Distinguishing Ionospheric Scintillation from Multipath in GNSS Signals Using Bagged Decision Trees Algorithm

Distinguishing Ionospheric Scintillation from Multipath in GNSS Signals Using Bagged Decision Trees Algorithm

Rayan Imam
dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
rayan.imam@polito.it

Fabio Dovis
dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
fabio.dovis@polito.it

***Abstract*—This paper presents a machine learning model able to distinguish between ionospheric scintillation and multipath in GNSS–based scintillation monitoring data. The inputs to the model are the average signal intensity, the variance in the signal intensity, and the covariance between the in–phase and the quadrature–phase outputs of the tracking loop of a GNSS receiver. The model labels the data as either scintillated, multipath affected, or clean GNSS signal. The overall accuracy of the model is 96% with 2% miss-detection rate and 0.0% false alarm rate for the scintillation class in particular. The gain in scintillation data is up to 17.5% that would have been discarded if an elevation mask of 30° was implemented.**

***Index Terms*—GNSS, ionospheric scintillation, multipath, machine learning, bagged decision trees**

I. Introduction

Global Navigation Satellite Systems (GNSS) provide a valuable tool for studying the upper layers of the atmosphere. GNSS trans–ionospheric signals, transmitted from approximately 20,000 km above the ground, carry some information about the propagation environment they pass through [1]. Thus, they have been utilized as signals of opportunity for characterizing these environments [2]. The ionospheric studies in particular benefited from these signals in monitoring the ionospheric irregularities near the magnetic equator and the magnetic poles [3].

When radio wave signals pass through disturbed ionosphere, the signal intensity and phase fluctuate as a result of reflection and refraction due to the irregularities in electrons densities [4]. These fluctuations are called ionospheric scintillations, and they are detectable and measurable. Ionospheric scintillation are a threat to GNSS service, leading to degraded performance and reliability of the positioning and timing services provided by GNSS systems [5]–[7].

Two indices are common for measuring scintillations: the S_4 index which measures the standard deviation of the signal intensity normalized by the average signal intensity, and the σ_ϕ index which measures the standard deviation of the detrended signal phase. A scintillation event is declared present on a signal if these indices pass pre-defined thresholds. These thresholds are not standard, but in general, 0.25 for S_4 and 0.15 for σ_ϕ are acceptable values [8].

However, these indices are triggered by causes other than scintillation, which make them suffer from false alarms. Multipath which results from signal reflection and refraction by the environment around the receiver is the most common and most dangerous because it inflates S_4 index in a similar way to scintillation [9]. In the literature, there are few techniques to detect multipath in scintillation monitoring data. The first, which is the most reliable, depends on characterizing the environment around the scintillation monitoring receiver and thus mapping the sources of multipath and consequenctly excluding the data contaminated by these sources [9], [10]. This method of course is location-dependent and have to be repeated for every monitoring station. The second technique relies on eliminating data below an elevation threshold where most data with potential multipath contamination exist. This method has two drawbacks: i) it results in reduced visibility of the sky [11] and thus reduces the amount of useful scintillation data up to 35-45% [10], and ii) it actually does not guarantee multipath free samples by assuming all the signals above the elevation threshold are multipath free. A common elevation threshold is 30° and as will be shown in the results section, sometimes multipath signals exceed that threshold resulting in false scintillation alarms. In this paper, we target those two drawbacks by developing a machine learning model, able to distinguish between multipath and scintillation in scintillation monitoring data.

The use of machine learning techniques to predict near–future ionospheric scintillation occurrences was pioneered in 2010 [12]. The proposed machine learning model relied on data from GNSS receivers and Digital Portable Sound (DPS) Digisondes installed in Brazil to forecast the ionospheric scintillation index S_4 for up to one day ahead. The bagged regression tree model developed also takes as inputs: i) the time of the day, ii) Kp data from World Data Center for Geomagnetism in Kyoto, and iii) the F10.7 solar flux with a resolution of 1 day obtained from National Geophysical Data Center. Since then, many articles investigated developing

machine learning based models to predict ionospheric scintillations based on GNSS measurements and/or non–GNSS based measurements (e.g [13]–[15]).

On the other hand, detecting ionoshperic scintillations is of great interest for GNSS–based scintillation monitoring activities. The ability to detect scintillation based on GNSS measurements alone is particularly favored because it eliminates the need of installing multiple instruments and/or for aiding the instruments with external data. In [16] and [17] support vector machines were trained to detect amplitude and phase scintillations respectively, while in [18] a decision tree was adopted for the classification task. Both models report high detection accuracy. However, in both models multipath effects were not considered, and thus a drawback of these models is that an elevation mask was used to minimize the effect of multipath but not eliminating it. This also limited the potential of implementing the models in operative scenrios where multipath is for sure expected to exist.

In this paper, we investigate the possibility to distinguish between multipath and scintillation using machine learning models. Approximately 45 hours of data captured in Hanoi, Vietnam in 2015 are exploited to train, test and validate a bagged trees model. We will detail the methodology and show the testing results of the model. Also we will show two operative examples to demonstrate the model ability to detect scintillations at low elevations, and also to detect multipath samples at the same elevations.

The rest of this paper is organized as follows. In section II we describe the data used in this work. In section III we detail the machine learning training and the obtained model. In section IV we present the results of testing the model. We conclude the paper in section V by summarizing the results.

II. Data Preparation

In this section we will describe the data we used in this article. We will focus on how this data was prepared for the machine learning task.

A. The data set

The data utilized in this work were collected in March, July and October 2015 in Hanoi, Vietnam as part of the ERICA project (Equatorial Ionosphere Characterization in Asia [19], [20]).

The data collection contains days with strong, moderate and weak–to–no scintillation. Moreover, some of the data collections coincide with days where geomagnetic storms were recorded.

Figure 1 shows the data collection we are using in this paper; giving the dates and local time of the files in addition to the scintillation strength. Each file is a 2 hours long recording of the GNSS spectrum. From the table we can see that the data of April are mainly strong and moderate scintillations, while the data of July and and October are mainly weak–to–no scintillation. We extracted scintillation samples from April data, while the clean and multipath signals were extracted from July and October for the reason to be clarified in the labeling methodology.

Data	Local Time (UTC+7)	Geom. event	Scint. strenght
April 06, 2015	19:28:11	--	strong
April 06, 2015	21:57:22	--	strong/moderate
April 07, 2015	00:32:16	--	low
April 09, 2015	20:24:04	✓	moderate/low
April 09, 2015	23:48:43	✓	low
April 13, 2015	20:40:22	--	strong
April 14, 2015	00:12:04	--	moderate
April 22, 2015	21:36:49	--	strong
April 23, 2015	20:18:05	--	strong/moderate
July 21, 2015	18:00:00	✓	low
July 22, 2015	18:00:00	✓	low
July 23, 2015	18:00:00	✓	low
October 05, 2015	06:00:00	--	low
October 06, 2015	07:00:00	--	low
October 07, 2015	07:00:00	--	low
October 07, 2015	18:00:00	--	low
October 08, 2015	06:00:00	--	low

Scintillation signals were extracted from these days

Multipath and clean signals were extracted from these days

Fig. 1. Selection of the data sets from Hanoi (Vietnam) data collection.

B. Data Processing and Labeling

We used a software defined GNSS receiver (SDR) to process these files and extract the post-correlation outputs of the tracking loop, i.e. the in–phase I and quadrature-phase Q outputs, at a rate of $50Hz$. From I and Q we are able to estimate the signal intensity SI, calculate the scintillation metric S_4, and the other metrics we need for training the machine learning model.

To label the data we use the following rules. We compared data from consecutive days for a PRN to check for multipath. If we see S_4 is inflated and is taking a similar trend in multiple days, this is an indication of multipath. If S_4 is inflated without sidereal repetition, we label it as scintillation. Thus, in the labeled data we only included data we could confidently label by comparing consecutive days. Finally, we label a sample as clean if the S_4 index is not inflated.

We have a total of 14.5 hours of data for each class: multipath, scintillation and clean. We used 10 hours for training and the remaining 4.5 hours for testing the models.

C. Machine Learning Attributes

From the definition of scintillation as fluctuations in signal intensity (and phase) we decided to use features that can capture fluctuations in general as inputs to train the machine learning algorithm. Thus we choose the average signal intensity, the variance in the signal intensity and the covariance between I and Q over a time window. The instantaneous signal intensity SI_i at time i is given by:

$$SI_i = I_i^2 + Q_i^2 \tag{1}$$

where I_i and Q_i are the in phase and quadrature phase components of the signal. The average value of SI is calculated from SI_i values over the previous time window of N samples as:

$$\overline{SI} = \frac{1}{N}\sum_{i=1}^{N} SI_i \tag{2}$$

Similarly, the variance of SI is calculated over the same time window as:

$$\sigma_{SI} = \frac{1}{N-1}\sum_{i=1}^{N}(SI_i - \overline{SI})^2 \quad (3)$$

In the case of the covariance, we choose to calculate the covariance between I^2 and Q^2 instead of I and Q because of the data bits that we have not removed from the received signal, and thus the covariance was calculated as:

$$cov(I^2, Q^2) = \frac{1}{N-1}\sum_{i=1}^{N}(I_i^2 - \overline{I^2})(Q_i^2 - \overline{Q^2}) \quad (4)$$

where $\overline{I^2}$ and $\overline{Q^2}$ are the average values of I^2 and Q^2 calculated over the same time window.

The choice of the averaging time window was experimentally investigated by preparing the training set many times, each one using a different averaging time window. The exact values of the averaging windows are $\{1,10,30,60,90,180\}$ s, where 60 s is the typical time window used in literature to calculate S_4 while 180 s is the typical duration of the shortest scintillation event used in literature. So, we envestigated the averaging windows around those two values. We finally down–sampled the data from $50Hz$ to $1Hz$.

III. The Machine Learning Model

In this section we will briefly describe the bagged trees concept. Then we will present the methodology we used to train a machine learning model able to distinguish between clean, scintillated and multipath affected GNSS signals. We will then introduce the trained model and its performance during the training phase.

A. Bagged Trees

The term Bagged trees comes from Bootstrap aggregation. It was introduced in 1996 [21] to reduce the variance of statistical learning methods, by building separate prediction models using subsets of the training set, i.e. bootstrapping the data. Then the average of all the models outputs in case of regression, or the voting result in case of classification, is taken as the final output of the model. Since each sub–model is trained using a subset of the training data, the remaining data of each subset (referred to as out-of-bag OOB observations) are used for validating that trained sub-model. This way, bagged trees always increase the classification trees accuracy, but on the other hand we loose the interpretability advantage of decision trees.

B. The Trained Model

We trained six bagged trees with 30 sub–trees each to distinguish between scintillated, multipath affected and clean GNSS signals. Each of these models was trained using one of the data sets that was prepared with the six averaging values explained in the previous section. Figure 2 shows the total accuracy (up) and the miss–detection rate of the scintillation class (bottom) for the different values of the averaging window.

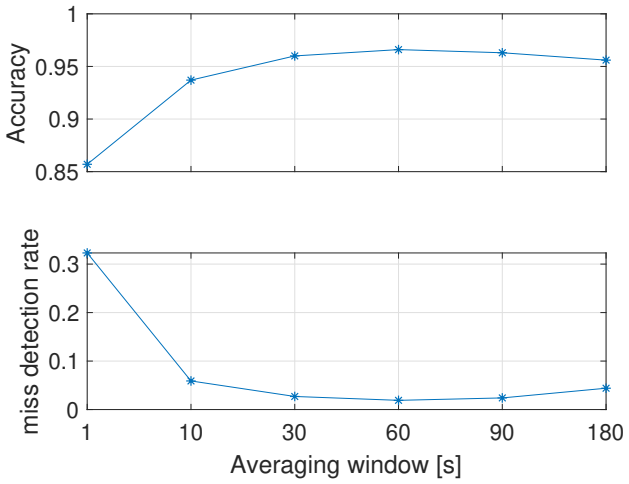

Fig. 2. Accuracy of the model when trained with data with different averaging time windows.

These accuracies and miss-detection rates were calculated from testing the model using the testing data set. Accuracy of a model is the ratio of the number of the correctly classified samples to the total number of samples. The miss–detection rate of a class is the number of wrongly classified samples of that class to the total number of samples in that class. Finally, the false alarm rate of a class is the number of wrongly classified samples as that class to the total number of samples classified as that class. These metrics are easily extracted from the confusion matrix of the classification results (an example of a confusion matrix is shown in the next section).

In Fig. 2 it can be seen that 30, 60 and 90 s averaging windows gave comparable good performance. We decided to adopt 60 s, which is the value in use in the literature for calculating S_4, for our model to distinguish between scintillation and multipath.

IV. Results

In this section we will show the results of testing the model using pre-labeled data. The first test was conducted using 13.5 hours of data, 4.5 hours for each of the three classes. The objective of showing this result is to quantitatively assess the machine learning model. In the next two tests, we will demonstrate the model capabilities under two scenarios i) detecting scintillation at low elevations, and ii) detecting multipath at above elevation threshold. These two scenarios challenge the model, because the first is the scintillation data that usually get discarded by the elevation angle threshold, while the second is the multipath data that raise false alarms in scintillation monitoring stations.

Figure 3 reports the confusion matrix when testing the model using a novel testing set. It can be seen that the model gave zero false alarm for scintillation. The miss-detection rate of scintillation is particularly low, with only 1.9% of the scintillation data miss-detected. The false alarm for the scintillation class is 0.0%. However, the model wrongly

classified multipath as clean signals with miss detection rate of 7.4%. Since our objective is to detect scintillation, and to distinguish scintillation from multipath in particular, this mix between the clean and multipath signals was acceptable for this work, and further investigating it can be objective of a future work. Overall, the testing of the model with novel data verifies the trained model, and thus we can rely on it to distinguish multipath from scintillation in operative scenarios.

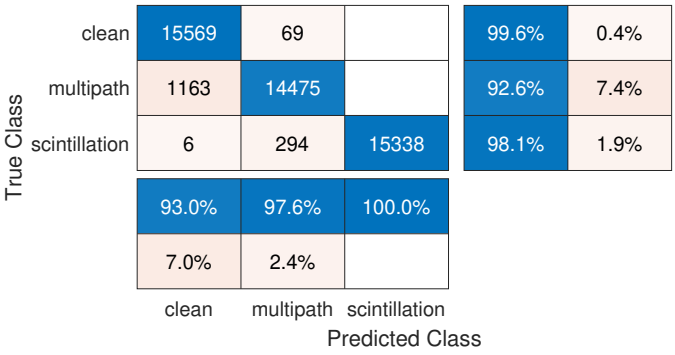

Fig. 3. The confusion Matrix of the testing dataset.

Figure 4 show the results of detecting scintillation at low elevation. Here, the signal intensity is rapidly fluctuating (2nd plot from top), leading to high S_4 (4th plot from top). However, the elevation of this signal is below 30° (bottom plot) and thus this data will probably be discarded by a scintillation monitoring station that relies on elevation masks to avoid multipath. Our model correctly classifies this data as scintillation as shown in the top plot of the figure, with a couple of miss–classified signals as multipath. Using our model, scintillation monitoring could gain more scintillation data by including low elevation measurements. In fact, in our testing dataset, by including low elevation data, we gained 17.5% more scintillation data that would have been discarded by a 30° elevation mask.

In Fig. 5 we show an example of a multipath signal. Here, the S_4 index is inflated (second plot from bottom) due to multipath up to an elevation angle of 35° (bottom plot). In fact, our model suspects all the data up to 14:45 UTC to be multipath affected (top plot), which means the slight enhancement on S_4 at 40° elevation is due to multipath. Using the model will thus reduce false scintillation alarms. This in return will lead to better automation of scintillation monitoring activities and better management of scintillation data repositories.

V. Conclusion

In this paper, we showed that machine learning models can be trained to distinguish between scintillation, multipath and clean GNSS signals. We trained a bagged trees models to distinguish between these classed based on features derived from GNSS receivers post–correlator outputs. The features we used are the average signal intensity, the variance in the signal intensity, and the covariance between the in–phase

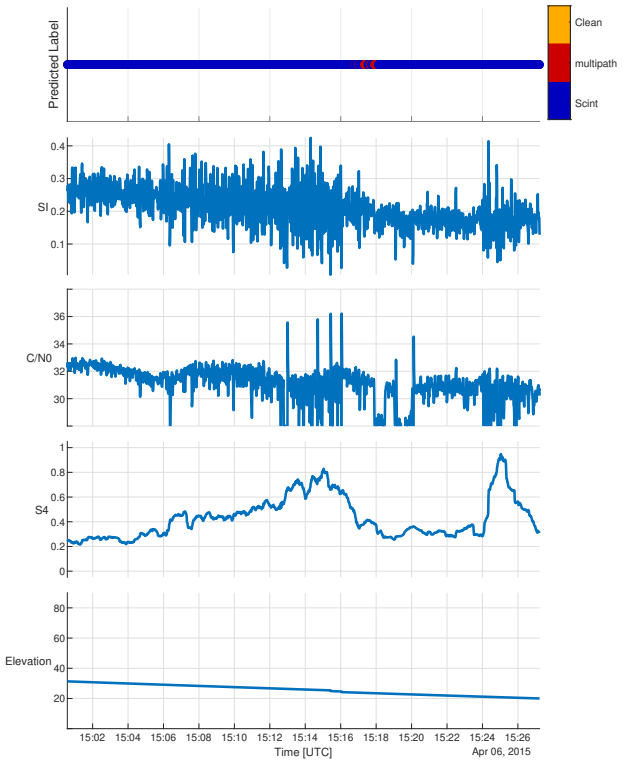

Fig. 4. Detecting Scintillation at Low Elevation.

and the quadrature–phase components of the GNSS signal at the correlator outputs. The averaging window was also investigated, and found that averaging windows between 30-90 s give comparable good performances, and consequenctly 60 s was adopted. The accuracy of the model developed is 96% with miss–detection rate on the scintillation class less than 2% and 0.0% false alarms when testing with a novel data set. In this dataset, 17.5% scintillation data gain was observed, resulting from including scintillation data that would have been discarded by a 30° elevation mask. Finally, two examples of testing the model in operative scenrios were shown, one for detecting scintillation at low elevations, and the other for identifying multipath signals that would have triggered a false alarm for a 30° elevation threhold system.

References

[1] F. Dovis, *GNSS Interference Threats and Countermeasures*. Artech House, 2015.

[2] D. Gebre-Egziabher and S. Gleason, *GNSS applications and methods*. Artech House, 2009.

[3] X. Pi, A. Mannucci, U. Lindqwister, and C. Ho, "Monitoring of global ionospheric irregularities using the worldwide gps network," *Geophysical Research Letters*, vol. 24, no. 18, pp. 2283–2286, 1997.

[4] Kung Chie Yeh and Chao-Han Liu, "Radio wave scintillations in the ionosphere," *Proceedings of the IEEE*, vol. 70, no. 4, pp. 324–360, 1982.

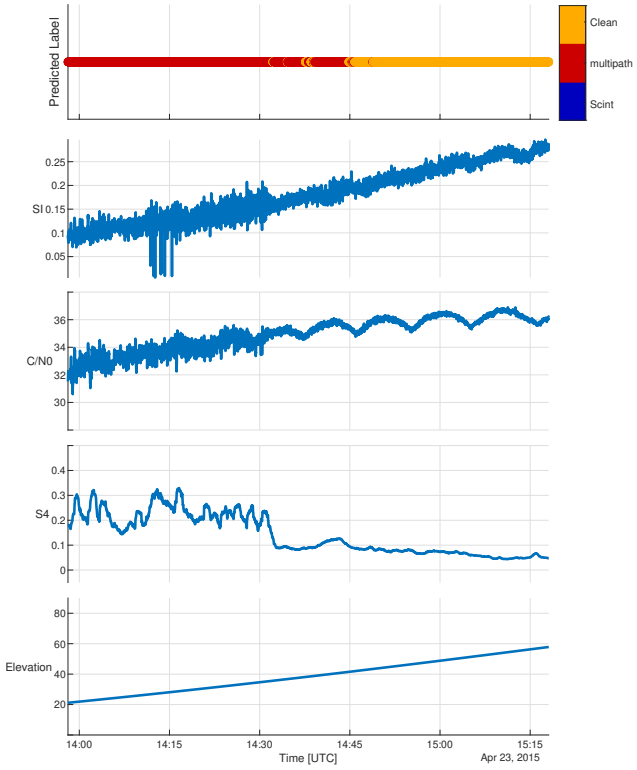

Fig. 5. Detecting Multipath at High Elevation.

[5] P. M. Kintner, B. M. Ledvina, and E. De Paula, "Gps and ionospheric scintillations," *Space weather*, vol. 5, no. 9, 2007.

[6] J. Lee, Y. T. J. Morton, J. Lee, H. Moon, and J. Seo, "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*, vol. 34, no. 5, pp. 96–110, 2017.

[7] N. Linty, A. Minetto, F. Dovis, and L. Spogli, "Effects of phase scintillation on the gnss positioning error during the september 2017 storm at svalbard," *Space Weather*, vol. 16, no. 9, pp. 1317–1329, 2018.

[8] L. Alfonsi, L. Spogli, G. De Franceschi, V. Romano, M. Aquino, A. Dodson, and C. N. Mitchell, "Bipolar climatology of gps ionospheric scintillation at solar minimum," *Radio Science*, vol. 46, no. 3, 2011. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010RS004571

[9] O. Olwendo *et al.*, "Elimination of superimposed multipath effects on scintillations index on solar quiet ionosphere at low latitude over the kenyan airspace from a lone positioned scinda system." Proceedings of the 23rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2010), 2010.

[10] L. Spogli, V. Romano, G. D. Franceschi, L. Alfonsi, E. Plakidis, C. Cesaroni, M. Aquino, A. Dodson, J. F. G. Monico, and B. Vani, "A filtering method developed to improve gnss receiver data quality in the calibra project," in *Mitigation of Ionospheric Threats to GNSS*, R. Notarpietro, F. Dovis, G. D. Franceschi, and M. Aquino, Eds. Rijeka: IntechOpen, 2014, ch. 9. [Online]. Available: https://doi.org/10.5772/58778

[11] N. Linty, F. Dovis, and L. Alfonsi, "Software-defined radio technology for gnss scintillation analysis: Bring antarctica to the lab," *GPS Solutions*, vol. 22, no. 4, p. 96, 2018.

[12] L. F. C. Rezende, E. R. de Paula, S. Stephany, I. J. Kantor, M. T. A. H. Muella, P. M. de Siqueira, and K. S. Correa, "Survey and prediction of the ionospheric scintillation using data mining techniques," *Space Weather*, vol. 8, no. 6, 2010. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009SW000532

[13] G. R. T. de Lima, S. Stephany, E. R. de Paula, I. S. Batista, and M. A. Abdu, "Prediction of the level of ionospheric scintillation at equatorial latitudes in brazil using a neural network," *Space Weather*, vol. 13, no. 8, pp. 446–457, 2015. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015SW001182

[14] R. M. McGranaghan, A. J. Mannucci, B. Wilson, C. A. Mattmann, and R. Chadwick, "New capabilities for prediction of high-latitude ionospheric scintillation: A novel approach with machine learning," *Space Weather*, vol. 16, no. 11, pp. 1817–1846, 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018SW002018

[15] K. Lamb, G. Malhotra, A. Vlontzos, E. Wagstaff, A. G. Baydin, A. Bhiwandiwalla, Y. Gal, A. Kalaitzis, A. Reina, and A. Bhatt, "Prediction of gnss phase scintillations: A machine learning approach," *arXiv preprint arXiv:1910.01570*, 2019.

[16] Y. Jiao, J. J. Hall, and Y. T. Morton, "Automatic equatorial gps amplitude scintillation detection using a machine learning algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 405–418, 2017.

[17] Y. Jiao, J. J. Hall, and Y. T. Morton, "Performance evaluation of an automatic gps ionospheric phase scintillation detector using a machine-learning algorithm," *NAVIGATION: Journal of the Institute of Navigation*, vol. 64, no. 3, pp. 391–402, 2017.

[18] N. Linty, A. Farasin, A. Favenza, and F. Dovis, "Detection of gnss ionospheric scintillations based on machine learning decision tree," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 303–317, 2019.

[19] G. Povero, M. Pini, F. Dovis, R. Romero, P. Abadi, L. Alfonsi, L. Spogli, D. Di Mauro, Le Huy Minh, La The Vinh, and N. Floury, "Ionosphere monitoring in south east asia: Activities in ginestra and erica projects," in *2015 International Association of Institutes of Navigation World Congress (IAIN)*, 2015, pp. 1–7.

[20] L. Spogli, C. Cesaroni, D. Di Mauro, M. Pezzopane, L. Alfonsi, E. Musicò, G. Povero, M. Pini, F. Dovis, R. Romero, N. Linty, P. Abadi, F. Nuraeni, A. Husin, M. Le Huy, T. T. Lan, T. V. La, V. G. Pillat, and N. Floury, "Formation of ionospheric irregularities over southeast asia during the 2015 st. patrick's day storm," *Journal of Geophysical Research: Space Physics*, vol. 121, no. 12, pp. 12,211–12,233, 2016. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JA023222

[21] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.