Near-real-time detection of co-seismic ionospheric disturbances using machine learning

Quentin Brissaud[1,1,1,1] and Elvira Astafyeva[2,2,2,2]

[1]NORSAR
[2]IPGP

September 7, 2023

Abstract

Tsunamis generated by large earthquake-induced displacements of the ocean floor can lead to tragic consequences for coastal communities. Ionospheric measurements of Co-Seismic Disturbances (CIDs) offer a unique solution to characterize an earthquake's tsunami potential in Near-Real-Time (NRT) since CIDs can be detected within 15 min of a seismic event. However, the detection of CIDs relies on human experts, which currently prevents the deployment of ionospheric methods in NRT. To address this critical lack of automatic procedure, we designed a machine-learning based framework to (1) classify ionospheric waveforms into CIDs and noise, (2) pick CID arrival times, and (3) associate arrivals across a satellite network in NRT. Machine-learning models (random forests) trained over an extensive ionospheric waveform dataset show excellent classification and arrival-time picking performances compared to existing detection procedures, which paves the way for the NRT imaging of surface displacements from the ionosphere.

submitted to *Geophys. J. Int.*

Near-real-time detection of co-seismic ionospheric disturbances using machine learning

Quentin Brissaud[1*] and Elvira Astafyeva[2]

[1] *NORSAR, Kjeller, Norway*

[2] *Université de Paris, Institut de Physique du Globe de Paris (IPGP),*

CNRS UMR7154, 35-39 Rue Hélène Brion, 75013 Paris, France

30 December 2021

SUMMARY

Tsunamis generated by large earthquake-induced displacements of the ocean floor can lead to tragic consequences for coastal communities. Ionospheric measurements of Co-Seismic Disturbances (CIDs) offer a unique solution to characterize an earthquake's tsunami potential in Near-Real-Time (NRT) since CIDs can be detected within 15 min of a seismic event. However, the detection of CIDs relies on human experts, which currently prevents the deployment of ionospheric methods in NRT. To address this critical lack of automatic procedure, we designed a machine-learning based framework to (1) classify ionospheric waveforms into CIDs and noise, (2) pick CID arrival times, and (3) associate arrivals across a satellite network in NRT. Machine-learning models (random forests) trained over an extensive ionospheric waveform dataset show excellent classification and arrival-time picking performances compared to existing detection procedures, which paves the way for the NRT imaging of surface displacements from the ionosphere.

Key words: Ionosphere/atmosphere interactions – Tsunami warning – Machine Learning

1 INTRODUCTION

Large seafloor displacements due to earthquakes are known to generate destructive tsunamis. Unfortunately, Near-Real-Time (NRT) mapping of the co-seismic surface displacements to characterize the earthquake tsunami potential

* Correspondence: quentin@norsar.no

is still challenging for conventional methods, especially for earthquakes with $M_w > 8$ (LaBrecque et al. 2019; Wright et al. 2012; Katsumata et al. 2013). In our definition, NRT corresponds to times within 15-20 minutes after the earthquake onset which is crucial for early-warning application as it gives several tens of minutes for populations to evacuate before the tsunami reaches the coasts.

Recently, several research groups have demonstrated that ionospheric measurements can offer an alternative to seismo-geodetic methods to estimate the tsunami potential of earthquakes. The ionosphere is an electrically charged atmospheric layer that is concentrated around 150-400 km of altitude. This layer is sensitive to the vertically propagating acoustic energy excited by natural hazards (earthquakes, tsunamis, volcanic eruptions) and man-made events (explosions, rocket launches, nuclear tests) (Heki 2006; Rolland et al. 2016; Komjathy et al. 2016; Shults et al. 2016; Astafyeva & Shults 2019; Astafyeva 2019). In particular, ionospheric signatures of earthquakes, known as co-seismic ionospheric disturbances (CID), can be detected 7-9 minutes after the earthquake. CIDs waveform characteristics are correlated to the seismic source properties. For instance, the amplitude of the CID scales almost linearly with the magnitude of an earthquake (Astafyeva et al. 2013b, 2014; Cahyadi & Heki 2015; Occhipinti et al. 2018; Heki 2021), or - for submarine earthquakes - with the tsunami wave height or volume of water that was displaced due to an earthquake (Kamogawa et al. 2016; Rakoto et al. 2018; Manta et al. 2020). Additionally, CID arrival times and detection coordinates provide strong constraints on the position of the seismic source, or the origin of tsunami (Afraimovich et al. 2006; Heki et al. 2006; Astafyeva et al. 2009; Tsai et al. 2011; Lee et al. 2018; Bagiya et al. 2020; Inchin et al. 2021; Zedek et al. 2021). Moreover, Astafyeva et al. (2011, 2013a); Astafyeva (2019) showed that the distribution of the first-detected CIDs match the position of the maximum displacement on the ground, and (Kakinami et al. 2021) showed that the initial point of CID matches the maximum vertical displacement of the tsunami source.

However, despite the high potential of seismo-ionospheric assessment of natural hazards, the detection and analysis of ionospheric disturbances still rely on human experts. This manual process is problematic for processing large data volume to detect CIDs and estimate seismic source parameters. Only a few studies have focused on the automatization of detection procedures in the ionosphere but only at low frequencies (Efendi & Arikan 2017; Belehaki et al. 2020). Ravanelli et al. (2021) investigated the use of both GNSS ground and ionospheric TEC measurements for NRT tsunami genesis estimation. However, Ravanelli et al. (2021) did not present any detection procedure for CIDs, but only showed TEC variations in NRT scenario. In addition, their TEC processing procedure included the use of 8th order polynomial fit in order to highlight the co-seismic signature. The latter is not possible in our definition of NRT mode, i.e. 15-20 minutes after the earthquake onset time. The first NRT-compatible method detecting CID was suggested by Maletckii & Astafyeva (2021). However, this study only showed good results on 1 Hz data with CIDs showing high temporal TEC derivative. Therefore, the community needs methods allowing for rapid automatic detection and recognition of CIDs for both future NRT developments and processing of large amount of TEC data retrospectively.

The problem of earthquake waveform detection has been investigated in the seismic community since the early days of modern computers (e.g., Allen 1982)). The automatization of waveform detection procedures has historically been performed in the seismic community using analytical methods such as the Short-Time Average / Long-Time Average (STA/LTA) filter(Allen 1982)). However, the high rate of false positives generated by these analytical filters has motivated the seismic community to implement Machine-Leaning (ML) approaches that combine both low computational time and high accuracy (Ross et al. 2018; Mousavi et al. 2020). Even when only small labelled waveform datasets are available, ML methods provide excellent classification results (Provost et al. 2017; Wenner et al. 2021). In particular, Random Forests (RF, Breiman 2001) show excellent generalization abilities, and do not require an extensive hyper-parameter tuning. Random-forest is an ensemble technique that build predictions by aggregating predictions from a set of decision trees. Aggregating results from individual decision trees built using bootstrap aggregation, that consist of randomly selecting input features to train each tree, makes RF particularly robust to new data.

To address the lack of automatic detection method, we build a RF-based architecture to classify TEC timeseries, pick arrival times, and associate detected arrivals. Random-forests are trained over an extensive CID waveform dataset from 12 large-magnitude earthquakes, to classify vTEC waveforms between CIDs and noise and pick arrival times in NRT. Our method is, to the best of our knowledge, the first reported machine-learning classifier and arrival-time picker of CIDs. In this paper, we first describe the generation of our waveform dataset, our detection procedure, and our machine-learning models. We show classification performance results over our testing dataset and against other analytical detection methods. We finally discuss the future implementation of such method for NRT applications.

2 DATA COLLECTION

The Global Navigation Satellite Systems (GNSS) are widely used to sound the ionosphere. GNSS signals transmitted by satellites and captured by ground-based dual-frequency GNSS receivers enable the estimation of the differential slant TEC (sTEC), that is equal to the number of electrons along a line-of-sight (LOS) between a satellite and a receiver. The sTEC is calculated from phase and code measurements (Hofmann-Wellenhof et al. 2008; Afraimovich et al. 2006; Shults et al. 2016). The phase measurements provide precise information about the ionospheric variations and disturbances, but they are biased by an unknown phase ambiguity constant. The code measurements are noisy and less precise, but are not ambiguous, which enables to estimate the bias by averaging the code values along the arc of measurements. The sTEC is then estimated by removing the bias from the phase measurements. However, in near-real-time scenario, since the CID and other disturbances are clearly seen in phase measurements, we suggest to calculate the sTEC using solely phase measurements that can be rapidly retrieved in real-time via the Networked Transport of RTCM via Internet Protocol (NTRIP):

$$sTEC_{ph} = \frac{1}{A} * \frac{f_1^2 * f_2^2}{f_1^2 - f_2^2} * (L_1 * \lambda_1 - L_2 * \lambda_2) \tag{1}$$

where A = 40.308 m^3/s^2, L_1 and L_2 are phase measurements, λ_1 and λ_2 are wavelengths at the two Global Positioning System (GPS) frequencies: $f_1 = 1227,60$ and $f_2 = 1575,42$ MHz. Once the sTEC is calculated, the first data value is subtracted from all data series to remove the unknown bias. Finally, because the sTEC is affected by the elevation angle of the LOS, we convert sTEC to vertical TEC (vTEC) by using the standard "mapping function":

$$vTEC = sTEC * \cos\left(\arcsin(\frac{R_e \cos\theta}{R_e + H_{\text{ion}}})\right), \tag{2}$$

where R_e is the Earth radius, θ is the LOS elevation angle, H_{ion} is the altitude of ionospheric detection. The H_{ion} cannot be known because the sTEC is an integral parameter. Based on the physical principles, the H_{ion} is presumed to be around the ionization maximum, i.e. around 250-350 km. Here we take H_{ion}=250 km for all events. This choice is reasonable from the point of view of the ionospheric physics, while determining of the real altitude of CID detection is out of the scope of this work. Moreover, once the system is trained, it can detect CID in TEC data series for any H_{ion} value. The total electron content is measured in TEC units (TECU), with 1 TECU$= 10^{16}$ electrons/m^2.

To construct our database, we collected GNSS-TEC data with CID signatures for 12 earthquakes that occurred between 2003 and 2016 (see Figure 1 and Table A1), including the M6.6 Chuetsu earthquake which is the smallest earthquake ever recorded by ionospheric GNSS data (Cahyadi & Heki 2015). The typical CID waveform are N-shaped and hump signatures (Figure 1b). However, CID waveforms also depend both on the magnetic field configuration in the epicentral region and on the geometry of the GNSS-sounding (Heki & Ping 2005; Astafyeva & Heki 2009; Rolland et al. 2013; Bagiya et al. 2019). Therefore, in order to correctly represent the large diversity of CID waveforms in our model, we included a variety of different TEC signatures that could be recorded after an earthquake (examples shown in Figures 1b to 1e).

The GNSS data used in this study were of 1, 15 and 30 second cadences (see Table A1). Following the NRT-compatible scenario, we did not apply band-pass filter to extract or amplify CID signatures, but only worked with raw relative vTEC.

3 AUTOMATIC DETECTION AND ASSOCIATION MODELS

We propose a multi-step RF-based procedure to detect and associate CIDs (see Figure 2): 1) selection of a time window, 2) data preprocessing, 3) waveform features extraction, 4) RF-based classification of inputs features between noise and CID classes, 5) if detection probability $> 50\%$ at step 4, RF-based arrival time picking, 6) if 3 successive time windows classified as CID, confirmation of the presence of an arrival and aggregation of arrival times, and 7) if a detection is confirmed at step 6, we then associate this arrival to previously detected CIDs. Finally, we shift the time window and repeat the procedure.

3.1 Preprocessing and feature extraction

To extract consistent waveform features in TEC data with different sampling times, we first downsample all waveforms down to 30 s (see Supplementary Section S6). Consistency in sampling rate is critical as the higher-frequency spectral content can lead to substantial variations in input features. For example, energy peaks at higher frequencies, that would

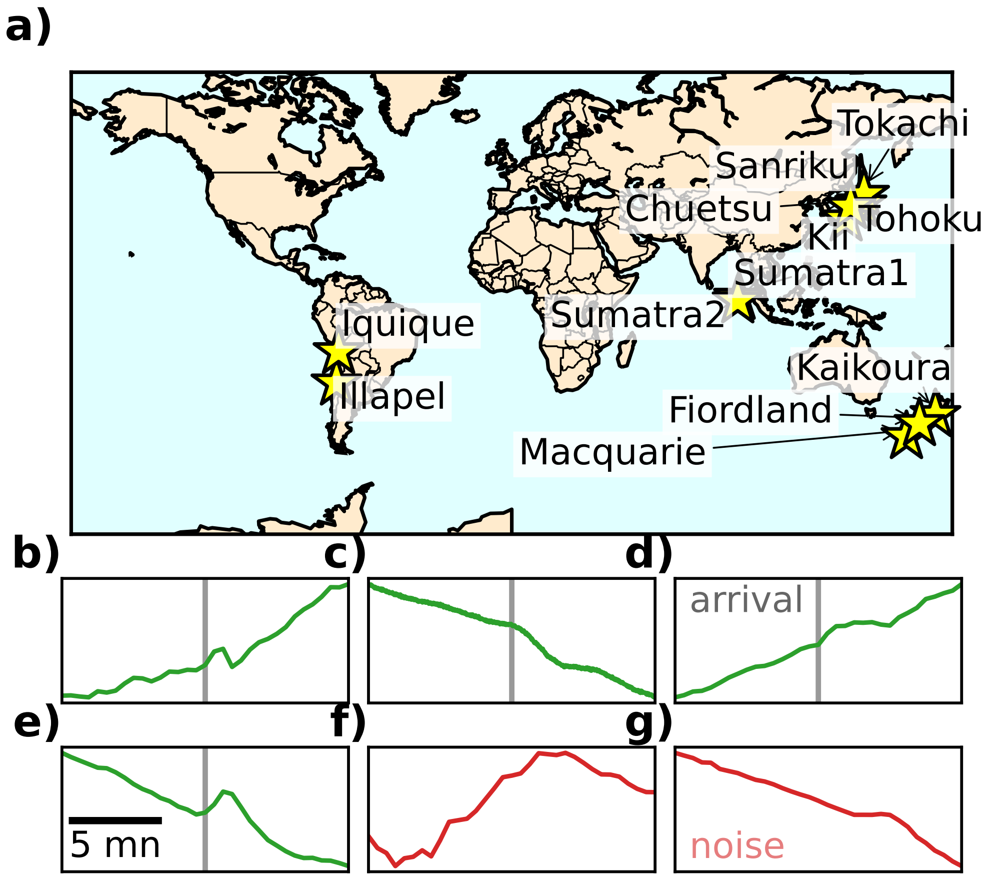

Figure 1. CID waveform dataset. (a) map showing the event included in the training dataset. Details about each event can be found in Table A1. (b) to (g) vTEC waveforms against time that include a CID arrival (panels b to e, green) and that only contain noise (panels f and g, red). The CID arrival time is shown as a grey vertical line in panels (b) to (e).

normally be smoothed out at lower frequencies, can drastically alter the envelope kurtosis and skewness. Additionally, TEC data may contain long-term trends due to GNSS satellite motion and other long-period TEC changes which can be considered as noise for the problem of CID detection. Therefore, we remove long-term trends (signals with periods typically greater than 30 mn) by first taking the time derivative of vTEC waveforms to remove long-wavelength trends and then performing a linear de-trending. Derivatives are computed using second order central differences in the interior points and second order one-sides (forward or backwards) differences at the boundaries. Once the TEC waveforms have been pre-processed, we extract 46 features calculated from the vTEC timeseries, spectra, and spectrograms (see Supplementary Section S1). These features are commonly used for signal classification tasks (e.g., Hammer et al. 2013; Hibert et al. 2014; Provost et al. 2017; Wenner et al. 2021).

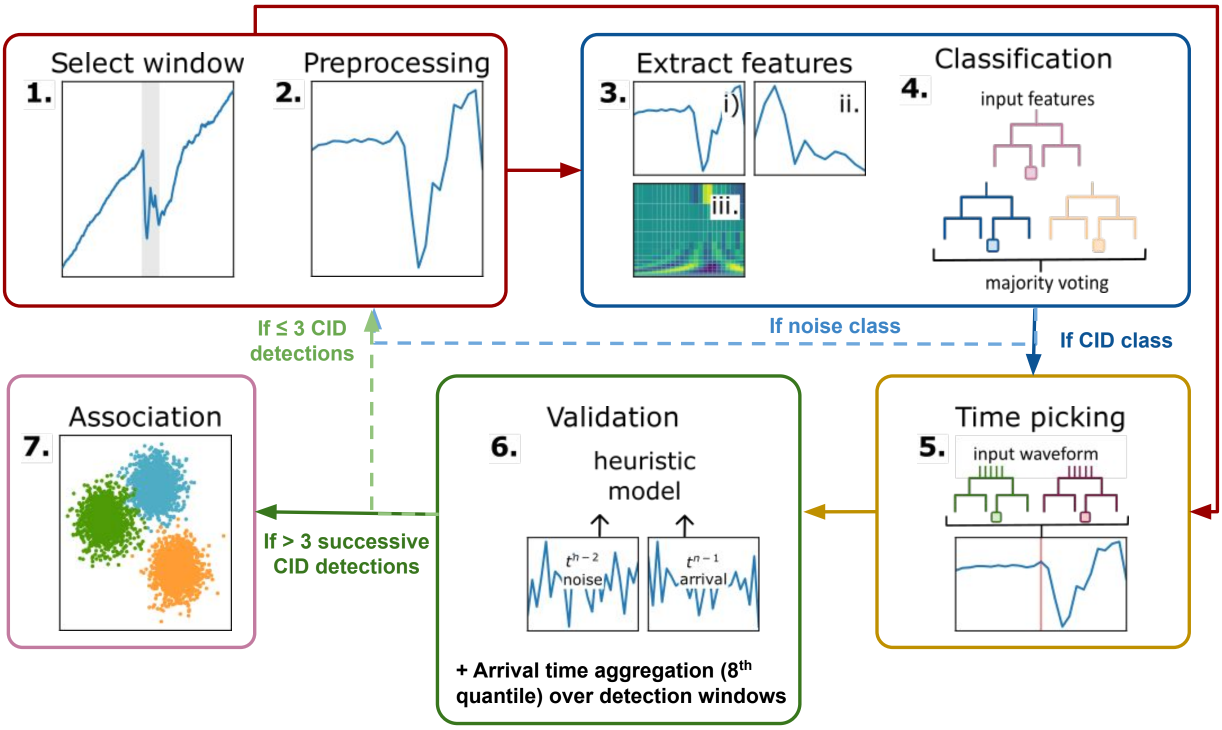

Figure 2. Detection and association procedures described in Section 3: 1) selection of a time window, 2) preprocessing of the waveform, 3) extraction of waveform features from i) time series, ii) spectrum, and iii) spectrogram, 4) RF classification of input waveform, 5) RF arrival time picking, 6) confirmation of an arrival if RF has classified three consecutive time windows (at times t^{n-2}, t^{n-1}, t^n) as arrival, and 7) association of arrivals across different satellites and stations.

3.2 Building a single-station CID detector

We selected a RF model (Breiman 2001) to discriminate vTEC signals between earthquakes and noise classes. Our RF model takes the features extracted from a given waveform at the previous step as inputs and outputs the probability of this waveform to be signal or noise. An input waveform is classified as CID if the detection probability predicted by the RF is over 50%. RFs predictions are constructed from average predictions from an ensemble of individual decision trees. Individual decision trees are built through bootstrap aggregation that consist of randomly selecting input features to train each tree. RFs have excellent generalization abilities, and do not require an extensive hyper-parameter tuning. We used the ”ExtraTrees” scikit implementation of the random forest (Pedregosa et al. 2011) which introduces an additional layer of randomness when building decision trees which allow for better generalization of the training dataset (Geurts et al. 2006). The training procedure relies on bootstrap samples to build each tree along with out-of-bag samples to estimate the generalization score. Bootstrapping makes decision trees less sensitive to the choice of training dataset which reduces the probability of overfitting. Additionally, the error computed from out-of-bag samples provides an excellent metric for RF’s classification performances.

We need to first build a dataset of features to train our RF classifier. This dataset building process is summarized in Figure 3. For each station, CID wavetrains are described by an arrival time and a duration. Wavetrain durations are considered uniform across satellites and stations for a given event (see Table A1). Signal durations are used to automatically label waveforms as CIDs, i.e., to build our training dataset. We consider a time-window to contain a CID if it overlaps the true wavetrain, i.e., CID confirmed by human analyst, by at least 70% which makes the RF more flexible to detect partial CID waveforms. Values picked for the duration correspond to estimates of the minimum duration of the CID across the network of satellites and stations. This choice ensures that at least the arrival time and/or the time at vTEC maximum are contained in the waveforms. Similar to Ross et al. (2018), we augment our training dataset by selecting four time-windows over each CID arrival by randomly shifting the beginning of the time window while still fulfilling the 70% overlap condition. Noise waveforms are selected randomly across all dataset with the condition that it should not overlap any CID wavetrain. Before extracting features, we add artificial Gaussian noise to the waveforms in the training dataset to reduce overfitting similar to Mousavi et al. (2020). We add Gaussian noise to both arrival and noise waveforms s so that the perturbed waveform $\overline{s}$ shows a specific Signal-to-Noise Ratio (SNR) such that $\overline{s} = s + \sqrt{\frac{\sigma^2}{\text{SNR}}}n$, where s is the original waveform, σ^2 is the variance of the original waveform, n is the added noise sampled from a normal distribution, and the SNR is picked within the range SNR $\in (1, 5)$. The final dataset consists of 2867 CIDs and 2867 randomly-picked noise waveforms.

3.3 Building an arrival-time picker

After the classification step, our detection algorithm needs to accurately select the arrival time in each window with a detection probability $> 50\%$. This time picking procedure remains challenging using threshold-based conditions such as STA/LTA filters (Allen 1982). False positives will degrade the arrival time estimate when using threshold-based methods since signal-to-noise ratio, signal duration and dispersion characteristics vary significantly between events. To overcome this problem, we build an automatic arrival-time picking procedure by using an ”ExtraTrees” RF regressor. Our RF takes a normalized pre-processed waveform as input (see Figure 2) and outputs offset in seconds from the window central time, i.e, a float number between -360 and 360. We trigger this arrival time picker only over windows where an arrival has been confirmed.

Similar to the RF classifier, we first have to build a waveform dataset to train our RF arrival-time picker (see Figure 3). We select arrival window for waveforms that overlaps the true wavetrain by at least 30%. This overlap is significantly lower than for the detector. This choice aims at training the RF to pick arrival times over the first detection window with incomplete CID waveforms. Similar to the training of the RF classifier in Section 3.2, we augment our training dataset by selecting four time-windows over each CID arrival by randomly perturbing the beginning of the time window while still fulfilling the 30% overlap condition which captures the uncertainty in arrival-time picking. The final dataset to train the arrival-time picker consists of 2867 CIDs.

3.4 Confirming a detection on a single station

Because of the natural variability of the ionosphere, false detections can still be present after the RF classification step. These false detections generally correspond to short-time spikes in RF detection probabilities while true detections show an increase in RF detection probabilities over longer time periods. To further remove false positives, we confirm a detection if 3 consecutive time windows are classified as CIDs. Variations of this value between 2 and 5 have a relatively small ($< 1\%$) influence on both recall and precision (see Supplementary Section S3). Short-time decrease in detection probabilities can occur within long CID wavetrains (generally caused by large earthquakes) compared to the processing time window. To reduce the number of false negatives, we notify the end of an CID wavetrain if 4 consecutive time windows show a detection probability below 50%.

Once a detection is confirmed, we must determine a single arrival time for the whole wavetrain. However, predictions in successive windows classified as CIDs and belonging to the same wavetrain might not have the same predicted time. Therefore, we determine the detected wavetrain's arrival time by computing the 8^{th} decile of the predicted arrival times over up to 10 successive CID windows. This choice of decile removes the influence of outliers in predicted arrival times made in early detection windows. We do not include predicted arrival times beyond 10 time steps, i.e. 300 s, since these arrivals might correspond to time windows that do not include the true arrival time.

3.5 Associating confirmed detections

Once arrival times are picked across a network of stations and satellites, their spatial distribution informs us about the nature of the detected disturbance. Because large-scale disturbances (e.g., geomagnetic storms, internal gravity waves) or false positives can still pollute the detection dataset after the confirmation procedure at step 5, it is critical to discriminate between CIDs and other sources. If the detected signals belong to a CID, arrival times should follow the geometry of the CID wavefront, whose geometry is controlled by local sound velocities (Inchin et al. 2021). Therefore, the difference in CID arrival times between two stations/satellites can not be lower than the time it takes an acoustic wavefront to propagate between these two stations/satellites at the local acoustic velocity. Furthermore, the spatial extent of the CID wavefront in the ionosphere is constrained by the dimensions of the activated faults at the ground (Inchin et al. 2021) which is generally below 1000 km. Arrivals detected at two stations/satellites located at large distances from each other (i.e., > 1000 km) are not likely to belong to the same CID wavefront. By ignoring combinations of detections that show un-realistic travel times, we further improve the quality of our detection dataset.

The association procedure is performed on a set of confirmed arrivals and consists of three steps: 1) for new detections $d_{current}$, give $d_{current}$ an unused association number $s_{current}$, 2) For each detection $d_{current}$ find other confirmed detections d_{accept} across the satellite network within an acceptable time range from the current detection $d_{current}$. By acceptable time range, we consider all arrivals with a time offset from the current detection $t_{offset} < r_{max}/c_{min}$, where $r_{max} = 500$ km is the maximum association range between two detection points, and $c_{min} = 0.65$ km/s is the minimum horizontal acoustic velocity. r_{max} is chosen as the maximum possible radius of a CID wavefront,

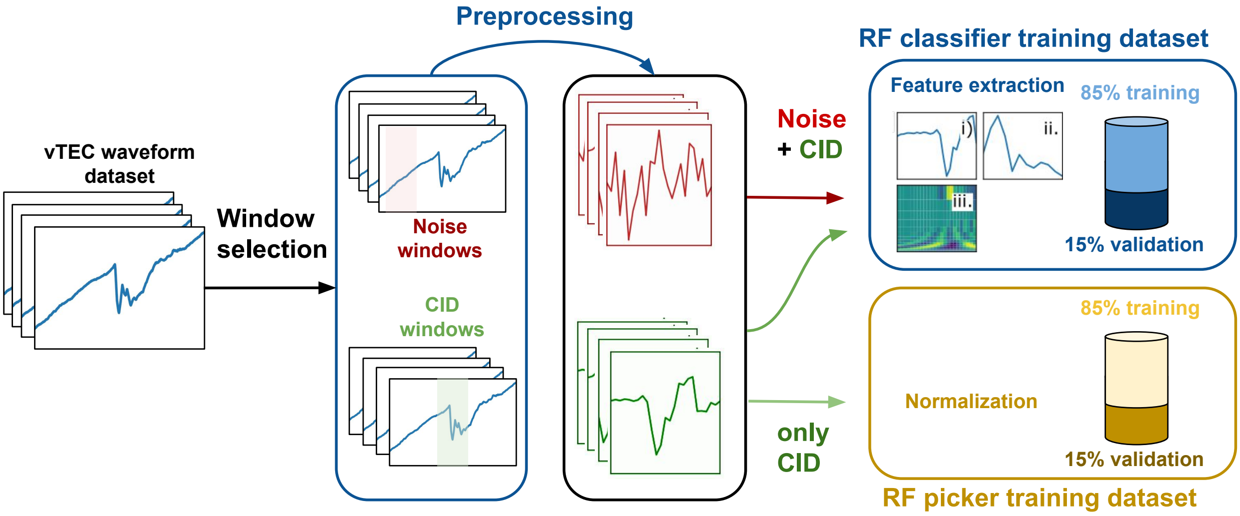

Figure 3. Building datasets to train our CID classifier and arrival-time picker. Each waveform in our vTEC dataset contains information about the CID arrival time and wavetrain duration. First, 4 CID windows and 4 noise windows are extracted from each vTEC waveform. CID windows must overlap the CID wavetrain by at least 70% while noise windows must start or end at least 1000 s, respectively, after or before the CID wavetrain. Each window is then pre-processed (derivative and linear detrending) to remove long-term trend. Features are extracted from the preprocessed CID and noise waveforms to build a training dataset for our RF classifier with 85% assigned to the training dataset and 15% to the validation dataset. To build our arrival-time picker RF model, preproccessed CID waveforms are normalized with 85% assigned to the training dataset and 15% to the validation dataset.

and c_{min} corresponds to the minimum acoustic velocity in the lower ionosphere. Finally, 3) for each detection in an acceptable time range d_{accept}, if detection has an association number s_{accept}, change $s_{current}$ to s_{accept}.

4 RESULTS

To optimize our ML models for detection and arrival-time picking, we split both datasets between 85% training data and 15% validation data (see Figure 3). The classifier's validation dataset is to calculate confusion matrices and measure the rate of false and true positives which is not accessible when bootstrapping samples. The performance of the classification procedure is sensitive to the window size used for training. In Figure 4a, we show both recall and precision metrics for both classes vs the choice of window size. Precision indicates the proportion of true detections relative to all detections (true positives plus false positives). Recall corresponds to the ratio of correct detections over all detections that should have been made (true positives plus false negatives). Because performances are also affected by the choice of overlap threshold used to build the training dataset, recall and precision are averaged over four overlaps between 30% to 90%. We observe that there is a clear improvement in both noise precision and arrival recall (up to $\sim 94\%$) with an increase in window size over the testing dataset up to 720 s. This owes to the higher number of incomplete CID wavetrain for smaller windows than larger ones. For larger time windows > 720 s, precision and recall values plateau as the predictive power of some input features computed over large time windows diminishes. We selected a time window of 720 s which gives excellent classification results while facilitating the arrival time picking procedure by decreasing the range of possible values compared to larger time windows.

The RF model can provide an estimate of the relative feature importance through the calculation of the Gini's impurity during training. The three best features (see Figure 4b) include two timeseries features (ratio of the envelope mean over the envelope maximum and the kurtosis of the timeseries) as well as a spectral feature (energy up to the Nyquist frequency, i.e., 0.0165 Hz), which differs from other signal classification studies (e.g., Wenner et al. 2021). However, the calculation of feature importance can be biased when considering continuous or high-cardinality categorical variables or when inputs features are co-linear. Co-linearity is present in our input dataset between spectral and time-series features (see Supplementary Section S2) which indicates a potential bias in variable importance results. The significant overlap between distributions supports the choice of a large number of features to properly discriminate between each class. Note that this overlap between clusters is also present when using other clustering methods such as such as Principal Component Analysis and t-distributed Stochastic Neighbor Embedding (see Supplementary Section S2), which further highlights the complexity of this classification problem.

The recall for our detection model, shown in Figure 4c, is high for a wide range of probability thresholds indicating that the RF rarely labels true arrivals as noise. We observe in Figure 4d that this value decreases rapidly for probability thresholds $> 50\%$ corresponding to a stricter classification. A threshold at 50% is a good trade-off to balance true and false positive rates. True and false positives show strong similarities in terms of amplitude and frequency content (see Figure 4e). However, with larger thresholds, the fall-out, i.e., the number of false alerts will also decrease. Changes in number of false alerts with variations in probability thresholds highlights that the threshold can be adapted to specific applications depending on the objective. For early warning applications, the number of missed alert should be low and lower thresholds could therefore be used. In contrast, when building arrival-time catalog to invert for source parameters, precision is key and false alerts should be avoided, which necessitates larger thresholds. Additionally, results indicate that RF outperforms the other analytical methods, including STA/LTA filters, in terms of both true and false positive rates (see Appendix B).

Detection results for a waveform recorded during the 2011 Sanriku earthquake (Figure 5a) show that both predicted (vertical grey line) and true (vertical red line in top panel) arrival times overlap, as the absolute error is low ($<$ 3 s). Note that the time used to plot detection probabilities corresponds to the end of the time window used for each classification. We observe that the duration of this wavetrain (~ 450 s) is much larger than the true wavetrain (~ 200 s), owing to the large time windows employed in our detection model. Outside of the detected wavetrain, detection probabilities generally remain low ($< 20\%$) in accordance to the high true negative rate shown in Figure 4c.

In addition to the classification of individual waveform snippets, accurate arrival times are crucial for NRT applications. We assess our model's arrival-time picking accuracy by computing the error between predicted and true arrival times. Arrival-time errors for each event in our CID dataset in Figure 5b indicate that most arrivals ($\sim 95\%$)

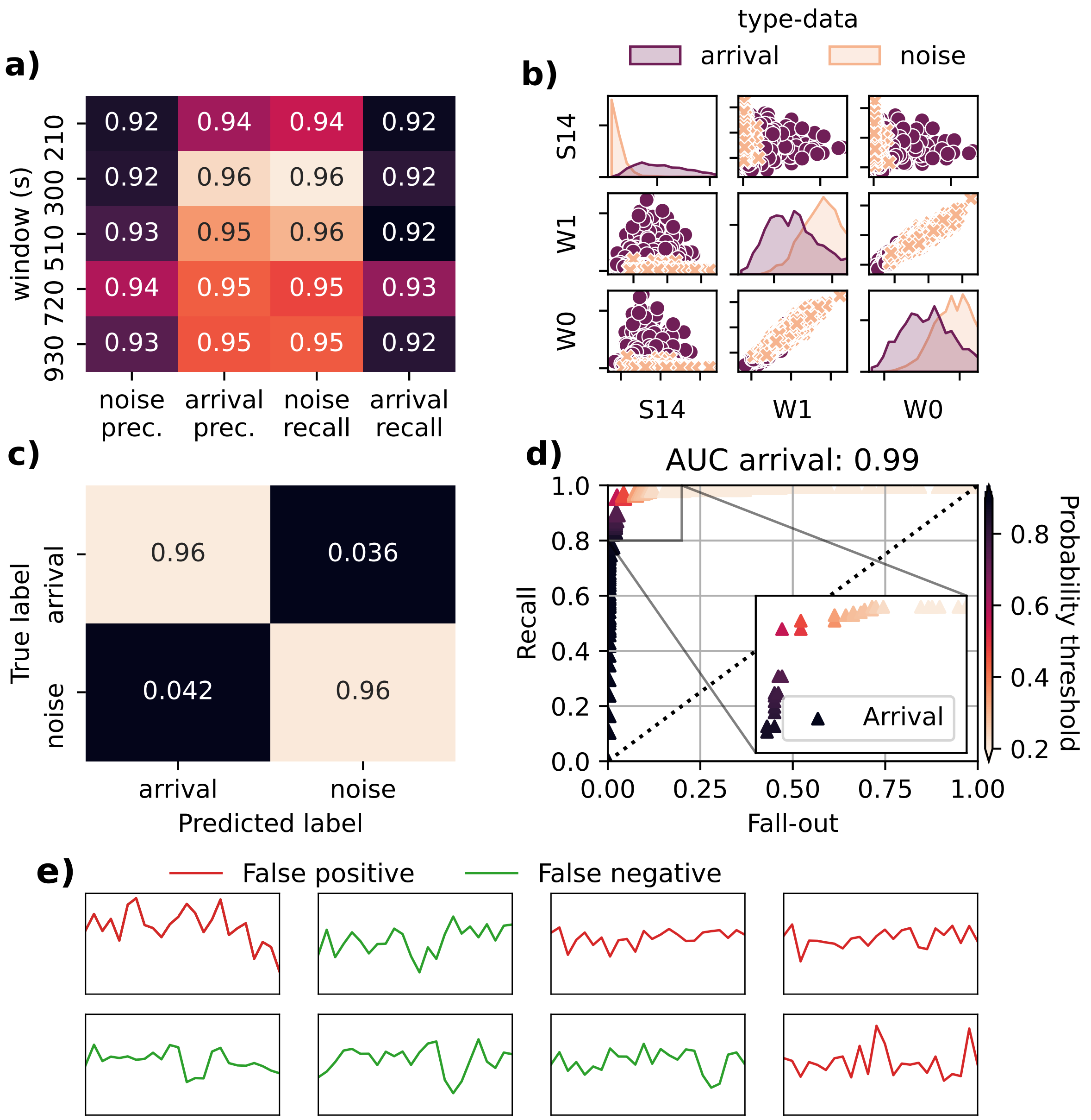

Figure 4. Sensitivity and accuracy of the RF classification step. (a) Precision (prec.) and recall for noise and arrival classes and various window sizes averaged over multiple overlap thresholds: $30\%, 50\%, 70\%$, and 90%. The following formula are used to compute recall and precision for arrival and noise: recall arrival = $TP/(TP + FN)$, recall noise = $TN/(TN + FP)$, precision arrival = $TP/(TP + FP)$, and precision noise = $TN/(TN + FN)$. TP, TN, FP, and FN correspond to True positive, True Negative, False positive, and False Negative. The correct detection of a CID corresponds to a TP. (b) Distribution of the three best features against each other. In the diagonal, we show univariate histograms for each feature. Best features are determined during training by calculating the Gini's impurity. W0 corresponds to the ratio of the envelope mean over the envelope maximum, W2 is the kurtosis of the timeseries, and S14 is the energy up to the Nyquist frequency, i.e., 0.0165 Hz. (c) Confusion matrix for the detection model with window size $w = 720$ s and an overlap of 70%. The confusion matrix is normalized over each row. (d) Arrival-class ROC curve using the detection model with window size $w = 720$ s. The Area Under Curve (AUC) value is shown above the panel. (e) examples of pre-processed waveforms corresponding to FP (red) and FN (green).

are captured with an absolute error $< 60s$, i.e., less than two time steps, and a large proportion of arrivals ($\sim 80\%$) are accurately reproduced with an absolute error < 30 s, which is below the sampling time in each CID waveform. Some outliers are present for both Illapel and Kaikoura events. Errors for the Kaikoura earthquake owe primarily to the high noise level in the waveforms (i.e., random fluctuations of TEC background) which leads to large variations in vTEC time derivatives. For Illapel, false positives are lumped together with the true detection windows and degrade the arrival-time picking performance over 4 time steps. However, the average arrival-time picking error across the whole dataset decreases significantly as the number of time steps increases, i.e., time since first detection (see Figure 5c).

Confirmed detections across multiple satellites/stations can be used to plot ionospheric maps for each event. Comparing Tohoku's ionospheric images in Figures 5d and 5e, we observe that the spatial distribution of arrival times is accurately reproduced by our detection model. The earliest arrival times match the location of maximum slip at the surface. The slight shift of the first arrivals to the south east owes to our choice of altitude of detection H_{ion} (Astafyeva et al. 2013a). However, some spurious arrivals are present in Figure 5e, west of the fault with early arrival times, and south-east of the fault with late arrival times. These false detections correspond to rapid changes in vTEC occurring more than 20 mn before or after the true arrival and classified as earthquake signals by our model.

Our association procedure enables the discrimination between detections belonging to the same wavefront and spurious arrivals. The distribution of association classes for the confirmed detections is shown in Figure 5f. Owing to the large time difference between spurious arrivals and the true arrivals, false detections are correctly classified in different association classes (see first vertical dark purple line in the inset plot in Figure 5f). The time evolution of the distribution of confirmed arrivals (see Supplementary Section S5) indicates that the entirety of the true arrivals were detected within 15 min after the event. Note that the position of ionospheric detection points is dependent on the altitude of detection H_{ion}, which could impact the association classes. However, while changing H_{ion} from 180 to 250 km for Tohoku affects the location of the ionospheric points, true CID arrivals are still correctly associated within the same class (see Supplementary Section S7).

New detections have also been reported by our model west of the epicenter (Figures 5d and 5e), in addition to the ones picked by human analysts, for the largest class corresponding the true CID (inset plot in Figure 5e and light purple class in Figure 5f). A low signal-to-noise ratio pulse is visible after the predicted arrival time (vertical line) at $t = 9.9$ mn after the earthquake, which is consistent with acoustic travel time from the source highlighted by other studies (e.g., Astafyeva et al. 2013a). Using our model also ensures consistency in the choice of arrival times, in contrast to human analysts who introduce a subjective uncertainty range when determining the true onset.

In order to further assess the ability of our model to detect arrivals on new unseen data, we processed waveforms recorded after the 2014 Iquique earthquake (see Table A1). In Figure 6a, we show the slip distribution of the Iquique earthquake along with the RF predicted arrivals times and association classes in Figures 6b and 6c. Predicted arrival times are coherent with the region of maximum slip at the surface despite a few false detections south of the fault. This confirms the excellent detection, arrival-time picking, and classification results on new data.

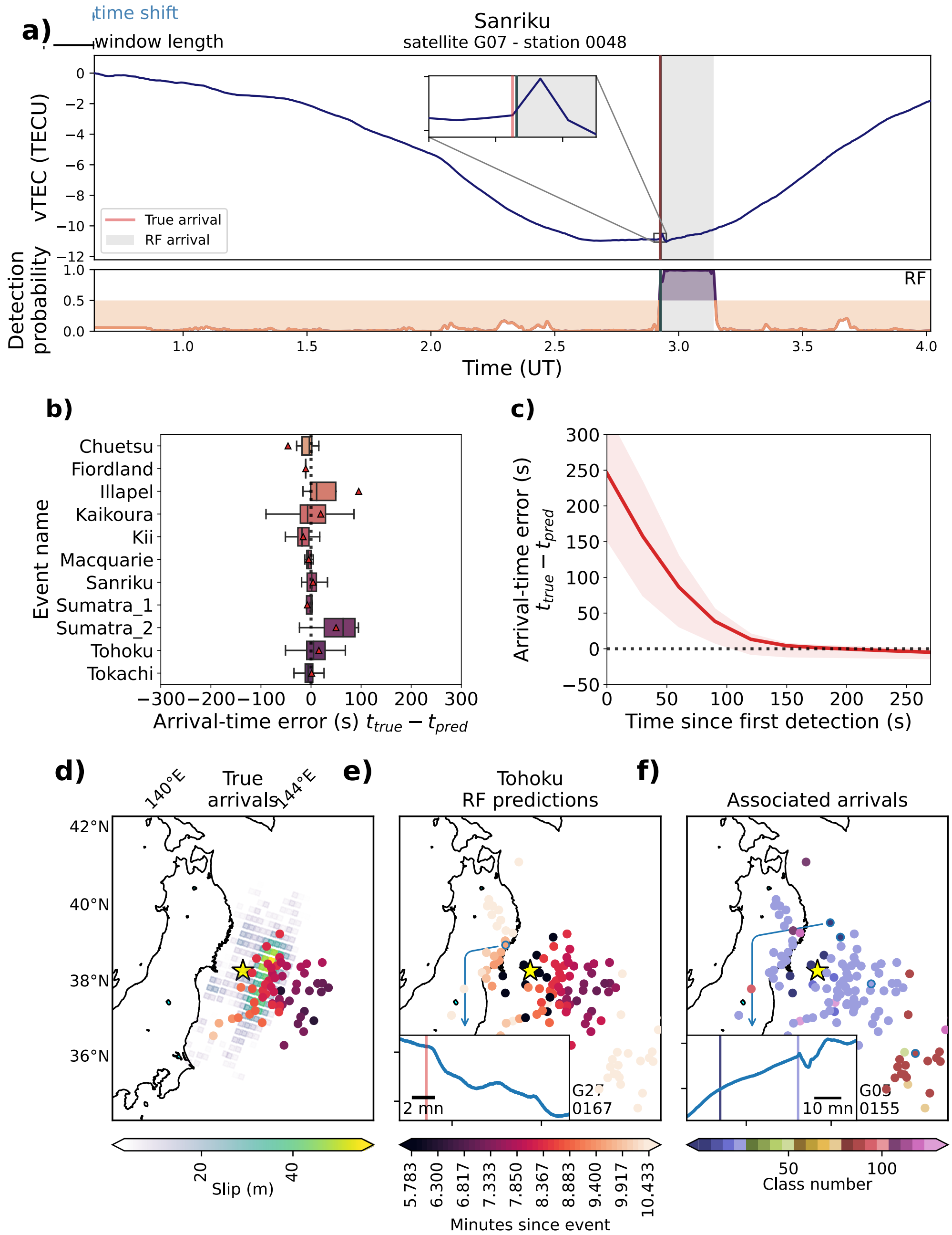

Figure 5. Performance assessment of arrival-time picking and association steps. (a) 4-h vTEC waveform for the Sanriku event, satellite G07, station 0048 along with RF detection probabilities. The time used to plot probabilities over each window is the window end time. The true arrival is shown as a red vertical line and the RF-predicted arrival time as a dark grey vertical line. The wavetrain detected by the RF and heuristic models is highlighted with a grey background. (b) box plot of arrival-time picking errors (in s) vs event after 3 mn since the first detection window. (c) Evolution of arrival-time picking error vs time delay since first detected window. The red curve shows the average error across all events. Red shaded background shows the 1^{st} to 3^{rd} quartile region computed across the events. (d) to (f) Tohoku's ionospheric maps with (d) hand-picked arrival times along with the epicenter location (yellow star), and surface projection of the fault slip (in m) as green to yellow patches, (e) RF-based arrival-time predictions for each confirmed detection with an inset plot showing a vTEC waveform for satellite G27 and station 0167 which was not reported by human analyst, and (f) association classes determined from confirmed detecytions, along with an inset plot showing the vTEC data for satellite G26, station 0155. The vertical lines correspond to the arrival times of the two detections at the station (first is a false detection; second is a true arrival). CID coordinates were calculated at the intersection point between the LOS and the ionospheric layer using $H_{\text{ion}} = 200$ km for lower elevations, and 250 km for higher elevations. These maps are generated 15 minutes after the event.

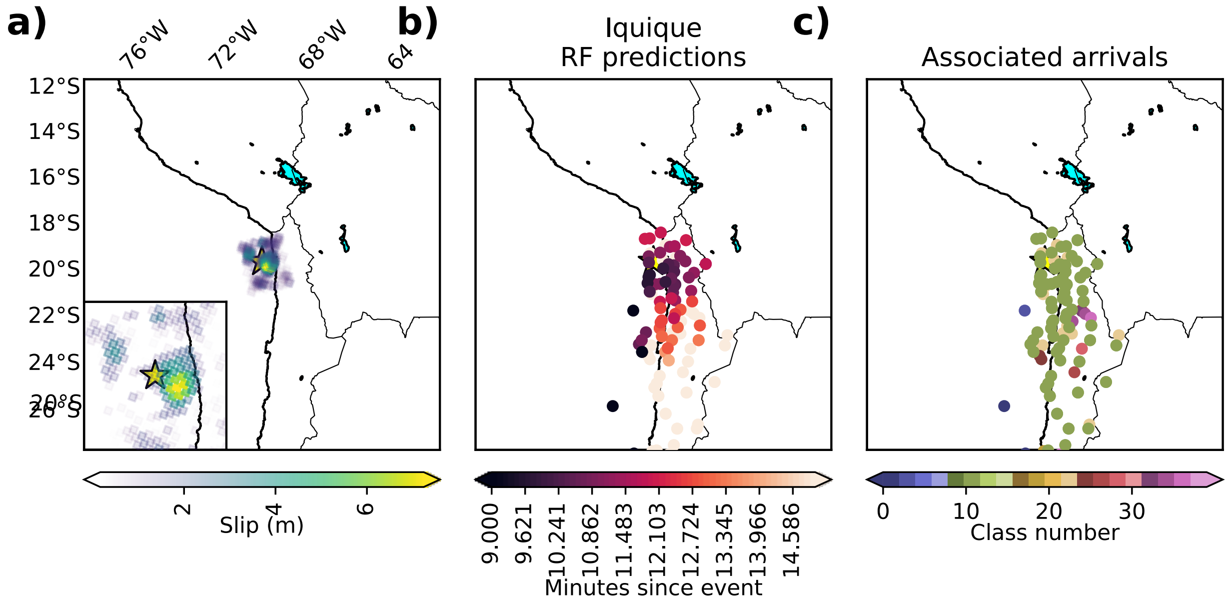

Figure 6. Ionospheric maps for the 2014 Iquique earthquake. (a) map showing the epicenter location (yellow star) and surface projection of the fault slip (in m) as green to yellow patches. (b) CID detections using our RF-based classifier and time picker, and (c) association classes determined from confirmed detections. CID coordinates were calculated at the intersection point between the LOS and the ionospheric layer using $H_{\text{ion}} = 250$ km. These maps can be generated 15 minutes after the event.

5 DISCUSSION

Monitoring procedures NRT-compatible require both high accuracy and low computational time. To provide an estimate of our algorithm's computational time, we show in Figure 7 the cost associated with detection, arrival-time picking, and association steps after the 2011 Tohoku event at station 0908 and satellite G05 (Figure 7a) on a single CPU (Dell T5610 Intel Xeon E5-2630 v2 2.6Ghz 6 CPUs 64GB RAM on CentOS 7). The computational time for feature extraction, classification, validation, and time picking for a single satellite/station pair is always below 1 s and is dominated by RF steps (Figure 7b). This result suggests that a similar detection methodology, trained with higher sampling-rate data, could be implemented for NRT applications up to 1 Hz. Note that the time picking step is only present when a detection occurs which explains the jump in computational cost around 7 mn after the earthquake.

We observe a significant increase in computational cost across the network 9 mn after the earthquake in Figure 7c. This jump in association cost corresponds to the earthquake-induced acoustic wave reaching the ionosphere which leads to a large number of detections at each combination of satellite/station (see Figure 7d). This association procedure is computationally expensive since it must scan through all possible neighbors of each new detection to update association classes, which scales linearly with the number of new detections. Yet, the maximum cost for one time step over the whole network is less than 6 s. It takes around 1 s to process 10 new detections, at a given time, over a

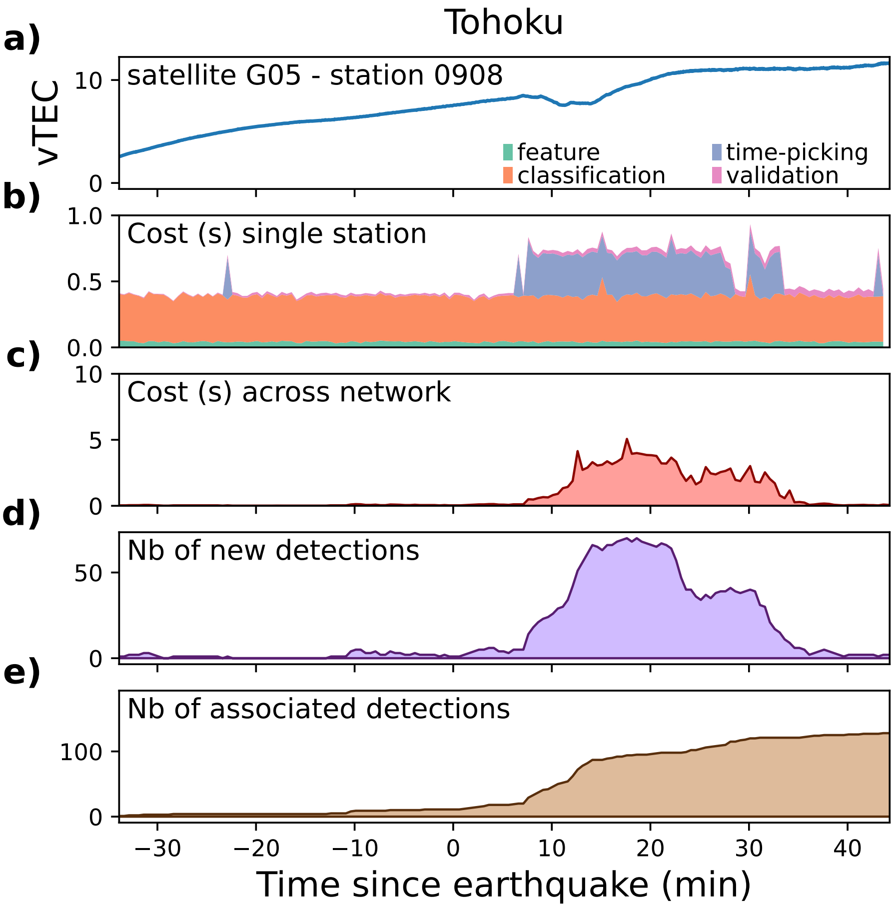

Figure 7. Computational cost associated with detection, arrival-time picking, and association steps after the 2011 Tohoku earthquake. (a) vTEC timeseries for satellite G07 and station 0048. (b) stack plot of computational time (s) for pre-processing and feature extraction (green), RF classification (orange), RF arrival-time picking (blue), and confirmation (pink) steps. (c) Computational cost (s) at each time iteration of the association procedure. (d) number of new detections per time iteration. (e) number of associated detections up to current time iteration.

network of about 100 satellites/stations. The number of associated detections reaches a plateau about 13 mn after the earthquake (see Figure 7e) which corresponds to the end of the association of all first CID arrivals.

The practical implementation of our detection/association procedure will require an efficient internet between the relevant GNSS stations to collect and extract timeseries for classification in NRT. However, because the overall computational cost of one time iteration using our method is below 6 s on a single CPU using non-compiled Python codes, at least 24 s are available for data acquisition and processing with waveforms sampled at 30 s. The association step is currently the most costly ($\sim 90\%$ of the total cost) but can be run in parallel to the other detection steps. Note that we also explored the feasibility of using our model to detect CIDs at a higher sampling rate by extracting input features without downsampling input data (see Supplementary Section S6). Our RF detection model always shows detection probabilities $> 50\%$ using a 1 s sampling time but still predict a strong increase in detection probability around the CID arrival. This suggests that increasing the detection threshold to higher values (e.g., from 50% to 70%) would enable implementation of our detection method at higher sampling-rates at the cost of a higher false positive likelihood.

Our model seems to be also able to detect vTEC variations associated with volcanic explosions and Rayleigh waves (see Supplementary Section S8). This suggests that a dataset of volcanic-induced and Rayleigh waves vTEC waveforms should be built and used to train an efficient discriminator between noise, earthquake, and volcanic phases. However, the discrimination between TEC signals from volcanic or seismic origin can easily be done by comparing the predicted arrival times at the ionospheric points to the distribution of seismic events in seismic catalogs which are available in NRT (Thompson et al. 2019).

6 CONCLUSIONS

We introduced an automatic procedure for detection, arrival-time picking, and association of CIDs. Detection and arrival time picking steps are performed using random forests trained over a CID dataset built from 12 earthquake events. These methods show excellent classification results with 96% true positive rate and 96% true negative rate, and arrival-time accuracy with an average error $<$ 20 s using a 120 s time delay since the first detection window. Our model also outperforms threshold-based detection methods in terms of both recall and precision. Our analytical classification procedure accurately associates all arrivals corresponding to the same wavefront. Classification results also indicate that low signal-to-noise ratio arrival that were not picked by human analysts could also captured by our RF detection model.

The performance of our automated procedure is promising for future NRT applications, including the use of CID arrival times for construction of ionospheric images of seismic sources. The first demonstration of seismo-ionospheric imagery was based on retrospective analysis of CID generated by the 2011 Tohoku earthquake (Astafyeva et al. 2011, 2013a). Here we show that our newly developed method can generate such images in NRT. Note that the position of ionospheric detection points is dependent on the altitude of detection H_{ion}. The latter parameter is not known precisely, but it is presumed to be around the height of ionospheric ionization maximum, i.e. around 250-350 km, depending on solar, geomagnetic, seasonal and diurnal conditions. Future studies should focus on determining the real H_{ion} in order to obtain accurate source locations.

Acquiring labeled vTEC data from additional events which will significantly improve the generalization abilities of our RF models. Additionally, the choice of features made in this paper could be further refined to obtain better accuracy (Han & Kim 2019). More accurate RF classifications could also alleviate the need for a validation step presented in Section 3.4. However, RF memory costs increase exponentially with tree depth, and consequently dataset size, $\sim 2^D$, with D the tree depth (Louppe 2014; Solé et al. 2014). The RF classification model is only about 70 mb but will grow considerably larger with new data. With a larger dataset, image segmentation ML techniques such as standard convolutional neural networks (Ross et al. 2018, 2019), transformers (Mousavi et al. 2020) or residual networks (Mousavi et al. 2019) applied on non-engineered inputs such as spectrograms could lead to substantial improvements in accuracy and memory costs for both classification and arrival time picking steps.

The proposed association algorithm does not incorporate any information about the source nor the atmospheric dynamics. This procedure could be improved by assessing the consistency of arrival time differences across a network of satellites and stations using a range of possible sources, similarly to the methods used for the automated production of seismic bulletins (Draelos et al. 2015). In contrast to seismic media, atmospheric velocities, i.e., winds, are time-dependent which introduces further complexity when computing theoretical source-receiver arrival times. Fast simulations of acoustic wave propagation up to the ionosphere with realistic atmospheric specifications would greatly improve the classification between true and false arrivals and enable the localization of the largest surface displacements (Bagiya et al. 2019; Inchin et al. 2021; Zedek et al. 2021). Finally, to confirm the detection of an earthquake across a given network and trigger an alert for human analysts, an additional heuristic could be implemented based, for example, on the number of detections per association class.

APPENDIX A: LIST OF EVENTS

The list of events compiled in our CID dataset is described in Table A1.

APPENDIX B: COMPARISON OF RF-BASED METHOD TO ANALYTICAL DETECTORS

To further assess the RF classification performance, we compare the results to two analytical detection methods: 1) a Short-Time Average / Long-Time Average (STA/LTA) detection method, and 2) a derivative-based threshold method. The STA/LTA method requires to set four parameters: the STA and LTA time windows and two thresholds to activate and deactivate the detection trigger. The STA window represents the average duration of expected earthquake signals while the LTA window captures the average TEC noise amplitude. The STA/LTA method employed here uses a 60 s STA window and a 400 s LTA window. A detection is triggered if the STA/LTA threshold reaches 2.5 while the end of a wavetrain is chosen where the threshold goes below 0.5. This trigger value of 2.5, lower than employed at seismic stations, is used to make sure we capture each arrival, i.e., to increase the true positive rate. Parameters are chosen empirically and could be improved with a thorough investigation of the STA/LTA accuracy over the whole dataset. However, fine tuning the hyperparameters increases the likelihood of over-fitting a specific dataset. This shows the advantage of using a ML-based approach that relies on an efficient optimization procedure enabling us to reach high accuracy without strong overfitting.

The analytical method used for comparison, referred to as ”AN”, is based on the analysis of TEC rate-of-change. Maletckii & Astafyeva (2021) noticed that, in a majority of cases, the CID are characterized by a rapid and high increase of TEC. To capture the CID arrival we therefore suggest to analyze the rate of TEC change between the two consecutive epochs, between every two and every three epochs:

$$\partial vTEC_1 = |vTEC_i - vTEC_{i+1}|, \tag{B.1}$$

$$\partial vTEC_2 = |vTEC_i - vTEC_{i+2}|, \tag{B.2}$$

$$\partial vTEC_3 = |vTEC_i - vTEC_{i+3}|, \tag{B.3}$$

$$\partial vTEC_4 = |vTEC_i - vTEC_{i+4}|, \tag{B.4}$$

where the subscript i corresponds to the time step t_i. The vTEC at epoch i is considered as the CID arrival if each slope $\partial vTEC_1, \partial vTEC_2$, and $\partial vTEC_3$ (and $\partial vTEC_4$ for 1s data) are greater that the thresholds shown in Table A2. These threshold values were determined analytically over multiple events. Detections are confirmed if 12 consecutive time steps fulfill the threshold conditions described in Table A2.

To assess the performance of each method, we determine the False and True negative and positive rates over the waveforms included in the testing dataset. To provide meaningful results, we scan entire waveforms (from 1-h to 2-h duration) instead of a few windows as done for RF training. Including entire waveforms means that more noise windows will be included than CID windows, which is an excellent test to assess the performance of each method in more realistic conditions (where CIDs are rare). We consider that a wavetrain, i.e., a time window characterized by an arrival time and a duration, classified as CID by any method is a true positive if it overlaps the true arrival by at least 70%.

Our RF-based detection method outperforms AN and STA in terms of true positive and negative rates (see Figures A1). We observe a lower true negative rate than determined during the RF validation step (see Figure 4c). This owes to the presence of much larger number of noise windows in the dataset. The STA/LTA filter also performs well to detect true arrivals. However, this high true positive rate comes at the cost of a low false positive rate, i.e., a large number of false alerts. The analytical method using only local time derivatives shows a large number of false negatives owing to presence of noise in the data.

ACKNOWLEDGMENTS

This work was supported by the French Space Agency (CNES, Project "RealDetect").

DATA AVAILABILITY

GNSS data are available from the following web-services: Japan GNSS Earth Observation System, GEONET (`http://datahouse1.gsi.go.jp/terras/terras_english.html`), GEONET Geological Hazard Information for New Zealand (`https://www.geonet.org.nz`), Scripps Orbit and Permanent Array Center (SOPAC, `http://sopac-old.ucsd.edu/dataBrowser.shtml`), National Seismological Centre, University of Chile (`http://gps.csn.uchile.cl`). Finite-fault data were downloaded from the US Geological Survey website (`https://earthquake.usgs.gov/`

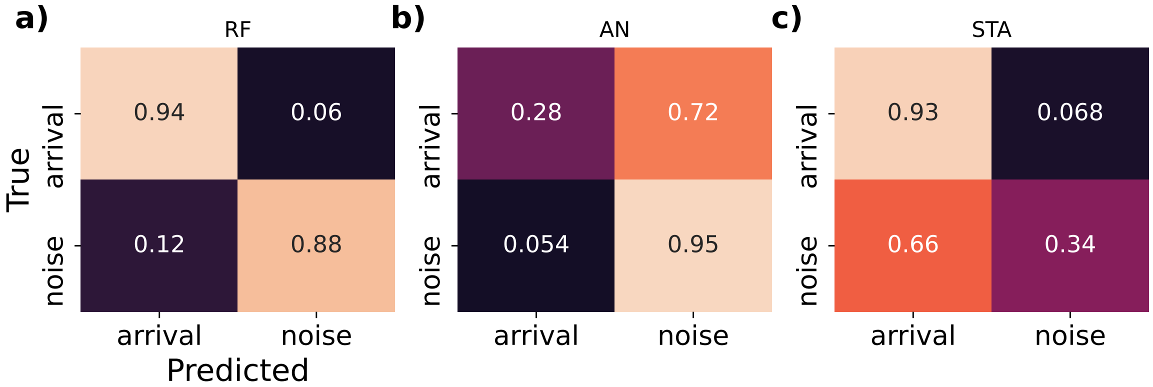

Figure A1. Confusion matrices calculated over the RF testing dataset consisting of 1-h to 2-h long waveforms for (a) the RF classification model, (b) the analytical time-derivative based model, and (c) the STA/LTA filter. Confusion matrices show from top to bottom and left to right, the True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), and True Negative Rate (TNR), such that: TPR = $TP/(TP + FN)$, TNR = $TN/(TN + FP)$, FPR = $TP/(TP + FP)$, and FNR = $TN/(TN + FN)$.

`earthquakes`). RF models, validation, and associations codes will be released upon publication on a FigShare and a GitHub repository.

REFERENCES

Afraimovich, E., Astafyeva, E., & Kiryushkin, V., 2006. Localization of the source of ionospheric disturbance generated during an earthquake, *International Journal of Geomagnetism and Aeronomy*, **6**, GI2002.

Allen, R., 1982. Automatic phase pickers: Their present use and future prospects, *Bulletin of the Seismological Society of America*, **72**(6B), S225–S242.

Astafyeva, E., 2019. Ionospheric detection of natural hazards, *Reviews of Geophysics*, **57**, 1265–1288.

Astafyeva, E. & Heki, K., 2009. Dependence of waveform of near-field coseismic ionospheric disturbances on focal mechanisms, *Earth, Planets, Space*, **61**, 939–943.

Astafyeva, E. & Shults, K., 2019. Ionospheric gnss imagery of seismic source: Possibilities, difficulties, and challenges, *Journal of Geophysical Research: Space Physics*, **124**(1), 534–543.

Astafyeva, E., Heki, K., Afraimovich, E., Kiryushkin, V., & Shalimov, S., 2009. Two-mode long-distance propagation of coseismic ionosphere disturbances, *J. Geophys. Res.*, **118**, A10307.

Astafyeva, E., Lognonné, P., & Rolland, L. M., 2011. First ionosphere images for the seismic slip on the example of the tohoku-oki earthquake, *Geophys. Res. Letters*, **38**, L22104.

Astafyeva, E., Rolland, L. M., Lognonné, P., Khelfi, K., & Yahagi, T., 2013a. Parameters of seismic source as deduced from 1hz ionospheric gps data: case-study of the 2011 tohoku-oki event, *Journal of Geophys. Research*, **118**, 5942–5950.

Astafyeva, E., Shalimov, S., Olshanskaya, E., & Lognonné, P., 2013b. Ionospheric response to earthquakes of different magnitudes: larger quakes perturb the ionosphere stronger and longer, *Geophys. Res. Letters*, **40**, 1675–1681.

Astafyeva, E., Rolland, L. M., & Sladen, A., 2014. Strike-slip earthquakes can also be detected in the ionosphere, *Earth and Planetary Science Letters*, **405**, 180–193.

Bagiya, M. S., Sunil, P. S., Sunil, A. S., & Ramesh, D. S., 2018. Coseismic contortion and coupled nocturnal ionospheric perturbations during 2016 kaikoura, mw 7.8 new zealand earthquake, *Journal of Geophysical Research: Space Physics*, **123**(2), 1477–1487.

Bagiya, M. S., Sunil, A., Rolland, L., Nayak, S., Ponraj, M., Thomas, D., & Ramesh, D. S., 2019. Mapping the impact of non-tectonic forcing mechanisms on gnss measured coseismic ionospheric perturbations, *Scientific reports*, **9**(1), 1–15.

Bagiya, M. S., Thomas, D., Astafyeva, E., Bletery, Q., Lognonné, P., & Ramesh, D. S., 2020. The ionospheric view of the 2011 tohoku-oki earthquake seismic source: the first 60 seconds of the rupture, *Scientific reports*, **10:5232**.

Belehaki, A., Tsagouri, I., Altadill, D., Blanch, E., Borries, C., Buresova, D., Chum, J., Galkin, I., Juan, J. M., Segarra, A., et al., 2020. An overview of methodologies for real-time detection, characterisation and tracking of traveling ionospheric disturbances developed in the techtide project, *Journal of Space Weather and Space Climate*, **10**, 42.

Breiman, L., 2001. Random forests, *Machine learning*, **45**(1), 5–32.

Cahyadi, M. N. & Heki, K., 2015. Coseismic ionospheric disturbance of the large strike-slip earthquakes in north sumatra in 2012 mw dependence of the disturbance amplitudes, *Geophysical Journal International*, **200**(1), 116–129.

Draelos, T. J., Ballard, S., Young, C. J., & Brogan, R., 2015. A new method for producing automated seismic bulletins: Probabilistic event detection, association, and location, *Bulletin of the Seismological Society of America*, **105**(5), 2453–2467.

Efendi, E. & Arikan, F., 2017. A fast algorithm for automatic detection of ionospheric disturbances: Drot, *Advances in Space Research*, **59**(12), 2923–2933.

Geurts, P., Ernst, D., & Wehenkel, L., 2006. Extremely randomized trees, *Machine learning*, **63**(1), 3–42.

Hammer, C., Ohrnberger, M., & Faeh, D., 2013. Classifying seismic waveforms from scratch: a case study in the alpine environment, *Geophysical Journal International*, **192**(1), 425–439.

Han, S. & Kim, H., 2019. On the optimal size of candidate feature set in random forest, *Applied Sciences*, **9**(5), 898.

Heki, K., 2006. Explosion energy of the 2004 eruption of the asama volcano, central japan, inferred from ionospheric disturbances, *Geophys. Res. Lett.*, **33**, L17101.

Heki, K., 2021. Ionospheric disturbances related to earthquakes in ionospheric dynamics and applications, *Geophys. Monograph, 260, edited by C. Huang, G. Lu, Y. Zhang, and L. J. Paxton*, pp. 511–526.

Heki, K. & Ping, J., 2005. Directivity and apparent velocity of the coseismic ionospheric disturbances observed with a dense gps array, *Earth and Planetary Science Letters*, **236**(3), 845–855.

Heki, K., Otsuka, Y., Choosakul, N., Hemmakorn, N., Komolmis, T., & Maruyama, T., 2006. Detection of ruptures of andaman fault segments in the 2004 great sumatra earthquake with coseismic ionospheric disturbances, *J. Geophys. Res.*, **111**, B09313.

Hibert, C., Mangeney, A., Grandjean, G., Baillard, C., Rivet, D., Shapiro, N. M., Satriano, C., Maggi, A., Boissier, P., Ferrazzini, V., et al., 2014. Automated identification, location, and volume estimation of rockfalls at piton de la fournaise volcano, *Journal of Geophysical Research: Earth Surface*, **119**(5), 1082–1105.

Hofmann-Wellenhof, B., Lichtenegger, H., & Wasle, E., 2008. *GNSS-Global Navigation Satellite System*, Springer.

Inchin, P., Snively, J., Kaneko, Y., Z., D., M., & Komjathy, A., 2021. Inferring the evolution of a large earthquake from its acoustic impacts on the ionosphere., *AGU Advances*, **2**.

Kakinami, Y., Saito, H., Yamamoto, T., Chen, C.-H., Yamamoto, M., Nakajima, K., Liu, J.-Y., & Watanabe, S., 2021. Onset altitudes of co-seismic ionospheric disturbances determined by multiple distributions of gnss tec after the foreshock of the 2011 tohoku earthquake on march 9, 2011, *Earth and Space Sciences*.

Kamogawa, M., Orihara, Y., Tsurudome, C., Tomida, Y., Kanaya, T., & Ikeda, D., e. a., 2016. A possible space-based tsunami early warning system using observations of the tsunami ionospheric hole, *Scientific Reports*, **6:37989**.

Katsumata, A., Ueno, H., Aoki, S., Yasushiro, Y., & Barrientos, S., 2013. Rapid magnitude determination from peak amplitudes at local stations, *Earth, Planets Space*, **65**, 843–853.

Komjathy, A., Yang, Y., Meng, X., Vekhoglyadova, O., Mannucci, A., & Langley, R., 2016. Review and perspectives: Understanding natural-hazards-generated ionospheric perturbations using gps measurements and coupled modeling, *Radio Science*, **51**, 951–961.

LaBrecque, J., Rundle, J., Bawden, G., Surface, E., & Area, I. F., 2019. Global navigation satellite system enhancement for tsunami early warning systems, *Global Assessment Report on Disaster Risk Reduction*.

Lee, R., Rolland, L., & Mykesell, T., 2018. Seismo-ionospheric observations, modeling and backprojection of the 2016 kaikoura earthquake, *Bulletin of the Seismological Society of America*, **108**(3B), 1794–1806.

Louppe, G., 2014. Understanding random forests: From theory to practice, *arXiv preprint arXiv:1407.7502*.

Maletckii, B. & Astafyeva, E., 2021. Determining spatio-temporal characteristics of coseismic travelling ionospheric disturbances (ctid) in near real-time, *Scientific Reports*, **11**.

Manta, F., Occhipinti, G., Feng, L., & Hill, E., 2020. Rapid identification of tsunamigenic earthquakes using gnss ionospheric sounding, *Scientific Reports*, **10:11054**.

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C., 2019. Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection, *Scientific reports*, **9**(1), 1–14.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nature communications*, **11**(1), 1–12.

Occhipinti, G., Aden-Antoniow, F., Bablet, A., Molinie, J.-P., & Farges, T., 2018. Surface waves magnitude estimation from ionospheric signature of rayleigh waves measured by doppler sounder and oth radar, *Scientific Reports*, **8:1555**.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E., 2011. Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**, 2825–2830.

Provost, F., Hibert, C., & Malet, J.-P., 2017. Automatic classification of endogenous landslide seismicity using the random forest supervised classifier, *Geophysical Research Letters*, **44**(1), 113–120.

Rakoto, V., Lognonné, P., Rolland, L., & Coisson, P., 2018. Tsunami wave height estimation from gps-derived ionospheric data, *J. Geophys. Res.*, **123**, 4329–4348.

Ravanelli, M., Occhipinti, G., Savastano, G., Komjathy, A., Shume, E. B., & Crespi, M., 2021. Gnss total variometric approach: first demonstration of a tool for real-time tsunami genesis estimation, *Scientific reports*, **11**(1), 1–12.

Rolland, L., Vergnolle, M., Nocquet, J.-M., Sladen, A., Dessa, J.-X., Tavakoli, F., Nankali, H., & Cappa, F., 2013. Discriminating the tectonic and non-tectonic contributions in the ionospheric signature of the 2011, mw7.1, dip-slip van earthquake, eastern turkey, *Geophys. Res. Lett.*, **40**.

Rolland, L. M., Occhipinti, G., Lognonné, P., & Loevenbruck, A., 2016. Ionospheric gravity waves detected offshore hawaii after tsunami, *Geophys. Res. Lett.*, **37**, L17101.

Ross, Z. E., Meier, M.-A., & Hauksson, E., 2018. P wave arrival picking and first-motion polarity determination with deep learning, *Journal of Geophysical Research: Solid Earth*, **123**(6), 5120–5129.

Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., Zhan, Z., Simons, M., Fielding, E. J., Yun, S.-H., et al., 2019. Hierarchical interlocked orthogonal faulting in the 2019 ridgecrest earthquake sequence, *Science*, **366**(6463), 346–351.

Shults, K., Astafyeva, E., & Adourian, S., 2016. Ionospheric detection and localization of volcano eruptions on the example of the april 2015 calbuco events, *Journal of Geophysical Research: Space Physics*, **121**(10), 10,303–10,315.

Solé, X., Ramisa, A., & Torras, C., 2014. Evaluation of random forests on large-scale classification problems using a bag-of-visual-words representation, in *CCIA*, pp. 273–276.

Thomas, D., Bagiya, M. S., Sunil, P. S., Rolland, L., Sunil, A. S., Mikesell, T. D., Nayak, S., Mangalampalli, S., & Ramesh, D. S., 2018. Revelation of early detection of co-seismic ionospheric perturbations in gps-tec from realistic modelling approach: Case study, *Scientific reports*, **8**(1), 1–10.

Thompson, E. M., McBride, S. K., Hayes, G. P., Allstadt, K. E., Wald, L. A., Wald, D. J., Knudsen, K. L., Worden, C. B., Marano, K. D., Jibson, R. W., & Grant, A. R. R., 2019. USGS Near-Real-Time Products—and Their Use—for the 2018 Anchorage Earthquake, *Seismological Research Letters*, **91**(1), 94–113.

Tsai, H.-F., Liu, J.-Y., Lin, C.-H., & Chen, C.-H., 2011. Tracking the epicenter and the tsunami origin with gps ionosphere observation, *Earth, Planets Space*, **63**, 859–862.

Wenner, M., Hibert, C., van Herwijnen, A., Meier, L., & Walter, F., 2021. Near-real-time automated classification of seismic signals of slope failures with continuous random forests, *Natural Hazards and Earth System Sciences*, **21**(1), 339–361.

Wright, T., Houlie, N., Hildyard, M., & Iwabuchi, T., 2012. Real-time, reliable magnitudes for large earthquakes from 1 hz gps precise point positioning: The 2011 tohoku-oki (japan) earthquake, *Geophys. Res. Lett.*, **38**(L12302).

Zedek, F., Rolland, L. M., Dylan Mikesell, T., Sladen, A., Delouis, B., Twardzik, C., & Coïsson, P., 2021. Locating surface deformation induced by earthquakes using gps, glonass and galileo ionospheric sounding from a single station, *Advances in Space Research*.

Table A1. List of events included in the dataset. Events are sorted by magnitude

Event Reference	Mag.	Lat. ; Lon.	Date (DD/MM/YY)	Time (UTC)	Min. signal duration (s)	Sat.	Samp
Tohoku Astafyeva et al. (2011, 2013a)	9.1	38.3 ; 142.37	11/03/2011	05:46:23	800	G26 G05	1s, 30s
Sumatra 1 Astafyeva et al. (2014)	8.6	2.35 ; 92.8	11/04/2012	08:38:37	300	G32	15s
Tokachi Heki & Ping (2005)	8.3	41.78 ; 143.90	25/09/2003	19:50:06	440	G13 G24	30s
Illapel Bagiya et al. (2019)	8.3	-31.57; -71.61	16/09/2015	22:54:32	600	G25,G12 G24	15s, 30s
Sumatra 2 Astafyeva et al. (2014)	8.2	0.90 ; 92.31	11/04/2012	10:43:09	300	G32	15s
Iquique Bagiya et al. (2019)	8.2	-19.61 ; -70.77	01/04/2014	23:46:47	700	G01,G20 G23	15s, 30s
Macquarie Astafyeva et al. (2014)	8.1	-49.91 ; 161.25	23/12/2004	14:59:03	550	G05	30s
Fiordland Astafyeva et al. (2013b)	7.8	-45.75 ; 166.58	15/07/2009	09:22:29	300	G20	30s
Kaikoura Bagiya et al. (2018)	7.8	42.757 ; 173.077	13/11/2016	11:02:56	550	G20 G29	1s, 30s
Sanriku Thomas et al. (2018); Astafyeva & Shults (2019)	7.3	38.44 ; 142.84	09/03/2011	02:45:20	200	G07, G10 G08	1s, 30s
Kii Heki & Ping (2005)	7.2	33.1 ; 136.6	05/09/2004	10:07:07	425	G15	30s
Chuetsu Cahyadi & Heki (2015)	6.6	37.54 ; 138.45	16/07/2007	01:12:22	300	G26	30s

Table A2. Slope parameters for different sampling rates used by the analytical detector AN.

Sampling (s)	s_1 (TECU/epoch)	s_2 (TECU/epoch)	s_3 (TECU/epoch)	s_4 (TECU/epoch)
1	0.017	0.027	0.045	0.05
15	0.08	0.125	0.12	-
30	0.11	0.18	-	-

Supporting Information for "Near-real-time detection of co-seismic ionospheric disturbances using machine learning"

Quentin Brissaud[1] and Elvira Astafyeva[2]

[1]NORSAR, Kjeller, Norway
[2]Université de Paris, Institut de Physique du Globe de Paris (IPGP), CNRS UMR7154, 35-39 Rue Hélène Brion, 75013 Paris, France

Contents of this file

1. Texts S1 to S8
2. Figures S1 to S10
3. Table S1

Introduction

This Supplementary file contains additional details about:

- Text S1 List of input features
- Text S2 R2 cross correlations of input features and clustering analysis
- Text S3 Sensitivity of classification accuracy to number of validation points
- Text S4 Arrival time picking optimization
- Text S5 Time evolution of detected arrivals
- Text S6 Detection of CIDs at higher sampling rates
- Text S7 Impact of H_{ion} on association classes
- Text S8 Detection of ionospheric signal from volcanic eruptions and Rayleigh waves

S1 List of input features

All input features used to train the RF classifier presented in Section 3 are described in Table table S1.

The distribution of input features over our training and testing datasets is shown in Figure S1.

S2 R2 cross correlations of input features and clustering analysis

The RF model can provide an estimate of the relative feature importance through the calculation of the Gini's impurity during training. Figure 4b shows that the three best features have been extracted from the timeseries in contrast to other signal classification studies Wenner et al. (2021). However, the calculation of feature importance can be biased when considering continuous or high-cardinality categorical variables or when inputs features are co-linear. To assess the input features correlations within our CID dataset, we show in Figure S2 the R2 cross correlations. Co-linearity is present in our input dataset between spectral and time-series features which indicates a potential bias in variable importance results.

Table S1: List of attributes. $Nyf = 0.0165$ Hz is the Nyquist frequency. These attributes are commonly-used in signal-classification studies. We refer the reader to the following references for more details: (Bessason et al., 2007; Curilem et al., 2009; Hammer et al., 2012; Hibert et al., 2014; Provost et al., 2017; Wenner et al., 2021)

Short name	Description
W0	Ratio of the mean over the maximum of the envelope signal
W1	Ratio of the median over the maximum of the envelope signal
W2	Kurtosis of the raw signal (peakness of the signal)
W3	Kurtosis of the envelope
W4	Skewness of the raw signal
W5	Skewness of the envelope
W6	Number of peaks in the autocorrelation function
W7	Energy in the first third part of the autocorrelation function
W8	Energy in the remaining part of the autocorrelation function
W9	W7/W8
W10	Maximum of the envelope signal
W11	Energy of the signal filtered in 0.001-0.005 Hz
W12	Energy of the signal filtered in 0.005-0.015 Hz
W13	Kurtosis of the signal filtered in 0.001-0.005 Hz
W14	Kurtosis of the signal filtered in 0.005-0.015 Hz
S0	Mean of the Fourier transform (FT)
S1	Maximum of the FT
S2	Frequency at the FT maximum
S3	Frequency at the FT centroid
S4	Frequency at the FT 1st quartile
S5	Frequency at the FT 2nd quartile
S6	Median of the normalized FT
S7	Variance of the normalized FT
S8	Number of Fourier transform peaks (> 0.75 FT max.)
S9	Mean of FT peaks (S8)
S10	Gyration radius
S11	Energy up to 0.5Nyf Hz
S12	Energy up to 0.75Nyf Hz
S14	Energy up to 1.0Nyf Hz
FT0	Kurtosis of the maximum of all Fourier transforms (FTs) as a function of time
FT1	Kurtosis of the maximum of all FTs as a function of frequency
FT2	Mean ratio between the maximum and the mean of all FTs
FT3	Mean ratio between the maximum and the median of all FTs
FT4	Number of peaks in the curve showing the temporal evolution of the FTs maximum
FT5	Number of peaks in the curve showing the temporal evolution of the FTs mean
FT6	Number of peaks in the curve showing the temporal evolution of the FTs median
FT7	ratio of FT4 over FT5
FT8	ratio of FT4 over FT6
FT9	Number of peaks in the curve of the temporal evolution of the FTs central frequency
FT10	Number of peaks in the curve of the temporal evolution of the FTs maximum frequency
FT11	FT9/FT10
FT12	Mean distance between the curves of the temporal evolution of the FTs maximum and mean frequency
FT13	Mean distance between the curves of the temporal evolution of the FTs maximum and median frequency
FT14	Mean distance between the 1st quartile and the median of all FTs as a function of time
FT15	Mean distance between the 3rd quartile and the median of all FTs as a function of time
FT16	Mean distance between the 3rd quartile and the 1st quartile of all FTs as a function of time

Additionally, the significant overlap between distributions in Figure 4b motivates the choice of a large number of features to properly discriminate between each class. However, multidimensional clusters in input data are difficult to represent in 2d which prevents human interpretation. Two standard methods to facilitate the visualization of clusters in data are called Principal Component Analysis (PCA), and T-Distributed Stochastic Neighbor Embedding (TSNE, Van der Maaten and Hinton (2008)). PCA consists in project the input features in onto a space with orthogonal, i.e., uncorrelated, vector basis such that the greatest variance of the data comes to lie on the first coordinate. TSNE builds probability distributions over pairs of high-dimensional vectors so that similar data points have a higher probability while dissimilar data points are assigned a lower probability. Then, TSNE constructs a similar probability distribution over the points in the low-dimensional space, and it minimizes the Kullback–Leibler divergence between the two distributions with respect to the locations of the points in the map. No obvious clusters can be identified in the two first components of the PCA (Figure S3a) which indicates that only complex nonlinear relationships can help discriminating signals between noise and CID classes. TNSE non-linear mapping suggests two clusters (Figure S3b). These clusters are particularly visible for events showing strong CID amplitudes such as Sanriku. However, events with low signal-to-noise ratio CIDs, such as Kii, do not show a significant overlap between noise and arrival clusters. This further highlights the complexity of this classification problem.

S3 Sensitivity of classification accuracy to number of validation points

The heuristic model presented in Section 3.4 relies on a single parameter to confirm a detection: the number of consecutive time steps with a detection probability $> 50\%$, referred to as N_d. To determine the optimal value of N_d we varied this parameter between 2 and 5, and computed the true and false positive and negative rates over our true-arrival dataset, i.e., 2000 s waveforms centered on each true arrival. In Figure S4, we observe that the variations in N_d (Nb points trigger) do not affect significantly the true and false positive rates. Because we observe a slight decrease in False positive rate with an increase in N_d, we select $N_d = 3$ as a trade-off between false alerts and time delay to confirm a detection.

S4 Arrival time picking optimization

The arrival time picking procedure is based on a RF model. This model takes vTEC time derivatives as an input and gives a time shift from the window central time as an output. The RF will therefore be sensitive to the window size, as larger windows increase the number of inputs and tend to complicate the picking procedure while small time windows lack data points to regularize the time picking problem. Additionally, the range of window overlap with the true wavetrain used for training plays a significant role on the RF performances. Using small overlaps will train the machine to pick arrivals on incomplete waveforms and therefore makes the problem more difficult. However this will enable the machine to more efficiently pick arrival times over the first detection time windows of a given wavetrain. We show in Figure S5, the variations in time picking accuracy with window size and overlap (called deviation). As a trade-off between errors and the ability of our RF model to pick arrival times over incomplete waveforms, we choose a window size similar to the RF classifier (see Section 3.2) and an overlap of 30%.

S5 Time evolution of detected arrivals

A requirement for NRT applications is to obtain alerts within 20mn after the event. Therefore, our detection and association procedure should trigger a valid alert as soon as possible in addition to providing accurate arrival times. In Figure S6, we show the evolution of the distribution of arrival times with time since the event for the earthquake Tohoku. We observe that after 12mn, we already observe a specific trend in arrival-time values highlighting that the acoustic energy is propagating from East to West. After 15mn, almost all hand-picked arrival times have been correctly determined by our model.

S6 Detection of CIDs at higher sampling rates

A machine learning model trained with data sampled at 30s might learn patterns that are invariant with frequency. To assess how our classification model performs on 1s data, we extracted features in each time window without downsampling waveforms. In addition, we used a 1s time shift between two consecutive time windows. Detection probability and picked arrival times are shown in Figure S7. Detection probabilities are always over 50%, i.e., RF classified the whole timeseries as a CID with the use of a detection threshold at 50%. Yet, we observe a significant increase in detection probability around 2.85 UT, from 60% to 95%, that matches the arrival of the CID. Jumps in detection probabilities indicates that using a larger detection threshold, such as $\geq 70\%$ instead of $\geq 50\%$, could enable the processing of higher sampling-rate data with of our algorithm. These larger probabilities owe to the additional noise introduced by higher frequencies when extracting input features. The higher-frequency spectral content can lead to substantial variations in certain input features. For example, energy peaks at higher frequencies, that would normally be smoothed out at lower frequencies, can drastically alter the envelope kurtosis and skewness, which are critical parameters for discrimination between noise and arrival windows. Nonetheless, the ability of our model to recover the true arrival time is extremely promising for near-real-time applications.

S7 Impact of H_{ion} on association classes

The position of ionospheric detection points is dependent on the altitude of detection H_{ion}, which could impact the association classes. To assess the sensitivity of the association classes on H_{ion}, we changed the altitude of the ionospheric points for the Tohoku event from 180 to 250 km. The location of the center of the main association class (light purple in Figure S8c) tends to shift towards the South-East with the increase in H_{ion}. While the location of the ionospheric points changes with H_{ion}, the true arrival times (Figure S8a) are still correctly associated in the same class (light purple in Figure S8c).

S8 Detection of ionospheric signal from volcanic eruptions and Rayleigh waves

Other low-frequency acoustic sources, such as volcanoes or surface Rayleigh waves can generate transient ionospheric perturbations. In particular, volcanic eruptions generate both infrasonic and gravito-acoustic signals in the 0.1-10 mHz frequency range known as Co-Volcanic Ionospheric Disturbances (CVID). While gravity waves show a much lower frequency content Hines (1960), near-epicentral CVID can show short-period signals with significant energy below 5 minutes Shestakov et al. (2021). We therefore first assessed the sensitivity of our RF model to travelling volcanic-induced ionospheric propagation using the example of the Calbuco volcanic eruption on April 22, 2015 Shults et al. (2016). In figure S9, we observe that the entire volcanic-induced gravito-acoustic wavetrain is classified as CID. This can be explained by the similarity of CIDs and CVIDs in the feature space due to significant energy at high frequencies corresponding to infrasound signals mixed with the graviy wavefield.

The atmospheric perturbations generated by seismic Rayleigh waves can also propagate to the ionosphere and be observed on TEC data (Rolland et al., 2011). Such signals typically show energy between XXX s and XXX s, similar to epicentral infrasound. Testing our method on a Rayleigh-wave signal observed after the XXX event, we observe that the transient signal is well captured and its arrival time accurately predicted (see Figure S10). This indicates that both epicentral and Rayleigh-wave infrasound can be observed and associated by our detection method.

References

E. Astafyeva, K. Heki, E. Afraimovich, V. Kiryushkin, and S. Shalimov. Two-mode long-distance propagation of coseismic ionosphere disturbances. *J. Geophys. Res.*, 118:A10307, 2009. doi: 10.1029/2008JA013853.

B. Bessason, G. Eiríksson, Ó. Thórarinsson, A. Thórarinsson, and S. Einarsson. Automatic detection of avalanches and debris flows by seismic methods. *Journal of Glaciology*, 53(182):461–472, 2007. doi: 10.3189/002214307783258468.

G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón. Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms. *Journal of volcanology and geothermal research*, 180(1):1–8, 2009. doi: 10.1016/j.jvolgeores.2008.12.002.

C. Hammer, M. Beyreuther, and M. Ohrnberger. A Seismic-Event Spotting System for Volcano Fast-Response Systems. *Bulletin of the Seismological Society of America*, 102(3):948–960, 06 2012. ISSN 0037-1106. doi: 10.1785/0120110167.

C. Hibert, A. Mangeney, G. Grandjean, C. Baillard, D. Rivet, N. M. Shapiro, C. Satriano, A. Maggi, P. Boissier, V. Ferrazzini, and W. Crawford. Automated identification, location, and volume estimation of rockfalls at piton de la fournaise volcano. *Journal of Geophysical Research: Earth Surface*, 119(5): 1082–1105, 2014. doi: https://doi.org/10.1002/2013JF002970.

C. O. Hines. Internal atmospheric gravity waves at ionospheric heights. *Canadian Journal of Physics*, 38 (11):1441–1481, 1960.

F. Provost, C. Hibert, and J.-P. Malet. Automatic classification of endogenous landslide seismicity using the random forest supervised classifier. *Geophysical Research Letters*, 44(1):113–120, 2017. doi: 10.1002/2016GL070709.

L. M. Rolland, P. Lognonné, and H. Munekane. Detection and modeling of Rayleigh wave induced patterns in the ionosphere. *J. Geophys. Res. Space Phys.*, 116(A5), 2011. doi: 10.1111/j.1651-2227.2005.tb01817.x.

N. Shestakov, A. Orlyakovskiy, N. Perevalova, N. Titkov, D. Chebrov, M. Ohzono, and H. Takahashi. Investigation of ionospheric response to june 2009 sarychev peak volcano eruption. *Remote Sensing*, 13 (4):638, 2021. doi: 10.3390/rs13040638.

K. Shults, E. Astafyeva, and S. Adourian. Ionospheric detection and localization of volcano eruptions on the example of the april 2015 calbuco events. *Journal of Geophysical Research: Space Physics*, 121(10): 10,303–10,315, 2016. doi: 10.1002/2016JA023382.

L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. URL `https://jmlr.csail.mit.edu/papers/volume9/vandermaaten08a/vandermaaten08a.pdf`.

M. Wenner, C. Hibert, A. van Herwijnen, L. Meier, and F. Walter. Near-real-time automated classification of seismic signals of slope failures with continuous random forests. *Natural Hazards and Earth System Sciences*, 21(1):339–361, 2021. doi: 10.5194/nhess-21-339-2021.

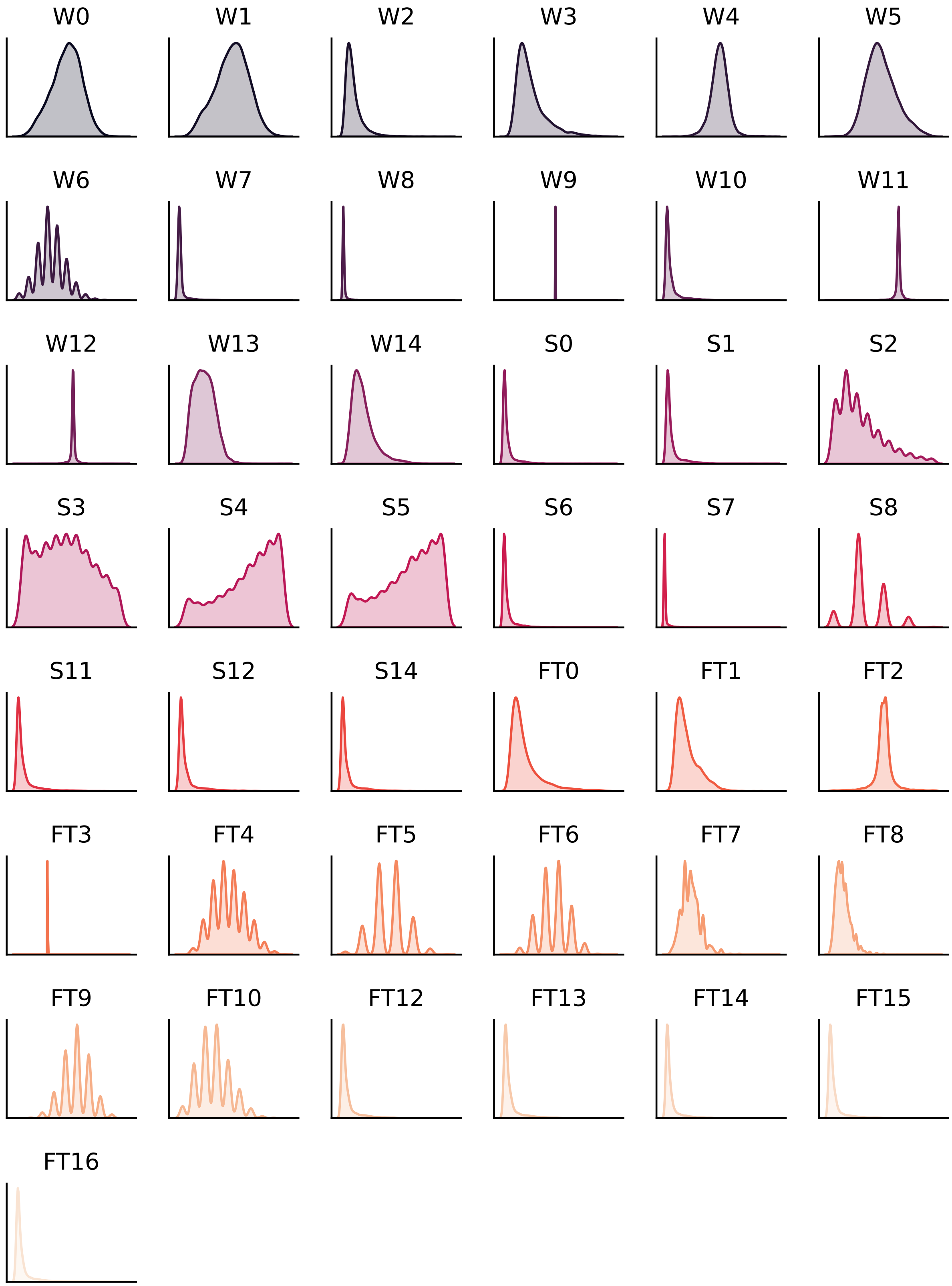

Figure S1: Probability density of each input features over our training and testing datasets. The short name of feature for each plot is shown above the plot. The description of each feature is given in Table table S1

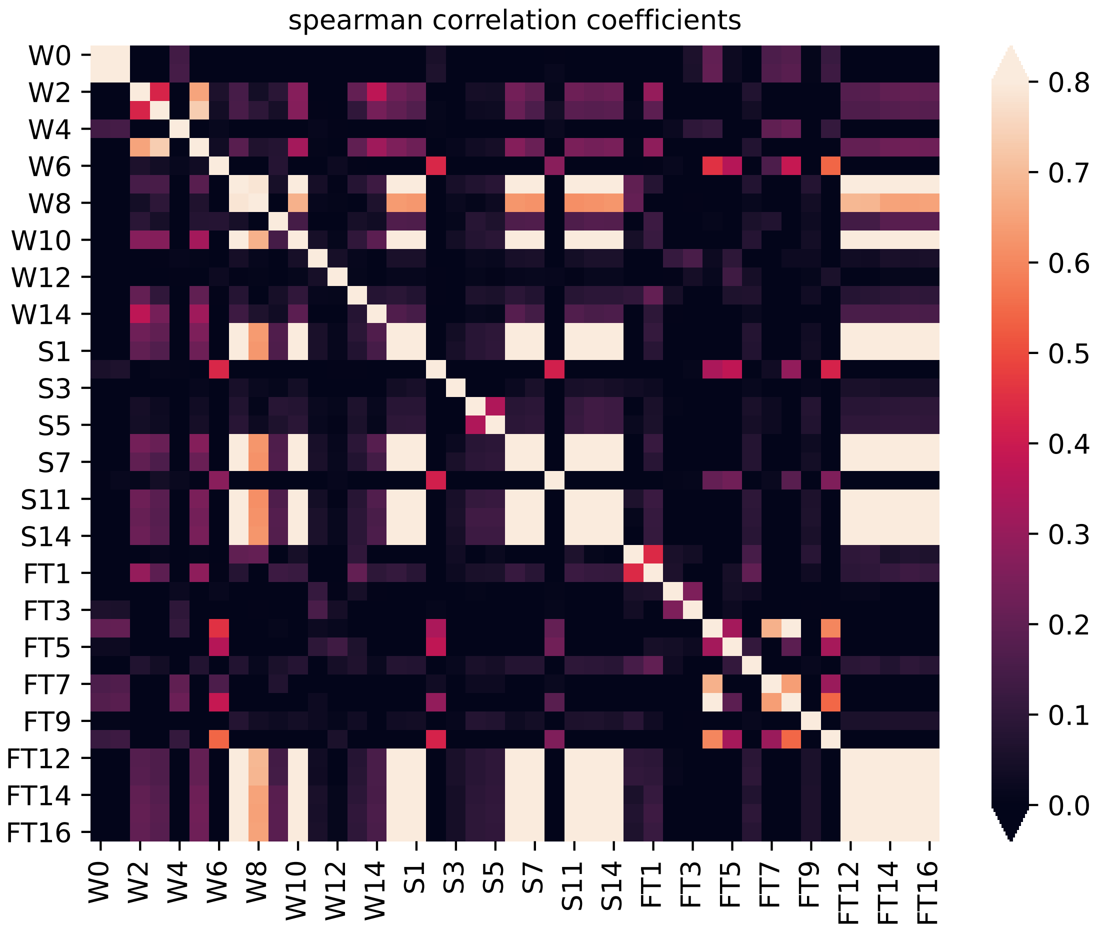

Figure S2: Spearman's correlation coefficients between each feature used for training. A description of each feature is given in Table table S1.

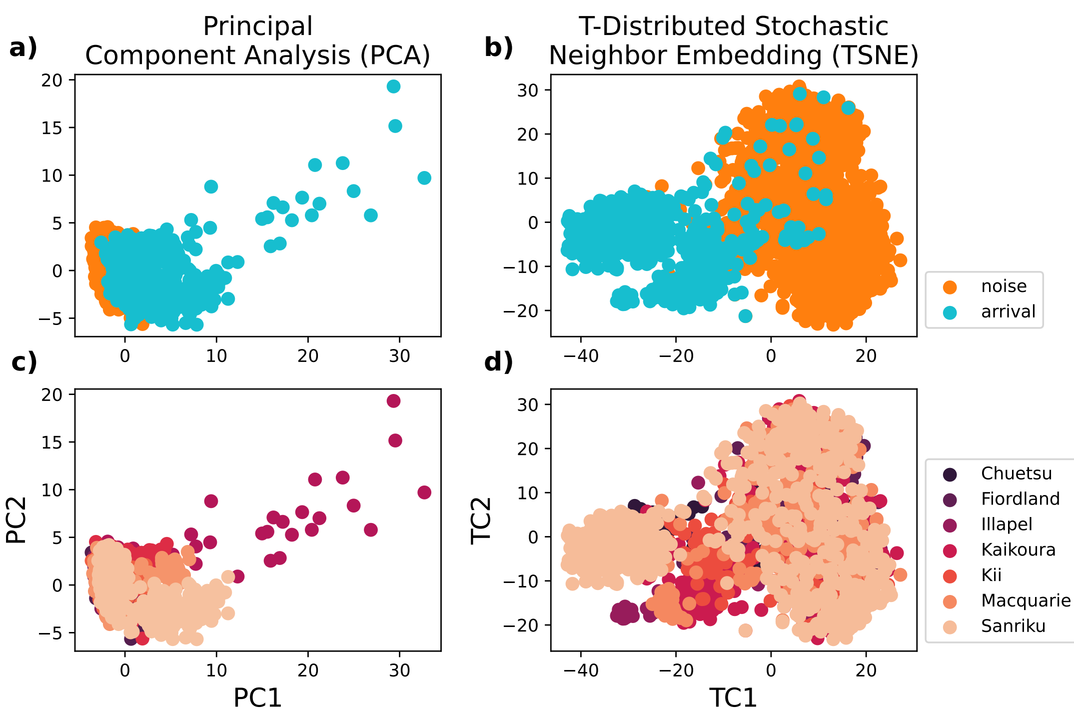

Figure S3: First versus second component of (a,c) a Principal Component Analysis (PCA) and (b,d) a T-Distributed Stochastic Neighbor Embedding (TNSE, Van der Maaten and Hinton (2008)). Points are colorcoded with (a,b) the detection class, and (c,d) the event name for the arrival class.

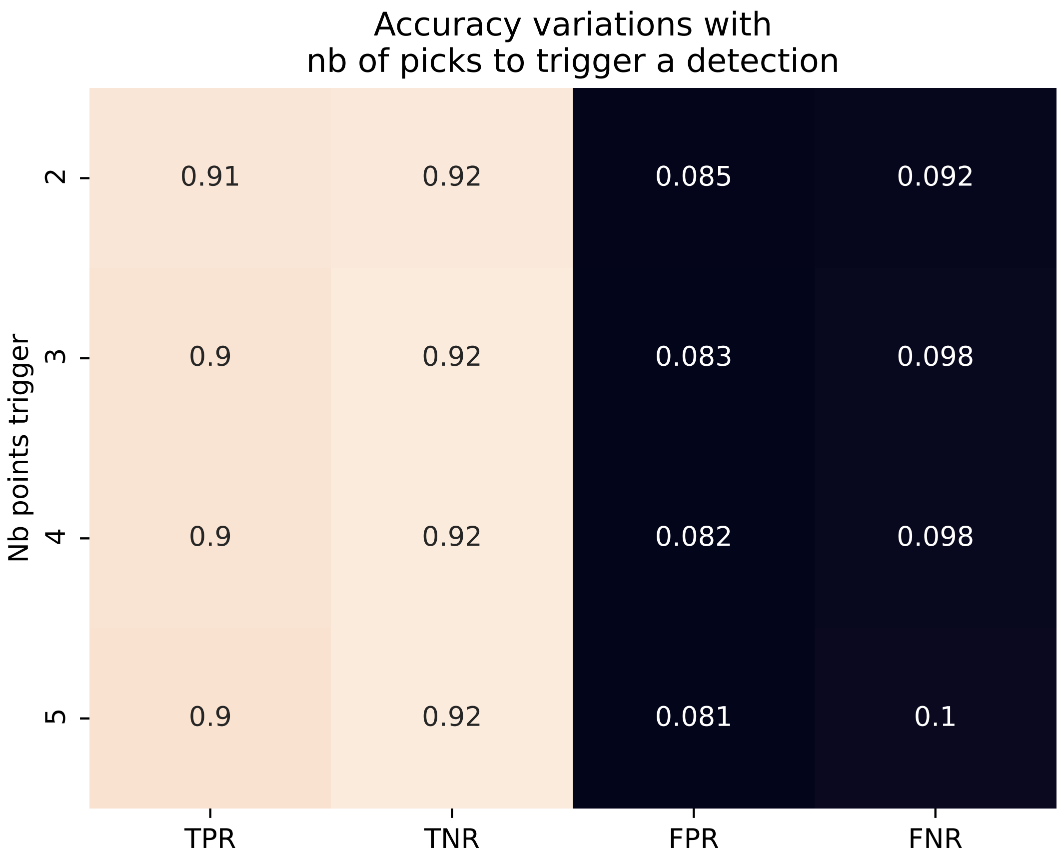

Figure S4: True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR) with the choice of number of time steps for validation in the heuristic model presented in Section 3.4

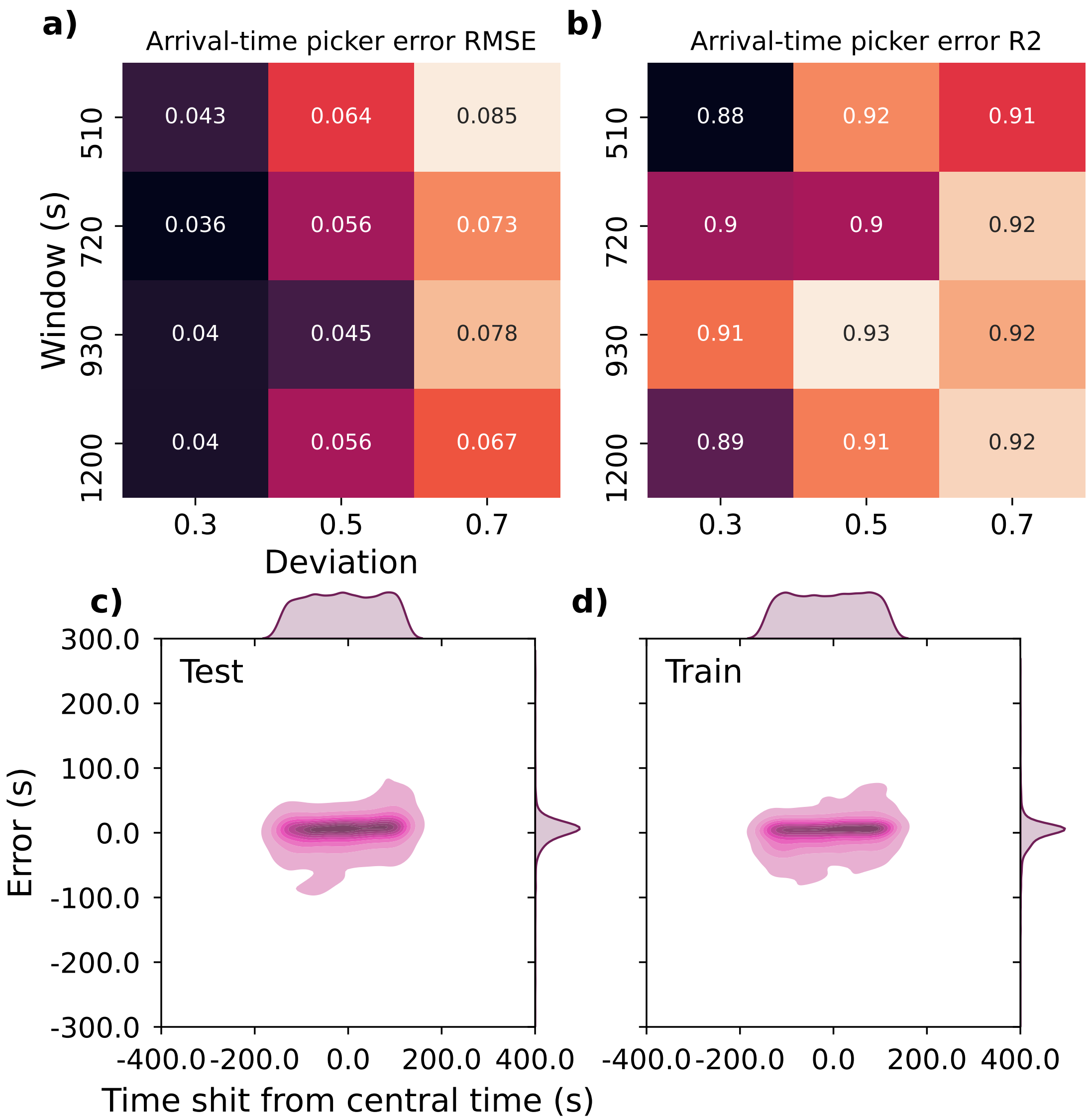

Figure S5: Performance of RF arrival time picker. (a) Root Mean Square Error (RMSE) vs minimum true-wavetrain overlap (deviation) and window size (s). The minimum true-wavetrain overlap corresponds to the minimum fraction of the wavetrain that has to be included in a window to be considered for training. (b) R2 error vs minimum true-wavetrain overlap (deviation) and window size (s). Bottom Distribution of arrival-time picking errors (s) vs true time shift from central time (s) over (c) the testing dataset, and (d) the training dataset.

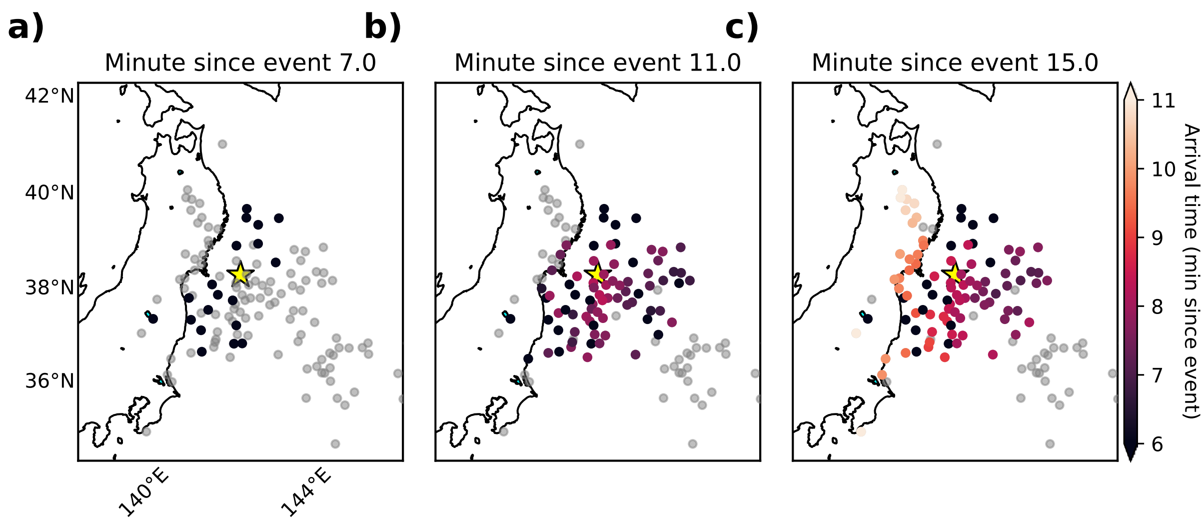

Figure S6: Ionospheric maps after the 2011 Tohoku earthquake generated at various times since the event. (a) to (c) Distribution of detected arrival times after (a) 7 minutes, (b) 11 minutes, and (c) 15 minutes since the event. CID coordinates were calculated at the intersection point between the LOS and the ionospheric layer using $H_{\text{ion}} = 250$ km. The colorcode corresponds to the predicted arrival time at each ionospheric point. Grey dots correspond to the location of ionospheric points where there is no detection yet but with detections after 20 mn.

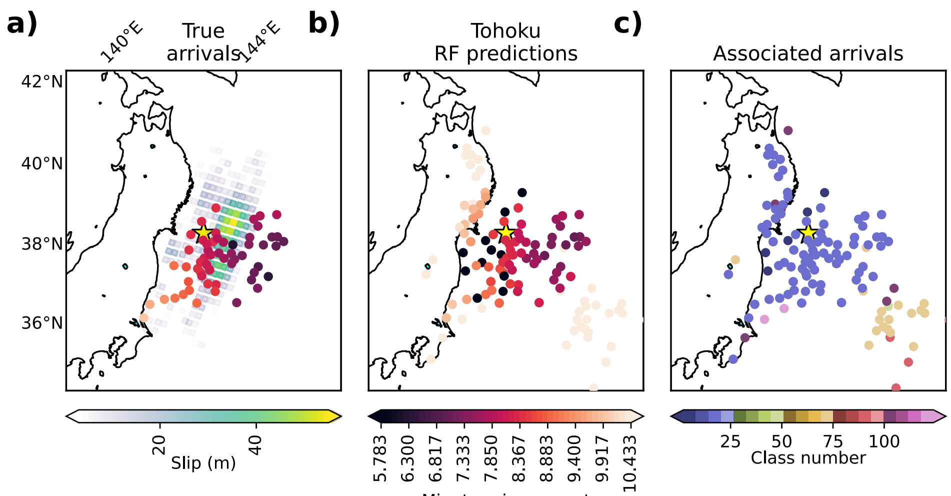

Figure S7: Tohoku's ionospheric arrival-time maps computed 14 minutes after the event for (d) hand-picked arrival times along with the epicenter location (yellow star), and surface projection of the fault slip (in m) as green to yellow patches, (e) RF-based arrival-time predictions, and (f) association classes determined from predicted arrival times.. CID coordinates were calculated at the intersection point between the LOS and the ionospheric layer using $H_{\text{ion}} = 180$ km.

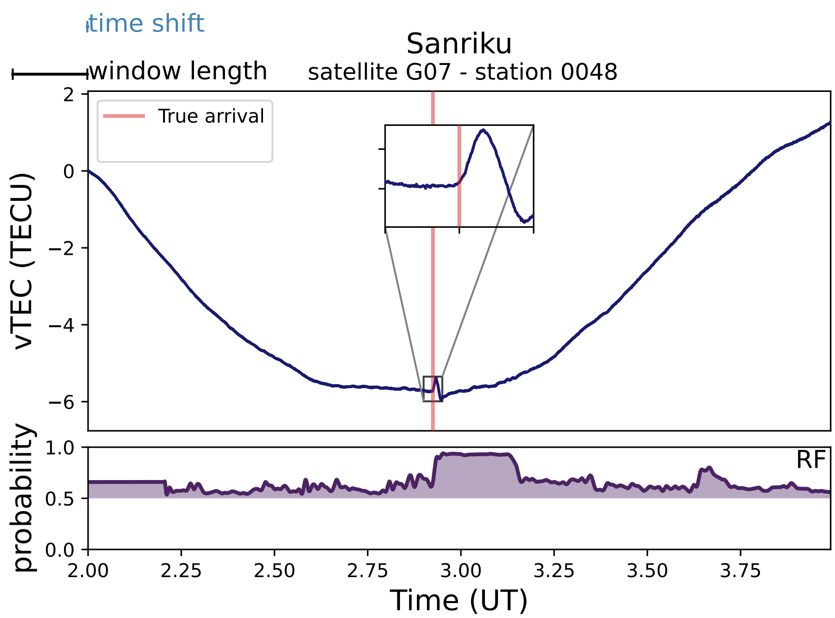

Figure S8: Performance assessment of RF detection and arrival-time picking at a higher sampling rate of 1s. 2-h vTEC waveform for the Sanriku event, satellite G07, station 0048 along with detection probabilities predicted by our RF detection model (bottom). The true arrival is shown as a red vertical line.

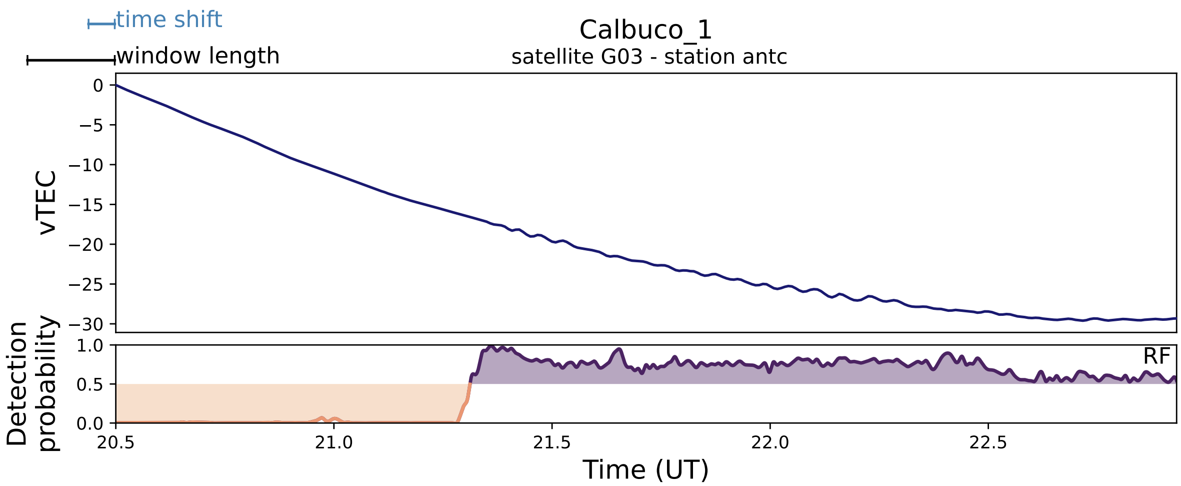

Figure S9: vTEC waveform for the Calbuco eruption, satellite G03, station antc along with detection probabilities predicted by our detection procedure (see Section 3) using a window size $w = 720$ s. Volcano-associated ionospheric perturbations are present between 21.3 and 22.5UT. The RF-predicted arrival time as a dark grey vertical line. The detected wavetrain using the RF is highlighted with a grey background.

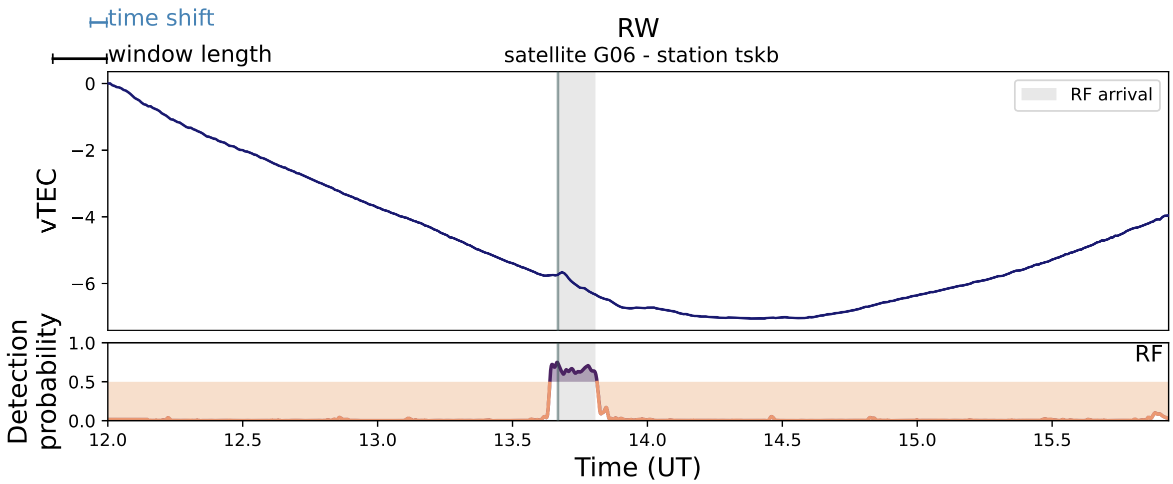

Figure S10: vTEC waveform from seismic Rayleigh waves recorded after the 1994 earthquake in Kuril Islands (Astafyeva et al., 2009), satellite G06, station tskb along with detection probabilities predicted by our detection procedure using a window size $w = 720$ s. Rayleigh-wave-associated ionospheric perturbations are present between 13.6UT and 13.8UT. The RF-predicted arrival time as a dark grey vertical line. The detected wavetrain using the RF is highlighted with a grey background.