POLITECNICO DI TORINO
Repository ISTITUZIONALE

Detection of GNSS Ionospheric Scintillations based on Machine Learning Decision Tree

Original
Detection of GNSS Ionospheric Scintillations based on Machine Learning Decision Tree / Linty, Nicola; Farasin, Alessandro; Favenza, Alfredo; Dovis, Fabio. - In: IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS. - ISSN 0018-9251. - ELETTRONICO. - 55:1(2019), pp. 303-317. [10.1109/TAES.2018.2850385]

Availability:
This version is available at: 11583/2712391 since: 2019-03-27T09:57:15Z

Publisher:
Institute of Electrical and Electronics Engineers Inc.

Published
DOI:10.1109/TAES.2018.2850385

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 December 2025

Detection of GNSS Ionospheric Scintillations Based on Machine Learning Decision Tree

NICOLA LINTY, Member, IEEE
Department of Electronics and Telecommunications Politecnico di Torino, Torino Italy

ALESSANDRO FARASIN
ALFREDO FAVENZA
Istituto Superiore Mario Boella, Torino Italy

FABIO DOVIS
Department of Electronics and Telecommunications Politecnico di Torino, Torino Italy

This paper proposes a methodology for automatic, accurate, and early detection of amplitude ionospheric scintillation events, based on machine learning algorithms, applied on big sets of 50 Hz postcorrelation data provided by a global navigation satellite system receiver. Experimental results on real data show that this approach can considerably improve traditional methods, reaching a detection accuracy of 98%, very close to human-driven manual classification. Moreover, the detection responsiveness is enhanced, enabling early scintillation alerts.

Manuscript received May 31, 2017; revised February 6, 2018 and June 1, 2018; released for publication June 5, 2018. Date of publication June 27, 2018; date of current version February 7, 2019.

DOI. No. 10.1109/TAES.2018.2850385

Refereeing of this contribution was handled by J. Morton.

Authors' addresses: N. Linty and F. Dovis are with the Department of Electronics and Telecommunications, Politecnico di Torino, Torino 10129, Italy, E-mail: (nicola.linty@polito.it; fabio.dovis@polito.it); A. Farasin is with the Istituto Superiore Mario Boella, Torino 10129, Italy, and also with the Department of Control and Computer Engineering, Politecnico di Torino, Torino 10129, Italy, E-mail: (alessandro.farasin@ismb.it); A. Favenza is with the Istituto Superiore Mario Boella, Torino 10129, Italy, E-mail: (alfredo.favenza@ismb.it). *(Corresponding author: Nicola Linty.)*

0018-9251 © 2018 CCBY

I. INTRODUCTION

The propagation through the atmosphere has a significant influence on radio signals broadcast by satellites toward the Earth. Irregularities and gradients of the ionization of the upper layer of the atmosphere, the ionosphere, a region characterized by a high concentration of free electrons, can cause fluctuations of the signal amplitude and phase, which are called *ionospheric scintillations*. Scintillations also affect Global Navigation Satellite System (GNSS) transionospheric signals. Under disturbed ionospheric conditions, GNSS receivers are more subject to phase errors, cycle slips, increased carrier Doppler jitter, and losses of lock, resulting in positioning errors of the order of tens of meters, and, in the most severe cases, even in the complete receiver outage [1]. Scintillations are a threat to GNSSs since they may have disruptive impact on the user receiver performance when high accuracy, reliability, and continuity of the positioning service are needed, as, for example, for critical applications and precise positioning [2]–[4].

As a consequence, scintillation monitoring and detection is a key aspect for improving the quality and reliability of GNSS observations [5]. Networks of GNSS receivers, specifically designed for accomplishing these tasks, have been installed in recent years, both at low and high latitudes, where scintillation is more likely to occur [6], [7]. The purpose is indeed twofold: on one side, observation of the signals themselves, which are a source of information for understanding and modeling the upper layers of the atmosphere [8]; on the other side, the signals can be used as detectors and triggers to raise warning and take countermeasures for GNSS-based operations. For this reason, it is important to design receivers robust to the presence of scintillation, but also to have proper algorithms for the detection of the event and its classification [9].

Early and accurate detection of scintillation events is a very important feature for space weather applications for atmospheric remote sensing and in general for all those data collection systems that automatically detect and record Intermediate Frequency (IF) raw samples [10], [11]. However, the scientific literature about scintillation detection techniques is limited. Most of the works are based on very simple event triggers, which are based on the comparison of scintillation indices provided by commercial receivers with preset threshold values [12]. However, this approach overlooks high-moment characteristics of the signals and requires detrending operations. Some alternatives to traditional scintillation indices were proposed, for instance, exploiting wavelet techniques [13], decomposing the Carrier-to-Noise density power ratio (C/N_0) by means of adaptive time-frequency methods [14] of evaluating statistical properties of the histogram of received samples [15]. The common drawback of such techniques is that they rely on complex and computationally expensive operations or on dedicated receiver architectures.

Recent studies have demonstrated that machine learning techniques can be exploited for scintillation detection. In [16], Rezend *et al.* propose a survey of data mining techniques, relying on observation and integration of GNSS receivers, other sensors and online forecast services. However, this approach relies on external data sources and instruments, which are not always available. A technique based on supervised machine learning Support Vector Machine (SVM) algorithm has been proposed for amplitude scintillation detection in [17] and [18]. This method has been extended to phase scintillation detection in [19] and [20]. The main limitations of these approaches are: they provide a detection trigger at a low rate; they are based on SVM models, which are computationally demanding; and they have been tested on a set of data prefiltered at an elevation mask of 30°; thus, discarding potentially useful and important data.

The work presented in this paper aims at proposing an alternative method for the detection of amplitude scintillation based on machine learning. The scope of this approach is multifold as follows:

1) to propose an alternative to the use of traditional scintillation indices, the performance of which may depend on algorithmic choices, such as detrending and average operations;
2) to use only common GNSS stand-alone receivers observables;
3) to be able to understand the presence of the scintillation event including the transient time before and after its strongest phase, thus providing an early run-time alert;
4) to provide an automatic method, resembling manual observation of the observables, while keeping the cost low in terms of human effort and enabling run-time detection;
5) to reduce the rate of false alarms due to the ambiguity between scintillation and other events, such as multipath, that may affect the assessment of the classical amplitude scintillation index, without the need of prefiltering data;
6) to reduce the missed detection caused by *a priori* filtering of data at low-elevation angle, often implemented to hard-cut multipath effect;
7) to use computationally efficient machine learning algorithms, such decision tree.

The paper is organized as follows. After Section I, which has outlined the scope of the work, Section II provides an overview on scintillation, its effects on GNSS signals and applications, and on machine learning algorithms, models and metrics for performance evaluation. Section III gives an overview of traditional scintillation detection techniques and analyzes their limitations on selected case studies. Section IV introduces machine learning detection, identifying two different sets of features based on different receiver measurements. Section V validates the proposed approaches, presenting quantitative, and qualitative results obtained running different machine learning algorithms on different sets of features. Finally, Section VI draws conclusions and outlines the future work.

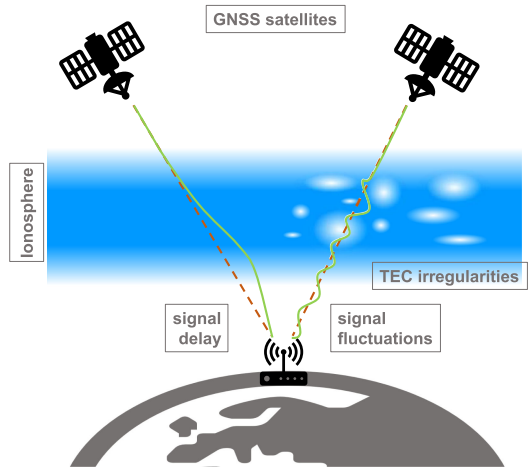

Fig. 1. Pictorial representation of ionospheric delay and scintillation phenomena. The red dashed lines are the line-of-sight signal paths from the GNSS satellites to the receiver on earth; the green continuous signal accounts for propagation distortions.

II. GENERAL OVERVIEW

This section presents a general overview on GNSS, ionospheric scintillation, and machine learning. A reader expert in the field can skip to Section III.

A. GNSS and Ionospheric Scintillations

GNSSs are radio-communication satellite systems that enable a generic user to compute position, velocity, and time at its current location, anywhere on the Earth, processing Radio Frequency (RF) signals transmitted from a constellation of satellites and performing trilateration with respect to the satellites, taken as reference points [21]. Despite being originally developed for localization, GNSSs are not limited to positioning purposes, but span an unlimited range of applications, including scientific observations.

One of the main characteristic of GNSS signals is their low-received power. For this reason, the accuracy, availability, and reliability of the position solution is threatened by potential errors, affecting the overall quality of the process. Ionospheric propagation is the major and more variable natural error source in GNSS signal processing at the receiver level. Propagation through this layer introduces a potentially strong degradation of the GNSS signal, as depicted in Fig. 1, causing significant errors in any GNSS-based application. The ionosphere affects the quality of GNSS signals both in terms of a temporal delay and of scintillations. While delay compensation techniques are nowadays applied in any GNSS receiver [21], scintillations is still an issue for both mass-market and professional devices. Occurrence of scintillations is very difficult to be modeled, due to their quasi-random nature [22]. Therefore, they remain, to this day, one of the major limiting factors for high-accuracy applications.

Scintillation monitoring is indeed a central activity, both in the GNSS and in the space weather community. The amount of scintillation affecting a satellite signal can be evaluated by exploiting the correlation output values. Two

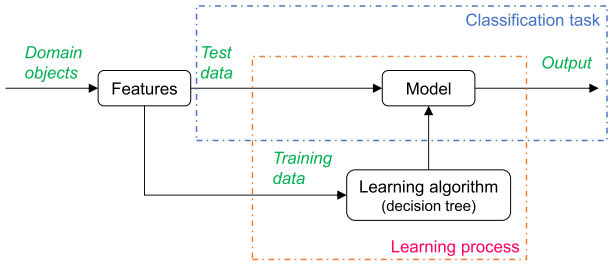

Fig. 2. Flow diagram of the machine learning process. In the first stage, the learning process, depicted in the red box, a model is defined, using training data, selected features and a specific learning algorithm, such as a decision tree. The model is then applied on the test dataset in the classification task, in the blue box, to generate the output.

indices are usually considered: S_4 for amplitude scintillation and σ_ϕ for phase scintillation. S_4 measures the amount of amplitude fluctuations due to scintillations in GNSS signals; it corresponds to the normalized standard deviation of the detrended Signal Intensity (SI) computed from the inphase (I) and quadrature-phase (Q) prompt correlation samples. σ_ϕ is calculated as the standard deviation of the detrended carrier phase measurements. Both indices are calculated over a varying observation interval, usually equal to 60 s.

Most of the works on scintillation monitoring are based on the comparison of the value of these two indices with predefined thresholds. Nevertheless, detection based on such fixed thresholds can be misleading, due to: the loss of the transient phases of the events, causing a delay in the raise of possible warning flags; the missed detection of weak events with high variance; or the signal distortions caused by other phenomena, such as multipath. The only reliable procedure is to entrust the detection of scintillation events to a human-driven visual inspection of data.

B. Machine Learning

Machine learning [23] is the systematic study of intelligent algorithms and systems that improve their knowledge or performance by experience. In its general concept, machine learning process (Fig. 2) refers to the ability of solving a task, processing right features describing the domain of interest, according to a model. The use of machine learning in Location-Based Services is also motivated by the increasing volume of available data collected at remote sites through low cost GNSS software receivers [24].

The problem under investigation in this paper is a typical binary classification task that can be undertaken exploiting a machine learning approach on a big dataset. The main elements of machine learning are [25] as follows:

1) *domain*, the problem to be solved (detection of ionospheric scintillation events in GNSS data collections);
2) *features*, the description of the objects of the domain (GNSS observables);
3) *task*, the abstract representation of the problem that reflects in the mapping between the input and the output (the automatic classification of data collection sample in scintillation/nonscintillation);
4) *model*, the output of the machine learning when the training set is fed to the algorithms.

The machine learning goal is to identify the right algorithm, or set of algorithms, to use the right set of features to build the right models that achieve the right tasks in terms of detection accuracy. This goal is achieved by feeding the machine learning algorithms with two different sets of data, which are as follows:

1) a *training-set* of historical, prelabeled data;
2) a *test-set*, nonlabeled, possibly real-time data.

Machine learning offers a large number of algorithms, to build models from a given dataset of input observations, and to make predictions and decision expressed as output. These algorithms are mainly grouped under three big families, which are as follows:

1) supervised learning, where input data (training set) has a known label or result;
2) unsupervised learning, where input data is not labeled and does not have a known result;
3) semisupervised learning, where input data is a mixture of labeled and unlabeled examples.

Since the goal of this research is classification of data on the basis of the detection of scintillation events and the input training set is fully labeled, this study will take into consideration two types of supervised learning algorithms: decision tree and random forest.

1) *Decision Tree:* Decision tree is one of the most commonly used classification technique [26]. It is based on tree structures, defined by recursively partitioning the input space: each internal node represents a certain feature of the domain; each branch, emanating from the node, is the outcome of the decision taken in the node according to a function; and each leaf represents a final classification decision, corresponding to the conjunction of single decisions taken during on the path from the root of the tree to the leaf. The learning takes place as the machine creates a set of rules defining a model, in terms of sequence of the features along the branches and functions for the decision criteria in each node. The rules are based on the concept of *utility* of a feature for the classification purpose.

Lets consider two classes, D^+ and D^-, and a boolean feature, D_1, which can take the values D_1^+ and D_1^-. The ideal situation is when $D_1^+ = D^+$ and $D_1^- = \emptyset$, or $D_1^+ = \emptyset$ and $D_1^- = D^-$: in this case, the branches are said to be pure. Typically this situation never happens, so the task is to measure the impurity of each feature and corresponding function. Extending the problem to a general case, the goal is to measure the impurity of a set of n^+ positives and n^- negatives, in terms of empirical probability $\dot{p} = n^+/(n^+ + n^-)$, to evaluate the best rule. A cost function is defined to achieve this task, thus, building the tree model. Among several possibilities, such as the minority class or the Gini index, the *entropy* $\mathcal{E}$ was chosen

$$\mathcal{E} = -\dot{p}\log_2(\dot{p}) - (1-\dot{p})\log_2(1-\dot{p}). \quad (1)$$

TABLE I
General Example of a Confusion Matrix

		Prediction	
		0	1
Truth	0	True negatives	False positives
	1	False negatives	True positives

By minimizing the entropy, the information gain brought by the tree is maximized. Further mathematical details on decisions trees can be found in [23], [27], and [28].

2) *Random Forest:* Random forest is an ensemble learning method for classification, based on the construction of a multitude of decision trees at training time [29]. It helps to overcome the problem of overfitting, as well as to reduce the variance of an estimate, exploiting averaging. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large and depends on the strength of the individual trees in the forest and the correlation between them.

3) *Performance Evaluation Metrics:* Performance of a machine learning classification algorithm is typically evaluated using statistical tools and metrics [28]. One of the most common is the confusion matrix, or contingency table; an example is reported in Table I. Each row refers to actual classes as annotated in the test case, each column to classes as predicted by the classifier. Last column and last row give the marginals, which are important to allow the statistical significance assessment. In order to distinguish the performance on the classes, correctly classified positives and negatives are referred to as true positives and true negatives, respectively; incorrect classified positives are called false negatives or missed detections; misclassified negatives are called false positives or false alarms. Consequently, the true positive rate is the proportion of positives correctly classified, also called sensitivity or recall, while the true negative rate is the proportion of negatives correctly classified.

Other metrics that can be used to assess the machine learning performance are as follows:

1) *Accuracy*: The percentage of correct predictions made by the model over a dataset, calculated as
True Positives + True Negatives.
2) *Precision*: The ratio of correct positive observations, calculated as
True Positives/ (True Positives + False Positives).
3) *Recall*: The ratio of correctly predicted positive events, also known as sensitivity, calculated as:
True Positives/ (True Positives + False Negatives).
4) *F-score*: An alternative measure of a test accuracy, taking into account also both false positives and false negatives. It corresponds to the weighted average of precision and recall:
(2 · Recall · Precision) / (Recall + Precision).

4) *k-Fold Cross Validation:* Cross validation is a validation technique, which measures how the results of a statistical analysis will generalize to an independent dataset. In fact, knowledge about the test-set can affect the model and the evaluation metrics decreasing generalization performance. This situation is typically called overfitting. A solution to this problem is a procedure called cross validation. In the basic approach adopted in this paper, denoted k-fold cross validation [30], the training set is split into k smaller sets, called folds. Afterwards, for each of the k folds, first a model is trained, using the remaining $k-1$ folds as training data. The resulting model is then used to test the untrained fold. Finally, the average of the values obtained in each iteration is computed.

III. TRADITIONAL SCINTILLATION DETECTION

In this section, the traditional state-of-the-art approaches for amplitude scintillation detection are presented. Two detection rules are considered and tested on two case studies, evaluating and carefully discussing their performance and limitations.

A. Description of Traditional Methods

The majority of the works on amplitude scintillation monitoring are based on the analysis of the value of the S_4 index. Scintillation is typically considered present if S_4 exceeds a predefined threshold $\mathcal{T}_{S_4}$. Different detection rules can be identified, according to the literature on the topic.

1) Hard Detection: A hard detection rule is defined by simply applying the threshold $\mathcal{T}_{S_4}$ on the estimated value of S_4. Ionospheric scintillation is present at epoch n if and only if

$$S_4[n] > \mathcal{T}_{S_4}. \tag{2}$$

It is a simple approach, easy to be implemented, but it might lead to an undesired and nonnegligible number of false alarms. The threshold is commonly assumed to be $\mathcal{T}_{S_4} = 0.4$ by Dubey *et al.* in [31]–[35]; other works consider scintillation moderate in the range between 0.2 and 0.5, and strong above 0.5 [36].

As the S_4 index is a measure of the variation of the amplitude of the GNSS signal, it is not unlikely that events other than ionospheric scintillation cause it to increase above the threshold, in a way similar to what scintillations do, thus, affecting the detection process. This is particularly frequent for low-elevation satellites, when the number of multipath reflected rays increases and, at the same time, the C/N_0 is lower.

2) Semihard Detection: In order to better characterize the scintillation phenomenon and to reduce the false alarm rate, more filters can be applied to the signal. For example, it is quite common to apply an elevation mask; most of the multipath-induced false alarms can be removed by considering only signals from satellites above a certain elevation angle. Further conditions can be defined on the C/N_0 value, so as to exclude noisy measurements, or on the satellites azimuth ϑ_{az}. According to the semihard detection rule,

scintillation is present at epoch n if and only if

$$S_4[n] > \mathcal{T}_{S_4} \wedge \theta_{\text{el}}[n] > \mathcal{T}_{\theta_{\text{el}}} \wedge C/N_0[n] > \mathcal{T}_{C/N_0}. \quad (3)$$

The value of the elevation threshold $\mathcal{T}_{\theta_{\text{el}}}$ is typically set to 30° [36], [37]. The definition of $\mathcal{T}_{C/N_0}$ is more complex, as the C/N_0 is the result of an estimation process and depends on the receiver implementation. The value 37 dBHz gives satisfactory results [38]. Nevertheless, it has been proven that the filter on the C/N_0 is not very discriminant in terms of detection results; lower values, such as 30 dBHz, which corresponds to the sensitivity of a standard tracking loop, lead to analog results.

3) Manual Detection: A "third" approach corresponds to human-driven manual and subjective identification of the portion of data affected by scintillation. This can be achieved by visual inspection of the S_4 and C/N_0 estimates, of the satellite azimuth and elevation and of the comparison of historical data. Even though this approach lacks scientific rigor, it can assure the best detection performance, provided that the person doing the manual annotation has enough knowledge and experience. However, it is time-consuming, subject to human errors, and not automatic.

In this paper, the manual annotation is considered as the reference *ground truth* for the detection performance analysis. The same approach has been used also in other works relying on machine learning for scintillation detection [17], [18].

B. Case Studies

The examples refer to GNSS data collections performed on March 26 and April 2, 2015, in Hanoi (Vietnam), at 11° 20′ N geo-magnetic latitude, using a customized Software Defined Radio (SDR)-based GNSS data grabber and software receiver [38]. Moderate and strong amplitude scintillation events were observed; the Dst index negative peaks, provided by the World Data Center of Kyoto, amount to −20 nT and −21 nT, respectively. Global Positioning System (GPS) L1 C/A signals are considered, respectively, of Pseudo-Random Noise (PRN) satellite 10 and 23. Fig. 3 shows the estimates of the S_4 index (top plot, black dots), of the C/N_0 and of the satellite elevation (bottom plot, blue and red lines, respectively) as computed by the software receiver. In addition, three horizontal lines are drawn in correspondence of the S_4, C/N_0 and elevation thresholds, used in the detection rules.

Traditional detection rules for scintillation monitoring are applied as follows.

1) All points for which scintillation is detected according to the hard rule, (2), are colored in cyan. In this case, $\mathcal{T}_{S_4} = 0.4$; therefore, all points with a S_4 higher than 0.4 are marked as scintillation.
2) All points for which scintillation is detected according to the semihard rule, (3), are colored in magenta. In this case, $\mathcal{T}_{\theta_{\text{el}}} = 30°$ and $\mathcal{T}_{C/N_0} = 37$ dBHz; as a consequence, only a subset of points of the previous case are marked as scintillation.

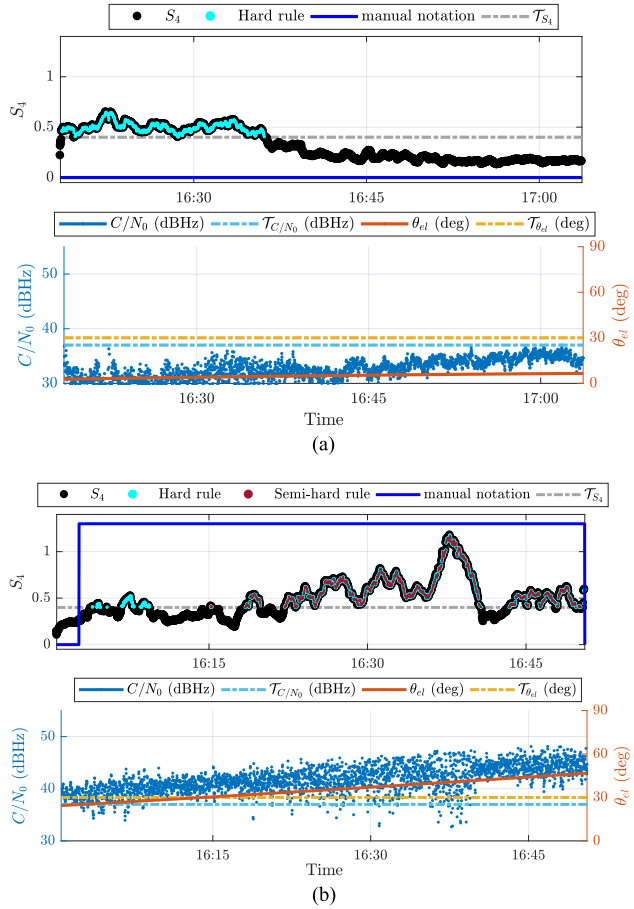

Fig. 3. Comparison of traditional scintillation detection methods for two different scintillation events. Top panels report the trend of the S_4, the manual annotation, the detection results of the hard and semihard rules and the value of the S_4 threshold $\mathcal{T}_{S_4} = 0.4$. Bottom panels report the elevation and C/N_0 trends, and their respective thresholds used in traditional rules, $\mathcal{T}_{\theta_{\text{el}}} = 30°$ and $\mathcal{T}_{C/N_0} = 37$ dBHz. (a) April 2, PRN 23. (b) March 26, PRN 10.

3) All points enclosed by the blue boxes are manually marked as scintillation by visual inspection.

From a careful analysis of the figure, it is clear that the hard and semihard rules fail in identifying scintillation, when compared to the manual annotation, considered as ground truth. In particular, in the first case, reported in Fig. 3(a), the high S_4 values in the time interval between 16:18 and 16:36 are likely due to multipath reflections. This is evident by considering the fact that the satellite is rising (elevation lower than 5°), but also exploiting *a priori* information on environmental conditions, such as the presence of obstacles in the satellite line-of-sight, or historical data analysis, such as the sidereal repetition of the event with the same S_4 pattern. The detection results of the hard rule are then characterized by a high false alarm rate. On the contrary, the semihard rule correctly marks all the points as no scintillation, thanks to the filter added by the threshold on the elevation.

On the other hand, in the second case, depicted in Fig. 3(b), both rules are too conservative. There are time windows, which are part of the same scintillation event, but

TABLE II
Confusion Matrices for the Hard and Semihard Rule, on the Examples of March 26, PRN 10 and April 2, PRN 23

		Hard				Semi-hard	
		0	1			0	1
Manual	0	32.9	19.2	Manual	0	52.1	0
	1	18.9	29.0		1	22.0	25.9

the values of S_4 and θ_{el} slightly lower than the corresponding thresholds classify them as nonscintillated time epochs, thus, generating a high-missed detection rate. The interval from 16:03 to 16.20 is the leading edge phase of the event detected starting from 16:20. Similarly, the interval from 16:40 to 16:44 can be considered scintillation, even if the S_4 value is slightly below the detection threshold; the presence of high S_4 values a few minutes before and after this time interval assures that it can be marked as scintillation.

Results can be formalized by means of the confusion matrix. In the first example, no scintillation is present, so the false positives and false negatives are equal to 0. While the semihard rule gives a True negatives rate of 100%, this percentage reduces to 61.6% for the hard rule. In the second example, the rate of true negatives and false positives is the same for both rules, and corresponds to 4.2% and 0%, respectively. However, in this case, the hard rule has a true positives rate of almost 58%, overcoming the semihard (51.8%), which in turn has a higher false positives rate.

The overall confusion matrices for detection results of the two examples and for both the hard and semihard rule are reported in Table II. The hard rule detection gives an overall percentage of correct estimations of only $32.9\% + 29\% = 61.9\%$. The false alarm rate amounts to 19.2%, while the missed detection rate is 18.1%. When moving to the semihard case, the number of points correctly estimated increases to $52.1\% + 25.9\% = 78\%$. As expected, the false alarms are 0%, but the missed detection rate increases to 22%.

C. Limitations

As demonstrated by the previous examples, the traditional thresholding-based approaches for automatic scintillation detection appear not to be able to fully characterize the event, since decisions based on hard thresholds do not take into account either the physics of the event, or the environmental conditions. Multipath, interference, and other nuisances might lead to erroneous scintillation detection, as S_4 overlooks the higher-moments characteristics of the signals.

Furthermore, it has to be remarked that the computation of S_4 is cumbersome and demanding: it requires complex averaging and detrending operations on the correlation outputs, in order to reduce noise and to remove the slow variations due to the signal dynamics. The choice of the best detrending technique is not trivial: several approaches, based on the use of high-order Butterworth filters, of wavelet transformations, and on simple averaging, have been described [39]–[41]. Nevertheless, it has been proven that different methods lead to different results [42]. It has also been proven that a different detrending shall be chosen for different geographical areas [43] and that detrending operations could introduce postprocessing artifacts [19].

TABLE III
Summary of the Traditional Scintillation Detection Approaches Strengths and Weaknesses

Detection	Strengths	Weaknesses
Hard	Very simple rule. Easy implementation also in real time.	Low detection accuracy Requires detrending. False alarms due to multipath reflections.
Semi-hard	Increased accuracy. Higher robustness to non-ionospheric impairments.	Not general approach. Location dependent. Higher computational burden.
Manual	High accuracy. Cross analysis of historical data.	Costly and time consuming. Dependent on human experience. Subjected to human errors. Non-real-time (post-processing).

Manual annotation can assure higher accuracy in the event classification, in terms of duration and continuity, at the expenses of a time-consuming human-driven visual inspection. Furthermore, it is a postprocessing analysis that is not suitable for a quasi-real-time detection. As a summary, strengths and weaknesses of these three approaches are reported in Table III.

This, in turn, justifies the investigation of different detection and classification techniques. The limitations of the aforementioned approaches can be mitigated exploiting machine learning techniques, able to learn from human processes to produce automatic high-accuracy detection and classification.

IV. MACHINE LEARNING SCINTILLATION DETECTION

The goal of the machine learning detection algorithm is to replicate the performance of the manual detection, without introducing additional human effort to manually classify the data. The algorithm, once trained on big datasets labeled by manual annotation, demonstrates better detection properties with respect to the hard and semihard approaches.

The machine learning algorithm considered is the *decision tree*, as it offers the best compromise between computational complexity, performance, and data preprocessing operations. k-fold cross validation, with $k = 10$, is performed on the dataset: for each run, 90% of input data are used in the training phase and 10% of data are used for the test set.

A. Correlation Matrix Analysis and Features Selection

The first necessary step is the selection of the features used to train the model. Features shall be selected between the measurements provided by a GNSS receiver. Nevertheless, their choice is not trivial, and the final performance of

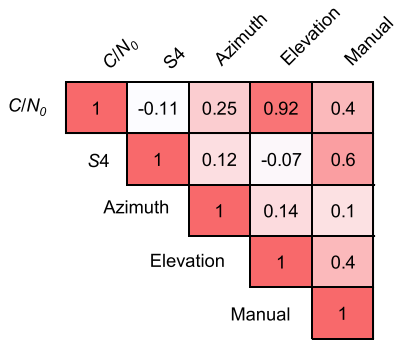

Fig. 4. Correlation matrix, considering the observables of the signal.

the algorithm, as well as the scalability and generality of the technique, depend on the features chosen.

The correlation matrix is a statistical tool used to underline the correlation between each couple of features. Each cell of the correlation matrix reports the Pearson correlation coefficient between variables X and Y, defined as follows:

$$\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

where x_i is a single record of the dataset and $\bar{x}$ is the sample mean. The correlation coefficient ranges from -1 to 1. Correlation between X and Y is absent when the correlation coefficient $\rho(X,Y)$ is equal to 0, weak when $|\rho(X,Y)| \leq 0.35$, moderate in the range 0.36–0.67 and strong when it exceeds 0.68 [44].

B. Observable-Based Features

Fig. 4 reports the matrix of the correlation $\rho(X,Y)$ between the manual ground truth and the 1 Hz observables provided by the GNSS scintillation monitoring receiver: S_4, C/N_0, θ_{el}, and ϑ_{az}. Manual annotation has a moderate correlation with C/N_0, θ_{el}, and S_4, and a very low correlation with ϑ_{az}.

A first set of features, $\mathcal{L}_1$, that looks to be a reasonable starting choice, is then defined as follows:

$$\mathcal{L}_1 = \{S_4,\ C/N_0,\ \theta_{el}\}. \tag{5}$$

It includes the observables having the highest correlation with the manual annotation. Furthermore, they are the same parameters used in the semihard detection rule (3).

The same two case studies considered in Section III-B are analyzed. The detection results of machine learning algorithm based on set $\mathcal{L}_1$ are reported in Fig. 5, along with the detection results of the hard and semihard rules. The blue line identifies the manual human-driven detection. The green line corresponds to the detection results of the machine learning decision tree algorithm. With respect to the hard and semihard cases, and compared to the manual detection, machine learning results show both a lower rate of false alarms, in the example of Fig. 5(a), and a lower rate of missed detection, in the example of Fig. 5(b).

The confusion matrix, reported in Table IV, summarizes the results. When compared to the case reported in Fig. 3, the missed detection rate is reduced to 7.4%, while there are no false alarms. The global success rate corresponds to $52.1\% + 40.5\% = 92.6\%$, meaning that for more than nine cases over ten the machine learning prediction matches the manual annotation. It is important to underline that these two examples, in particular the first, reflect a *worst case* situation, in which the presence of multipath introduces a further level of complexity. More simple cases, relative to portions of the dataset affected by scintillation only, show a higher success rate, and are not reported here.

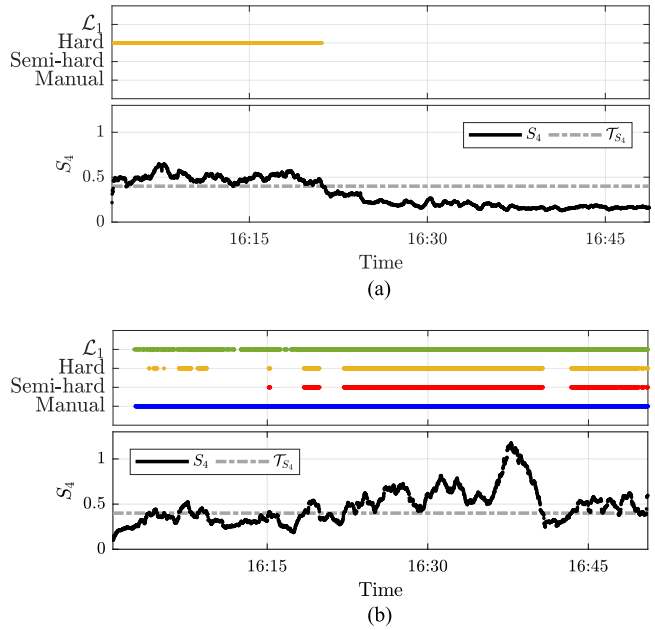

Fig. 5. Decision tree detection results for two different scintillation events and for set $\mathcal{L}_1$. Top panels report the manual annotation (ground truth), the detection results of the hard and semihard rules, and the machine learning detection results. Bottom panels report the trend of the S_4 and the value of the S_4 threshold $\mathcal{T}_{S_4} = 0.4$. (a) April 2, PRN 23. (b) March 26, PRN 10.

TABLE IV
Confusion Matrix for Machine Learning Prediction and for the Set $\mathcal{L}_1$, on the Examples of March 26, PRN 10 and April 2, PRN 23

		Prediction	
		0	1
Manual	0	52.1	0
	1	7.4	40.5

Nevertheless, although the results in terms of scintillation detection are good, the approach based on the signal observables reveals some limitations. On one side, the use of S_4 should be avoided, as it already corresponds to the output of the traditional approach. In addition, its derivation is computationally demanding and involves complex operations that could introduce postprocessing artifacts and location dependent solutions. At the same time, the use of C/N_0 could lead to misleading detection result, as, in the presence of scintillation, it might be affected by a bias or

even provide completely wrong results [18]. Finally, the use of satellites elevation can potentially lead to overfitting, making the model suitable only for the specific location of the data collection used for the training.

C. Signal-Based Features

In order to further improve the machine learning scintillation detection performance, it is possible to exploit, as features, not the final observables and the final scintillation index, but rather their components; in other words, to unpack the S_4 formulation [45]. In this section, the raw GNSS signal measurements at the output of the receiver tracking stage in time domain are used as features, the 50 Hz I and Q correlators output. They correspond to the higher rate observables that can be provided by a commercial receiver, and thus to the most accurate representation of the original GNSS signal. It will be proven that a machine learning approach that uses raw observables as features rather than the scintillation indices not only is able to detect scintillations, but offers a higher performance. Furthermore, such an approach overtakes the problem of computing the scintillation indices and is able to exclude the side effects and artifacts introduced by the postprocessing.

I and Q values cannot be directly injected in the learning, but they have to be averaged, in order to reduce the impact of thermal noise and to highlight the scintillation phenomenon. Therefore, new quantities are defined, based on a short observation window, T_{obs}. N samples of I and Q are averaged over the observation period, where $N = T_{\text{obs}} \cdot 50\ Hz$. The averaged correlations samples, denoted $\langle I \rangle$ and $\langle Q \rangle$, are then defined as follows:

$$\langle I \rangle = \frac{1}{N} \sum_{n=1}^{N} I_n \tag{6}$$

$$\langle Q \rangle = \frac{1}{N} \sum_{n=1}^{N} Q_n \tag{7}$$

where I_n and Q_n are, respectively, the I and Q correlator outputs at time n.

In order to combine the information brought by phase and quadrature components, the SI, at time n, can be computed as follows:

$$SI_n = I_n^2 + Q_n^2. \tag{8}$$

Similarly, SI is averaged over the observation period

$$\langle SI \rangle = \frac{1}{N} \sum_{n=1}^{N} SI_n = \frac{1}{N} \sum_{n=1}^{N} \left(I_n^2 + Q_n^2 \right). \tag{9}$$

This suggests to compute other two additional features, i.e., the average over the observation period of the square of the

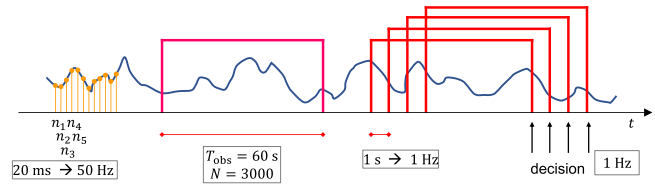

Fig. 6. Illustration of the samples averaging and window overlapping procedure used in the work.

I and Q correlators

$$\langle I^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} I_n^2 \tag{10}$$

$$\langle Q^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} Q_n^2. \tag{11}$$

By analogy with the previous cases and with the algorithm used to compute S_4, a last feature is identified as the average over the observation windows of the square of the SI

$$\langle SI^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} SI_n^2. \tag{12}$$

A new subset of features, including combinations of the I and Q correlators, is introduced

$$\mathcal{L}_2 = \left\{ \langle I \rangle, \langle Q \rangle, \langle SI \rangle, \langle I^2 \rangle, \langle Q^2 \rangle, \langle SI^2 \rangle \right\}. \tag{13}$$

On the other hand, previous works on the topic consider as a feature only the Fourier transform of the SI [19], [20]. Furthermore, it is important to mention that, contrary to what is typically done in literature [18], [46], no prior elevation-based filtering aiming at reducing the effect of multipath is performed. Indeed, applying a mask partially hides the signal distortion phenomenon that are under study.

The features in (13) are obtained by averaging $N = 3000$ values of the 50 Hz correlator stage outputs. The observation windows is set to $T_{\text{obs}} = 60$ s, in agreement with common scintillation observation algorithms [22]. In addition, a sliding and overlapping windowing technique is applied, depicted in Fig. 6: by shifting the observation window of 1 s, the resolution of the observation is increased to 1 Hz. The same average operation over 60 s of values is performed and a set of new features is injected in the machine learning algorithm at the end of each window, thus, at a 1 Hz rate.

The correlation matrix for the set of features $\mathcal{L}_2$ is reported in Fig. 7. From the table, it emerges that a moderate correlation is experienced between the manual scintillation annotation and the variables $\langle SI \rangle$, $\langle I^2 \rangle$ and $\langle SI^2 \rangle$.

The results obtained running decision tree on set $\mathcal{L}_2$ on the selected case studies will be presented in Section V-B, respectively, in Fig. 11(a) and (b).

V. RESULTS

In this section, a complete performance analysis of different machine learning algorithms, on the full dataset, and

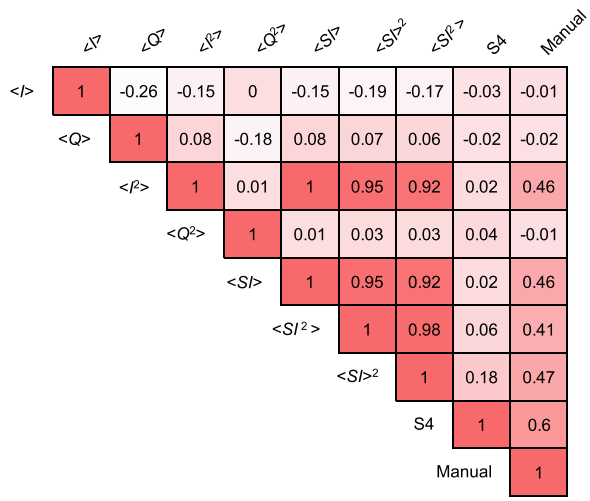

Fig. 7. Correlation matrix considering the features defined in set $\mathcal{L}_2$.

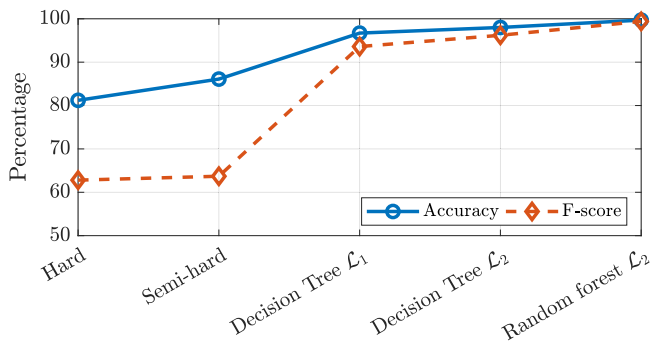

Fig. 8. Summary of the accuracy and F-score obtained for the different scintillation detection techniques presented.

TABLE V
Confusion Matrices for the Hard and Semihard Rule, on the Complete Dataset

		Hard				Semi-hard	
		0	1			0	1
Manual	0	65.45	9.97	Manual	0	73.97	13.61
	1	8.78	15.8		1	0.26	12.16

TABLE VI
Summary of the Overall Detection Results for Different Algorithms Over Different Set of Features

Detection rule	Set	Accuracy	Precision	Recall	F-score
Hard	S_4	81.2%	64.3%	61.3%	62.8%
Semi-hard	S_4, C/N_0, θ_{el}	86.1%	97.9%	47.2%	63.7%
Decision tree	$\mathcal{L}_1$	96.7%	93.4%	93.8%	93.6%
Decision tree	$\mathcal{L}_2$	98.0%	96.3%	96.1%	96.2%
Random forest	$\mathcal{L}_2$	99.7%	99.4%	99.3%	99.4%

TABLE VII
Confusion Matrix for the Complete Dataset Using Decision Tree Over Set $\mathcal{L}_1$

		Prediction	
		0	1
Manual	0	72.2	1.7
	1	1.6	24.5

considering different sets of features, is provided and compared to the traditional detection methods. First, quantitative results in terms of confusion matrices, accuracy, precision, recall and F-score are reported and commented. Then, qualitative results focusing on the punctual analysis of some false negatives and false positives predictions are proposed, with a focus on the run time detection capabilities. To conclude, results of a test on novel, untrained data, collected in a different location, are presented.

Data are part of the same data collection described in Section III-B (Hanoi, Vietnam, at 11° 20′ N geo-magnetic latitude, in March and April 2015). They include a total of 169 955 entries of data at 50 Hz resolution, spanning a total time interval of about 6 h, and including in total 20 different satellites. The rate of scintillation events is about 1:4.

A. Quantitative Results

1) Hard and Semihard Rules: First, summary results related to the hard and semihard rules tests are reported. The confusion matrices are depicted in Table V. The number of false positives and of false negatives is quite high in both cases, respectively, 18.75% and 13.87%. In the semihard rule case, the percentage of false positives is reduced to only 0.26%, at the expenses of a higher rate of false negatives. This is justified recalling that the semihard rule is a conservative approach.

Accuracy, precision, recall, and F-score are reported in the first two rows of Table VI. Although also precision and recall are reported, a fair performance analysis shall be made focusing on the the accuracy, and in particular on the F-score. As detailed in Section II-B3, precision and recall give a partial overview on the goodness of the algorithms, while the F-score is the most complete metric, taking into account also the numerosity of the points denoted as scintillated. The accuracy is about 5% higher in the second case. However, the lower value of the recall is the sign of an uneven evaluation of the performance. Indeed, the F-score for the two approaches is similar; the semihard algorithm only improves this indicator by 1%. The two approaches can then be defined similar in terms of detection performance.

2) Decision Tree on Set $\mathcal{L}_1$: The central row of Table VI reports the results obtained running the decision tree algorithm over the set of features $\mathcal{L}_1$, on the complete dataset, using a standard ten-fold cross validation approach. The confusion matrix is reported in Table VII, and is the result of the average of the ten confusion matrices generated during the ten-fold cross validation process. Despite being a nonstandard approach, this is allowed as each fold has the same proportion of scintillation and nonscintillation points.

The number of false positives and false negatives is reduced to about 3.3%. The accuracy, corresponds to 96.7%,

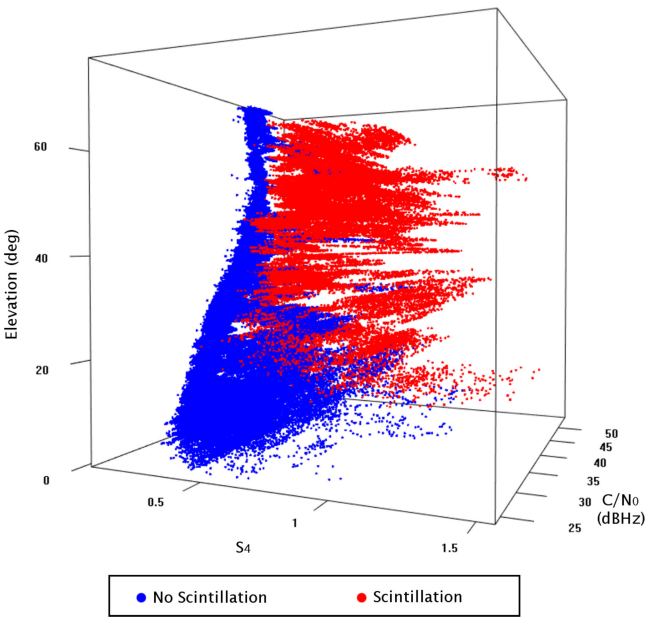

Fig. 9. Representation of the capabilities of the features defined in $\mathcal{L}_1$ to detect scintillation events.

TABLE VIII
Confusion Matrix for the Complete Dataset Using Decision Tree For Set $\mathcal{L}_2$

		Prediction	
		0	1
Manual	0	72.9	1
	1	1	25.1

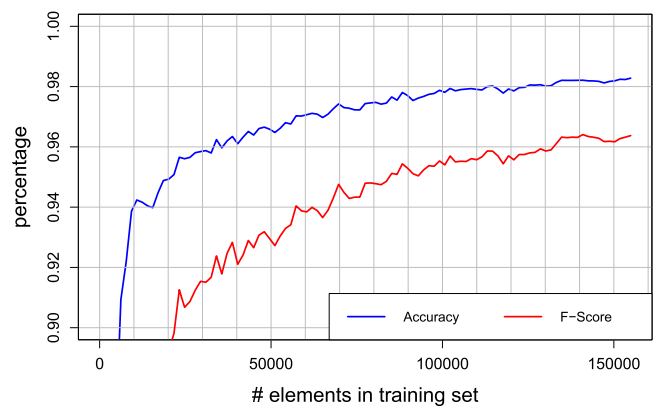

Fig. 10. Accuracy and F-score of the decision tree algorithm versus the number of points used in the training set.

meaning that for 164 299 over 169 955 points the machine learning prediction matches the manual annotation. The improvement with respect to the semihard rule is larger than 10%.

Complementary information is reported in Fig. 9. The figure shows the three dimensional space defined by $\mathcal{L}_1$ and reports the machine learning detection results. Blue points correspond to the portions of the signal for which no scintillation is detected, red points correspond to points for which machine learning algorithm detects scintillation. The two different regions appear to be quite well separated, sign that the global classification performance is good.

3) Decision Tree on Set $\mathcal{L}_2$: The last rows of Table VI report the results of the machine learning approaches on the set of features $\mathcal{L}_2$, identified in (13), and have been obtained performing a ten-fold cross validation.

Machine learning approaches overcome both traditional methods based on hard and semihard thresholding. The improvement, in terms of F-score, is about 30%. In addition, the signal-based set of features $\mathcal{L}_2$ outperforms the observable-based set of features $\mathcal{L}_1$. This is important considering the fact that $\mathcal{L}_2$ only includes features obtained by averaging and summing the I and Q correlator outputs, and that more elaborated features, such as S_4 and C/N_0, are not considered. This overcomes the problem of computing the scintillation indices and in turn reduces the computational burden of the detector. No complex averaging and detrending operations are required to compute S_4 and C/N_0 values, making this approach more generic and flexible. Furthermore, the use of $\mathcal{L}_2$ makes the technique location independent, as elevation is not used.

The confusion matrix obtained averaging the ten folds of the decision tree approach is reported in Table VIII. When compared to the confusion matrices for the set $\mathcal{L}_1$, the number of false positives passes from 1.7% to 1.0%, and the number of false negatives passes from 1.6% to 1.0%.

Fig. 10 reports the accuracy and the F-score obtained running decision tree algorithm on set $\mathcal{L}_2$ and using ten-fold cross validation over a different number of input points. The graph shows that the value of the metrics increases as more points are used in the training phase, because the model defined by the machine learning algorithm is more complete. In order to obtain an accuracy of at least 98% and a F-score of at least 96%, a training dataset of at least 140 000 points shall be used.

Table VI also shows that random forest technique further improves the results, as it evaluates the decisions obtained by several decision trees, at the expenses of a higher computation burden in the training phase.

B. Qualitative Analysis of the False Predictions

Despite being easy to read and offering a quick quantitative overview of the overall performance of machine learning, the metrics reported in Section V-A deserve a deeper analysis. This section focuses on a few test cases, correspondent to different scintillation events, analyzing in detail the most relevant examples of false negatives and false positives. All plots show the detection results of machine learning decision tree tests, performed on set $\mathcal{L}_2$ (green points), compared to the the hard (orange points), semihard (red points), and manual/ground truth (blue points). The value of θ_{el} and of C/N_0, along with the thresholds used in the semihard rule, are reported as a reference in bottom panels, for a better interpretation of the results.

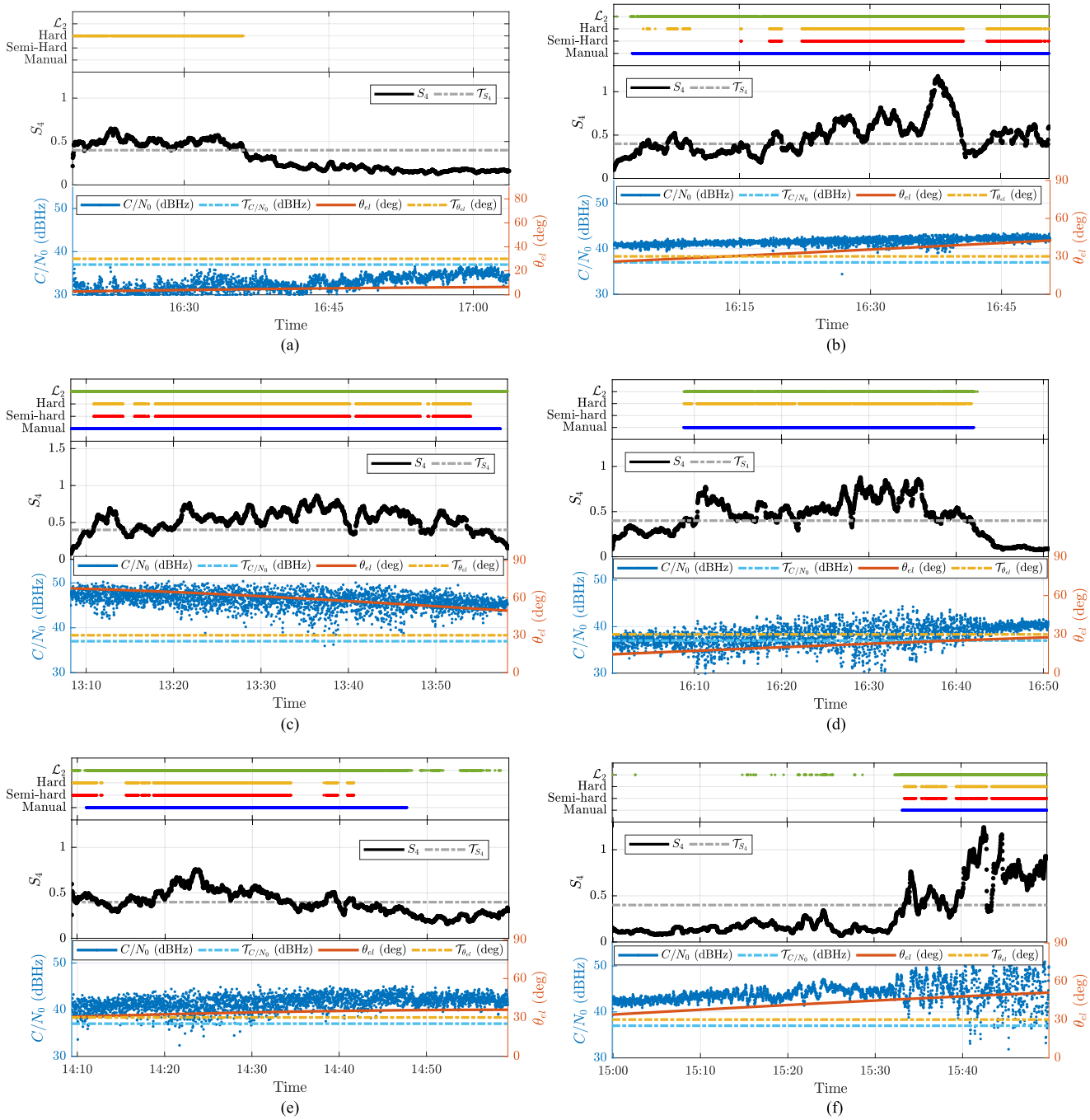

Fig. 11. Decision tree detection results for different test cases. Top panels report the manual annotation, the detection results of the hard and semihard rules, and the machine learning detection results, computed over set $\mathcal{L}_2$. Middle panels report the trend of the S_4 and the value of the S_4 threshold $\mathcal{T}_{S_4} = 0.4$. Bottom panels report the elevation and C/N_0 trends, and their respective thresholds used in traditional rules, $\mathcal{T}_{\theta_{\text{el}}} = 30^\circ$ and $\mathcal{T}_{C/N_0} = 37$ dBHz. (a) April 2, PRN 23. (b) March 26, PRN 10. (c) April 2, PRN 1. (d) March 26, PRN 2. (e) April 2, PRN 3. (f) March 26, PRN 6.

First, the same two examples analyzed in Section III are reported in Fig. 11(a) and (b). The detection performance of decision tree computed over set $\mathcal{L}_2$ is in line with the case over set $\mathcal{L}_1$, reported in Fig. 5. There are no false alarms, and the manual ground truth perfectly matches in both cases, confirming the goodness of the features of set $\mathcal{L}_2$.

Fig. 11(c) proves that the machine learning approach solves some of the the problems due to the predefined thresholds in hard and semihard rules, thus, reducing the missed detection rates and improving the overall accuracy. In this case, many points are wrongly not considered as scintillation by traditional approaches, as the value of S_4 is below $\mathcal{T}_{S_4}$. This happens both in the middle and at the borders of the event, although from visual inspection it is clear that all the points belong to the same 40-min long scintillation event. In particular, the raising and falling edges of the event are very well classified by machine learning

with respect to hard and semihard rules. Also in this case, the decision tree prediction perfectly matches the manual annotation.

Fig. 11(d) is a clear example of the importance of avoiding the *a priori* filtering of data according to an elevation mask. In this case, a scintillation event can be clearly identified, despite the elevation of the satellite is below 30°. The semihard rule, as well as any algorithm, which applies an elevation mask to exclude low elevation satellites, would fails in detecting the event. On the contrary, the decision tree model detects the scintillation and perfectly matches the manual annotation.

In Fig. 11(e), machine learning detection gives a certain number of false positives, when compared to the manual annotation. However, a better analysis of the scintillation event could raise doubts about the correctness of the visual inspection: some of the false positives are concentrated in portions of signal for which the manual annotation could be called into question. Although the S_4 value is below the threshold, the points declared as scintillation by the decision tree algorithm could actually mark the start of a second and lighter event. Similarly, Fig. 11(f) reports a certain number of false alarms. Also, in this case, the points in which scintillation has been detected could actually be precursors of the consecutive long event.

In general, it is possible to state that the cases in which the decision tree approach fails are not serious failures. In some cases, they correspond in ambiguous situations, and can be the consequence of carelessness in the visual inspection procedure, which is one of the limitations of human-based manual annotation. In other situations, missed detections and false alarms are isolated events, and could be completely eliminated, for instance, eliminating scintillation occurrences lasting less then a few seconds. It can be proved that an accurate postprocessing of the machine learning results can lead to even higher accuracy and F-score values, further confirming the validity of the machine learning approach for autonomous scintillation detection.

C. Run Time Events Detection

In addition to the good performance in terms of accuracy and detection rates, the results reported above show that machine learning offers a valid solution also for run time detection of the events. An increasing number of applications requires to rapidly raise an alert in case of scintillation, potentially leading to poor quality of the GNSS positioning solution. As traditional methods mostly rely on S_4, no alerts can be raised during the leading edge phase of a scintillation event. Indeed, when compared to the hard and semihard rules, decision tree approach allows an earlier detection of the event.

This is confirmed by many experimental results; one example is depicted in Fig. 12. In this specific case, machine learning can detect the scintillation event with an advance of 9 min, with respect to hard and semihard rules. Similar considerations could be drawn by considering the detection

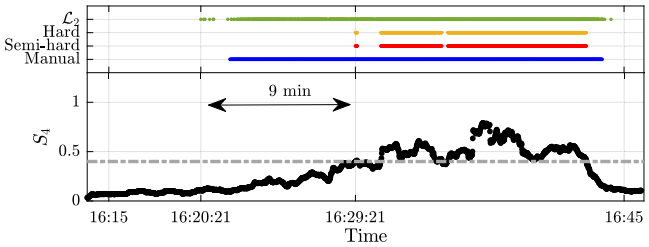

Fig. 12. Example of early scintillation detection, April 2, PRN 17.

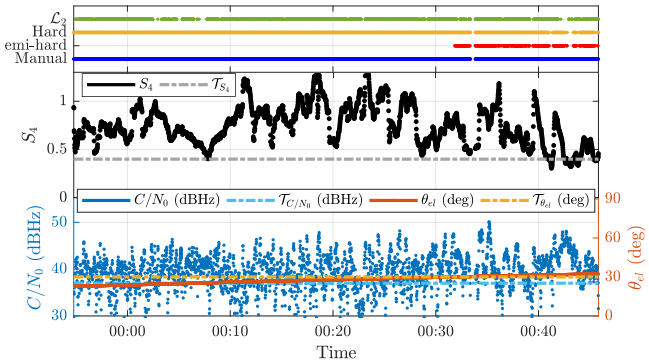

Fig. 13. Example detection for novel and untrained data, March 25, PRN 9.

results reported in Fig. 11(b), (c), and (f). It is important to point out that this cannot be considered a prediction of the scintillation event, but rather an early detection. Scintillation prediction is out of the scope of this paper.

D. Test on Novel Data

This section presents some test results on novel GNSS data affected by scintillation, in order to demonstrate the effectiveness of the decision tree approach. Data were acquired using a similar acquisition system, on March 25, 2015, in a different location (Presidente Prudente, Brasil, at a geomagentic latitude of 12° 90′ S). A fragment of 1 h of data, on seven different GPS satellites, corresponding to 28 066 entries has been considered. As in the previous case, a manual ground truth has been defined by visual inspection. It is important to underline that such data are not involved in the training phase, and thus in the construction of the model. The model applied is the same considered in all previous tests, defined over the signal-based features of set $\mathcal{L}_2$.

The overall F-score is higher than 90%, while for the semihard rule only 80.1% is reached. An example of detection results is reported in Fig. 13. The results for the other PRN affected by scintillation are similar and are not reported. Machine learning decision tree detection results are in line with the manual human-based annotation. All the data between 23:56 and 00:32 are below 30° of elevation. Detection techniques based on data prefiltering would have completely discarded from the analysis the data in this interval, while rules including an elevation threshold would have wrongly marked those events as nonscintillation.

TABLE IX
Features Identified to be Used in the Machine Learning Approach

	Name	Description
Observable-based features	S_4	The typical amplitude scintillation index used in scintillation monitoring. It is equal to zero for non-scintillation amplitude events, and increases during a scintillation event.
	C/N_0	The carrier over noise power density ratio, a low rate averaged measure of the signal strenght, based on the value of the I and Q output correlators.
	θ_{el}	The satellite's elevation angle above the horizon.
	ϑ_{az}	The satellite's azimuth with respect to the North.
Signal-based features	$\langle I \rangle$	The in-phase correlator output averaged over the observation window, defined in (6).
	$\langle Q \rangle$	The quadra-phase correlator output averaged over the observation window, defined in (7).
	$\langle I^2 \rangle$	The in-phase correlator output squared and averaged over the observation window, defined in (10).
	$\langle Q^2 \rangle$	The quadra-phase correlator output squared averaged over the observation window, defined in (11).
	$\langle SI \rangle$	The signal intensity, averaged over the observation window, defined in (9).
	$\langle SI^2 \rangle$	The signal intensity squared and averaged over the observation window, defined in (12).

VI. CONCLUSION

This paper proposes a novel approach to systematically detect ionospheric scintillation events affecting the GNSS signals amplitude, characterized by a higher level of accuracy, reliability, and readiness. This goal is achieved by leveraging machine learning techniques, able to learn from historical preclassified data, and to perform automatic classification on new data. A theoretical background on GNSS, ionospheric scintillations, and machine learning is given in the first part of the paper, along with a description of the classical detection techniques based on the computation of the amplitude scintillation index and on the hard thresholding rules. Several tests have been carried out and are reported in Section V. The results demonstrate that this approach outperforms state-of-the art techniques in terms of accuracy and F-score, and that it can reach the levels of a manual human-driven annotation. Furthermore, it has been proven that signal-based features (set $\mathcal{L}_2$) outperform observable-based features (set $\mathcal{L}_1$).

In addition, this solution offers the following added values:

1) increased generalization and location-independence, by avoiding the use of scintillation indices;
2) independence from detrending and average operations, required to compute S_4;
3) independence from *a priori* filtering of data at a certain elevation mark, thus, preserving potential useful information;
4) capability to avoid false alarms due to multipath reflections;
5) improved computational speed and costs optimization in terms of human effort and time;
6) potentiality to raise early run-time scintillation alerts.

Machine learning can facilitate the work of analyzing big sets of GNSS data affected by amplitude scintillation, leading to a better understanding of the physical phenomenon, and to a potential improvement of the robustness of GNSS receivers.

APPENDIX SUMMARY OF ALL THE FEATURES

All the features considered in this paper are summarized in Table IX.

ACKNOWLEDGEMENTS

The authors would like to thank James T. Curran for providing an initial dataset of GNSS data affected by scintillated events and for supporting the research work with enthusiasm and with precious suggestions.

REFERENCES

[1] P. Kintner, B. Ledvina, and E. De Paula
GPS and ionospheric scintillations
Sp. Weather, vol. 5, no. 9, pp. 1–23, 2007.

[2] V. Sreeja, M. Aquino, K. de Jong, and H. Visser
Effect of the 24 September 2011 solar radio burst on precise point positioning service
Sp. Weather, vol. 12, no. 3, pp. 143–147, 2014.

[3] B. Moreno, S. Radicella, M. C. de Lacy, M. Herraiz, and G. Rodriguez-Caderot
On the effects of the ionospheric disturbances on precise point positioning at equatorial latitudes
GPS Solutions, vol. 15, no. 4, pp. 381–390, 2011.

[4] X. Zhang, F. Guo, and P. Zhou
Improved precise point positioning in the presence of ionospheric scintillation
GPS Solutions, vol. 18, no. 1, pp. 51–60, 2014.

[5] J. Lee, Y. T. J. Morton, J. Lee, H.-S. 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 Process. Mag., vol. 34, no. 5, pp. 96–110, Sep. 2017.

[6] C. Cesaroni *et al.*
Monitoring ionosphere over South America: The MImOSA and MImOSA2 projects
in *Proc. Int. Assoc. Inst. Navig. World Congr.*, Oct. 2015, pp. 1–7.

[7] N. Linty *et al.*
Ionospheric scintillation threats to GNSS in polar regions: the DemoGRAPE case study in Antarctica
in *Proc. Eur. Navig. Conf.*, 2016, pp. 1–7.

[8] L. Spogli *et al.*
Formation of ionospheric irregularities over Southeast Asia during the 2015 St. Patrick's Day storm
J. Geophys. Res.: Sp. Phys., vol. 121, no. 12, pp. 211–233, 2016.

[9] J. Vila-Valls, P. Closas, C. Fernandez-Prades, and J. T. Curran
On the ionospheric scintillation mitigation in advanced GNSS receivers
IEEE Trans. Aerosp. Electron. Syst., to be published.

[10] G. Lachapelle and A. Broumandan
Benefits of GNSS IF data recording
in *Proc. Eur. Navig. Conf.*, May 2016, pp. 1–6.

[11] C. Cristodaro, F. Dovis, N. Linty, and R. Romero
Design of a configurable monitoring station for scintillations by means of a GNSS software radio receiver
IEEE Geosci. Remote Sens. Lett., vol. 15, no. 3, pp. 325–329, Mar. 2018.

[12] S. Taylor, Y. Morton, Y. Jiao, J. Triplett, and W. Pelgrum
An improved ionosphere scintillation event detection and automatic trigger for GNSS data collection systems
in *Proc. Int. Tech. Meet. Inst. Navig.*, 2012, pp. 1563–1569.

[13] W. Fu, S. Han, C. Rizos, M. Knight, and A. Finn
Real-time ionospheric scintillation monitoring
in *Proc. 12th Int. Tech. Meet. Satellite Div. Inst. Navig.*, 1999, vol. 99, pp. 14–17.

[14] S. Miriyala, P. R. Koppireddi, and S. R. Chanamallu
Robust detection of ionospheric scintillations using MF-DFA technique
Earth, Planets Sp., vol. 67, no. 98, pp. 1–5, 2015.

[15] R. Romero, N. Linty, F. Dovis, and R. V. Field
A novel approach to ionospheric scintillation detection based on an open loop architecture
in *Proc. 8th ESA Workshop Satellite Navig. Technol. Eur. Workshop GNSS Signals Signal Process.*, Dec. 2016, pp. 1–9.

[16] L. Rezende *et al.*
Survey and prediction of the ionospheric scintillation using data mining techniques
Sp. Weather, vol. 8, no. 6, pp. 1–10, 2010.

[17] Y. Jiao, J. J. Hall, and Y. T. Morton
Performance evaluations of an equatorial GPS amplitude scintillation detector using a machine learning algorithm
in *Proc. 29th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Sep. 2016, pp. 195–199.

[18] Y. Jiao, J. J. Hall, and Y. T. Morton
Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm
IEEE Trans. Aerosp. Electron. Syst., vol. 53, no. 1, pp. 405–418, Feb. 2017.

[19] Y. Jiao, J. J. Hall, and Y. T. Morton
Automatic GPS phase scintillation detector using a machine learning algorithm
in *Proc. Int. Tech. Meet. Inst. Navig.*, Monterey, CA, USA, Jan. 2017, pp. 1160–1172.

[20] 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, vol. 64, no. 3, pp. 391–402, 2017.

[21] P. W. Ward, J. W. Betz, and C. J. Hegarty
Satellite signal acquisition, tracking, and data demodulation
in *Understanding GPS: Principles and Applications*, E. Kaplan and C. Hegarty Eds. Norwood, MA, USA: Artech House, pp. 153–241, 2006.

[22] A. Van Dierendonck, J. Klobuchar, and Q. Hua
Ionospheric scintillation monitoring using commercial single frequency C/A code receivers
in *Proc. 6th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Salt Lake City, UT, USA, Sep. 1993, vol. 93, pp. 1333–1342.

[23] P. Flach
Machine Learning: The Art and Science of Algorithms That Make Sense of Data. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[24] A. Favenza, N. Linty, and F. Dovis
Exploiting standardized metadata for GNSS SDR remote processing: a case study
in *Proc. 29th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Portland, Oregon, Sep. 2016, pp. 77–85.

[25] T. M. Mitchell
Machine Learning. New York, NY, USA: McGraw Hill, 1997.

[26] J. R. Quinlan
Induction of decision trees
Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[27] R. Tibshirani and J. Friedman
The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Berlin, Germany: Springer, 2001.

[28] N. Japkowicz and M. Shah
Evaluating Learning Algorithms: A Classification Perspective. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[29] L. Breiman
Random forests
Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[30] P. Refaeilzadeh, L. Tang, and H. Liu
Cross-validation
in *Encyclopedia of Database Systems*, L. Liu and M.-T. Özsu Eds. Berlin, Germany: Springer, 2009, pp. 532–538.

[31] S. Dubey, R. Wahi, E. Mingkhwan, and A. Gwal
Study of amplitude and phase scintillation at GPS frequency
Indian J. Radio Sp. Phys., vol. 34, pp. 402–407, Dec. 2005.

[32] E. Aon, A. R. Othman, Y. H. Ho, and R. Shaddad
A study of ionospheric GPS scintillation during solar maximum at UTeM station
Jurnal Teknologi, vol. 73, pp. 123–128, 2015.

[33] R. M. Romero Gaviria
Estimation techniques and mitigation tools for ionospheric effects on GNSS receivers
Ph.D. dissertation, Politecnico di Torino, Torino, Italy, 2015.

[34] R. W. Middlestead
Digital Communications With Emphasis on Data Modems: Theory, Analysis, Design, Simulation, Testing, and Applications. Hoboken, NJ, USA: Wiley, 2017.

[35] A. Adewale, E. Oyeyemi, A. Adeloye, C. N. Mitchell, J. A. Rose, and P. Cilliers
A study of L-band scintillations and total electron content at an equatorial station, Lagos, Nigeria
Radio Sci., vol. 47, no. 2, pp. 1–12, 2012.

[36] Y. Jiao, Y. T. Morton, S. Taylor, and W. Pelgrum
Characterization of high-latitude ionospheric scintillation of GPS signals
Radio Sci., vol. 48, no. 6, pp. 698–708, 2013.

[37] P. Abadi, S. Saito, and W. Srigutomo
Low-latitude scintillation occurrences around the equatorial anomaly crest over Indonesia
Annales Geophysicae, vol. 32, no. 1, pp. 7–17, 2014.

[38] J. T. Curran, M. Bavaro, A. Morrison, and J. Fortuny
Developing a multi-frequency for GNSS-based scintillation monitoring receiver
in *Proc. 27th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Tampa, FL, USA, Sep. 2014, pp. 1142–1152.

[39] M. Najmafshar, S. Skone, and F. Ghafoori
GNSS data processing investigations for characterizing ionospheric scintillation
in *Proc. 27th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Tampa, FL, USA, Sep. 2014, pp. 1190–1202.

[40] M. H. Mokhtar
Mitigation of scintillation effects on GPS satellite positioning
Ph.D. dissertation, Univ. Leeds, Leeds, U.K., 2012.

[41] S. C. Mushini, P. Jayachandran, R. Langley, J. MacDougall, and D. Pokhotelov
Improved amplitude-and phase-scintillation indices derived from wavelet detrended high-latitude GPS data
GPS Solutions, vol. 16, no. 3, pp. 363–373, 2012.

[42] F. Niu
Performances of GPS signal observables detrending methods for ionosphere scintillation studies
Ph.D. dissertation, Miami Univ., Nashville, TN, USA, 2012.

[43] B. Forte and S. M. Radicella
Problems in data treatment for ionospheric scintillation measurements
Radio Sci., vol. 37, no. 6, pp. 8-1–8-5, 2002.

[44] R. Taylor
Interpretation of the correlation coefficient: a basic review
J. Diagnostic Med. Sonography, vol. 6, no. 1, pp. 35–39, 1990.

[45] A. Favenza, A. Farasin, N. Linty, and F. Dovis
A machine learning approach to GNSS scintillation detection: Automatic soft inspection of the events
in *Proc. 30th Int. Tech. Meet. Satellite Div. Inst. Navig.*, Sep. 2017, pp. 4103–4111.

[46] T. Y. Atilaw, P. Cilliers, and P. Martinez
Azimuth-dependent elevation threshold (ADET) masks to reduce multipath errors in ionospheric studies using GNSS
Adv. Sp. Res., vol. 59, no. 11, pp. 2726–2739, 2017.

Nicola Linty received the M.Sc. degree in communication engineering from Politecnico di Torino, Torino, Italy, in 2010 and the Ph.D. degree in electronics and telecommunication engineering from Politecnico di Torino, in 2015.

He is currently an Assistant Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy, and a Member of the Navigation Signal Analysis and Simulation research group. His research interests include the field of signal processing and simulation, applied to satellite navigation, design and development of innovative algorithms for GPS and Galileo receivers.

Alessandro Farasin received the Master's degree in computer science (Computer Vision & AI) from the Università degli Studi di Torino, Turin, Italy, in 2016. He is working toward the Ph.D. degree in machine learning algorithms, exploited and realized for the analysis of geospatial data with the Department of Automation and Informatics, the Politecnico of Torino, Torino, Italy.

He is currently a Researcher and a Software Developer with Istituto Superiore Mario Boella, Turin, Italy. He is working with the Microsoft Innovation Center area, ISMB, architecting and developing solutions based on Microsoft technologies, mainly focused on cloud, big data, and machine learning related topics.

Alfredo Favenza received the M.Sc. degree in computer science from the Università degli Studi di Torino, Torino, Italy, in 2007.

He is currently a Senior Researcher and a Project Manager with Istituto Superiore Mario Boella, Turin, Italy. In 2007, he started his research activity with ISMB in the area of Navigation technologies developing strong expertise on GNSS Software Defined Radio, and Augmentation Systems (EGNOS/EDAS). He is currently a Researcher of Mobile Solutions Research Area with ISMB interested in cross technologies domains where location-based services, big data, and navigation technologies can be combined to generate added-value innovation.

Fabio Dovis received the M.Sc. degree in electronics engineering and the Ph.D. degree in electronics and communications engineering from Politecnico di Torino, Torino, Italy.

He is currently an Associate Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy, and a Member of the Navigation Signal Analysis and Simulation group. He has a relevant experience in European projects in satellite navigation as well as cooperation with industries and research centers. His research interests include the design of GPS and Galileo receivers and advanced signal processing for interference and multipath detection and mitigation.