivers. *IEEE Transactions on Aerospace and Electronic Systems, 54*(4), 1692–1708. https://doi.org/10.1109/taes.2018.2798480

Vilà-Valls, J., Linty, N., Closas, P., Dovis, F., & Curran, J. T. (2020). Survey on signal processing for GNSS under ionospheric scintillation: Detection, monitoring, and mitigation. *NAVIGATION, 67*(3), 511–536. https://doi.org/10.1002/navi.379

Yeh, K. C., & Liu, C.-H. (1982). Radio wave scintillations in the ionosphere. *Proceedings of the IEEE, 70*(4), 324–360. https://doi.org/10.1109/proc.1982.12313

Publisher's Note