AGU ADVANCING EARTH AND SPACE SCIENCE

Radio Science

RESEARCH ARTICLE
10.1029/2019RS006931

Key Points:
- Based on observed anomalies in ionospheric measurements, an earthquake precursor detection technique (EQ-PD) is proposed
- In EQ-PD, machine learning techniques are used to provide statistically reliable precursor detections
- Proposed EQ-PD provides results in near real time

Correspondence to:
A. A. Akyol,
akyol@ee.bilkent.edu.tr

Citation:
Akyol, A. A., Arikan, O., & Arikan, F. (2020). A machine learning-based detection of earthquake precursors using ionospheric data. *Radio Science*, *55*, e2019RS006931. https://doi.org/10.1029/2019RS006931

Received 27 JUL 2019
Accepted 20 OCT 2020
Accepted article online 17 NOV 2020

A Machine Learning-Based Detection of Earthquake Precursors Using Ionospheric Data

A. A. Akyol[1], O. Arikan[1], and F. Arikan[2]

[1]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey, [2]Department of Electrical and Electronics Engineering, Hacettepe University, Ankara, Turkey

Abstract Detection of precursors of strong earthquakes is a challenging research area. Recently, it has been shown that strong earthquakes affect electron distribution in the regional ionosphere with indirectly observable changes in the ionospheric delays of GPS signals. Especially, the total electron content (TEC) estimated from GPS data can be used in the seismic precursor detection for strong earthquakes. Although physical mechanisms are not well understood yet, GPS-based seismic precursors can be observed days prior to the occurrence of the earthquake. In this study, a novel machine learning-based technique, EQ-PD, is proposed for detection of earthquake precursors in near real time based on GPS-TEC data along with daily geomagnetic indices. The proposed EQ-PD technique utilizes support vector machine (SVM) classifier to decide whether an observed spatiotemporal anomaly is related to an earthquake precursor or not. The data fed to the classifier are composed of spatiotemporal variability map of a region. Performance of the EQ-PD technique is demonstrated in a case study over a region covering Italy in between the dates of 1 January 2014 and 30 September 2016. The data are partitioned into three nonoverlapping time periods, that are used for training, validation, and test of detecting precursors of earthquakes with magnitudes above 4 in Richter scale. The EQ-PD technique is able to detect precursors in 17 out of 21 earthquakes while generating 7 false alarms during the validation period of 266 days and 22 out of 24 earthquakes while generating 13 false alarms during the test period of 282 days.

1. Introduction

Earthquakes (EQs) are one of the oldest and most dangerous natural disasters. They can cause massive damage on buildings resulting in fatal injuries and loss of human lives. In spite of the significant advancements in the construction technologies of the buildings, losses due to strong EQs are still at an unacceptably high level. It is hard to prevent these losses due to uncertainty of EQ epicenter, magnitude, and time of onset. EQ prediction techniques are expected to estimate the time, the location, and the magnitude of seismic activity (FAQs, Can you predict earthquakes?, U.S. Geological Survey, n.d.). Unfortunately, an accurate EQ prediction technique has not been developed yet. Therefore, EQ prediction is one of the challenging research areas with the objective of reducing these uncertainties.

EQ prediction is also a highly controversial topic that there is an ongoing discussion regarding underlying EQ preparation mechanisms, possible EQ precursors and reliability of EQ prediction techniques (Wyss, 1997; Wyss et al., 1997). Since underlying seismic activities and seismic activity triggered natural phenomenon are still not fully understood, ongoing discussion has not been finalized yet. Additionally, most of the literature on EQ prediction depends on empirical case studies that aim to find precursory signals in the aftermath of a strong EQ in a retrospective manner. However, behavior of these precursory signals during long time intervals is not investigated. Hence, there are criticisms regarding repeatability and reliability of these techniques (Geller, 1997).

Proposed techniques in the literature that contribute to the EQ prediction can be classified into two main categories: model-based and precursor-based techniques. Model-based techniques hypothesize a mathematical or machine learning model for the unknown mechanism between past seismic activities and upcoming strong EQs. Some of the proposed mathematical models can be listed as fault line strain related force models that are investigated to suggest a periodicity in EQ appearances (Bendick & Bilham, 2017), Fibonacci, Lucas, Dual (FDL) numbers that are embedded in the occurrence times of old EQs to predict the upcoming EQ onsets (Boucouvalas et al., 2015), spatial connection model that fits to the EQ occurrence pattern around the

fault zones (Kannan, 2014), and an empirical probabilistic model that had been proposed to predict onset and magnitude of an upcoming EQ (Papazachos & Papaioannou, 1993). Several machine learning models are also implemented on past seismic activity data to predict EQ onset, magnitude, or epicenter such as decision trees, random forest, AdaBoost, information network, multiobjective info-fuzzy network, *k*-nearest neighbors, support vector machine (SVM), artificial neural networks (Asencio-Cortés et al., 2015, 2016; Last et al., 2016; Mahmoudi et al., 2016; Moustra et al., 2011) , support vector regressors and hybrid neural networks trained on an EQ catalog (Asim et al., 2018), deep neural networks (Panakkat & Adeli, 2009; Wang et al., 2017), and convolutional neural networks to predict an upcoming EQ with seismic waveforms of length of 100 s Ibrahim et al. (2018).

Precursor-based techniques rely on observation of changes in the EQ precursors such as the state of the ionosphere (Gulyaeva et al., 2017b; Karatay et al., 2010; Namgaladze et al., 2009; Oyama et al., 2016; Pulinets et al., 2004), radon gas emissions (Allegri et al., 1983; Hartmann & Levy, 2005), chemical composition of underground water (Asteriadis & Livieratos, 1989; Hartmann & Levy, 2005), temperature (Pulinets et al., 2006; Tronin et al., 2002, 2004), strange lights Fidani (2010), and unusual animal behavior (Grant & Halliday, 2010) prior to strong EQs.

Some studies suggest possible mechanisms between pre-EQ seismic activities and electromagnetic state of the ionosphere. Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) model proposes a possible coupling mechanism between Earth's lithosphere and ionosphere (Pulinets & Ouzounov, 2011). In the model, pre-EQ seismic activities that are triggered by tectonic plate movements in lithosphere generate radon gas emissions from Earth's crust. Radon gas emissions ionize air in atmosphere resulting with observable anomalies in *F*2 layer ionospheric parameters such as the total electron content (TEC), the maximum ionization height (*hmF*2), and the critical frequency of maximum ionization layer (*foF*2). TEC is defined as the line integral of electron density on a raypath, and it has proven itself as one of the main observables of ionosphere and plasmasphere (Arikan et al., 2004). In Pulinets and Davidenko (2018), another physical mechanism based on radon gas emissions is also proposed for ionospheric anomaly generation process that is associated with the atmosphere planetary boundary layer (PBL) behavior in local time. The mechanism proposes a fact that low height of PBL upper boundary induces the accumulation of radon released from the Earth's crust in the surface layer that results with sharp raises in the ionization level of ionosphere during nighttime.

There are detailed investigations reporting possible correlations between the observable anomalies in *F*2 layer ionospheric parameters and seismic activities. In Pulinets et al. (2002), a preliminary statistical analysis is performed on ionospheric *foF*2 critical frequencies measured at Chung-Li (Taiwan Island) ground-based station to reveal ionospheric precursors to EQs that had taken place in between the period from 1978 to 1986 with magnitudes greater or equal to 4. In this study, ionospheric data for the selected seismoactive region (Taiwan Island) which is obtained 6 days prior to the event with a magnitude of above 4 in the period from 1978 to 1986 are investigated. They revealed a valuable conclusion that deep-focus EQs ($60\,\text{km} < z < 300\,\text{km}$) in the chosen Taiwan region generate significant responses on the ionosphere. In another study (Liu et al., 2004), *foF*2 and TEC anomalies are observed 1 to 7 days to EQs which occurred in 86 days in the Taiwan area during the 3 year period of 1997–1999. In Liu et al. (2006), an empirical evidence of the pre-EQ ionospheric anomalies (PEIAs) that are reported by statistically investigating the relationship between variations of the plasma frequency at the ionospheric *F*2 peak *foF*2 and 184 EQs with magnitude $M \geq 5.0$ during 1994–1999 in the Taiwan area is presented. The study revealed the fact that multiple PEIAs appeared 1 to 5 days before the 184 $M \geq 5.0$ investigated EQs. Another comprehensive study (Le et al., 2011) is conducted on the pre-EQ ionospheric anomalies by using the TEC data from the global ionosphere map (GIM). A total of 736 $M \geq 6.0$ EQs in the global area during 2002–2010 is selected, and TEC anomaly days are defined. Then the occurrence rates of abnormal days for both the days within 1–21 days prior to the EQs (*PE*), and the background days (*PN*) are calculated. They revealed remarkable conclusions that rate of daily TEC anomalies are increasing for 10 days prior to strong EQs. They also provided the ratio of occurrence rates *PE*/*PN* which indicates that its 1.8 to 3.6 times more likely to observe TEC anomalies prior to strong EQs than observing these anomalies during background or quiet days. In another statistical study by Kon et al. (2011), pre-EQ ionospheric anomalies in time series are examined, and a statistical test by using TEC derived from GIMs around the Japan area is performed. The result of the statistical study indicates that GIM-TEC anomalies tend to appear 1 to 5 days before the investigated EQs. Since Kon et al. (2011), Le et al. (2011), Liu et al. (2004, 2006), and Pulinets et al. (2002) reported that statistical investigations on the anomaly

detection performances are limited to seismic activity periods, potential false anomaly detections or false alarms of the proposed techniques over the seismically inactive days remain uncertain.

There are also several other studies supporting these alternative mechanisms that ionospheric disturbances in ion temperatures, *foF*2, and TEC had been observed prior to strong EQs around Japan (Karatay et al., 2010; Ouzounov et al., 2011; Oyama et al., 2016; Zolotov et al., 2013), Chile (Namgaladze et al., 2009; Yadav et al., 2015), Italy (Akhoondzadeh, 2016; Davidenko & Pulinets, 2019; Kouris et al., 2006; Pulinets, 1998), El-Salvador (Plotkin, 1999), India (Trigunait et al., 2004), the United States (Kouris & Fotiadis, 2002; Pulinets et al., 2007), Mexico (Pulinets, 2004; Pulinets et al., 2004, 2005), Taiwan (Liu et al., 2000, 2004, 2010), Indonesia (Devi et al., 2013), Nepal (Li et al., 2016), China (Karatay et al., 2010; Zhang et al., 2004), and Turkey (Akyol, 2013; Akyol et al., 2013; Arikan et al., 2012). In these studies, statistical methods have been used for the identification of the influence of seismic activities on the ionospheric variations such as TEC difference and variation analysis (Arikan et al., 2012; Karatay et al., 2010; Namgaladze et al., 2009; Ouzounov et al., 2011; Plotkin, 1999; Yadav et al., 2015; Zolotov et al., 2013), equatorial TEC abnormality analysis (Devi et al., 2013), ionospheric correction Trigunait et al. (2004), correlation analysis between TEC and foF2 or different pairs of GPS receivers (Pulinets, 2004; Pulinets et al., 2004, 2005, 2007), generation of ionospheric-precursor masks based on variation of TEC and foF2 (Davidenko & Pulinets, 2019), interquartile range and percentage analysis (Liu et al., 2000, 2004, 2010, 2016; Oyama et al., 2016; Zhang et al., 2004), relative deviation of daily ionospheric parameters from their corresponding monthly medians (Kouris & Fotiadis, 2002; Kouris et al., 2006), and seismoionospheric variation investigation Pulinets (1998) and detection (Akhoondzadeh, 2016).

Considering the impact of statistically reliable real-time precursor detection of EQs, an objective investigation on the precursor detection and false alarm performances of a TEC-based EQ precursor detection technique had been performed in Akyol (2013) and Akyol et al. (2013). These investigations revealed correlation between ionospheric TEC anomalies and strong EQs by thresholding TEC-based anomaly detection signals. In this study, real-time seismic precursor detection performance of a machine learning-based ionospheric EQ precursor detection technique, EQ-PD, is presented. For EQ prediction, the proposed EQ-PD contributes to the literature in the following aspects.

1. EQ-PD performs detections of EQ precursors in the ionospheric data by using machine learning techniques.
2. Availability of ionospheric data over large areas and time durations enables reliable training, validation, and test of the EQ-PD over nonoverlapping time intervals. Hence, statistical evaluation of the EQ-PD can be performed reliably.
3. When daily available geomagnetic parameters and ionospheric TEC data are used, EQ-PD provides daily EQ precursor decisions.
4. When compared to other machine learning-based techniques, EQ-PD can be classified as a Boosting technique which combines multiple EQ precursor detections to obtain a final EQ precursor detection.

In this work, performance of the proposed EQ-PD technique is investigated over a region around Italy for a time duration of 2 years and 10 months. In this investigation, data obtained over this region are partitioned over three nonoverlapping time intervals for training, validation, and testing of the EQ-PD technique.

This paper is organized as follows. The proposed EQ-PD technique that consists of a sequence of processing phases is presented in detail in section 3. Precursor detection performance of the EQ-PD is presented in section 4. Conclusions and future directions of research are given in section 5.

2. Data Used by the Proposed EQ-PD Technique

Global Navigation Satellite System (GNSS) networks with precisely known receiver positions provide measurements on the ionospheric phase delays of satellite signals to improve GNSS-based position estimates (Davies & Hartmann, 1997). Reliable estimates for slant TEC (STEC), which is the total number of electrons in a cylinder of 1 m^2 cross section connecting the GPS satellite to the GPS receiver, can be produced by using provided phase delays. For each receiver position on the GNSS network, a vertical TEC (VTEC or TEC in short) estimate can be generated from the available STEC estimates at the receiver site, TEC is measured in units of TECU which is TECU = 10^{16} el/m^2 (Arikan et al., 2004).

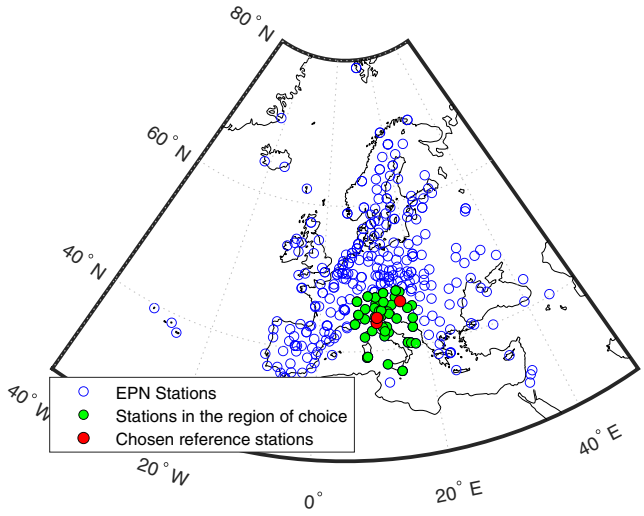

Figure 1. Positions of 320 EPN reference stations over Europe including region of choice covering Italy shown with green circles and chosen reference stations with red circles.

By using interpolation techniques, 2-D TEC maps can be produced from the TEC data obtained by using a network of spatially distributed GPS receivers (Sayin et al., 2008). These maps can be updated every 2 hr or as frequently as 2.5 min (Arikan et al., 2003; IONOLAB, Ionospheric Research Laboratory, n.d). Therefore, TEC is a reliable data source for constantly monitoring the state of the ionosphere and detection of anomalies. In this study, TEC data of EUREF Permanent GNSS Network (EPN) reference stations that are deployed in a seismically active region covering Italy are used (Bruyninx et al., 2012).

EPN is operated by the International Association of Geodesy (IAG) Regional Reference Frame sub-Commission for Europe (EUREF) with the collaboration of more than 100 European agencies and universities. As shown in Figure 1, The EPN consists of 320 Continuously Operating GNSS Reference Stations (CORS) located over European continent. To obtain reliable TEC data for 30 s of time separation between its samples, available measurements of each EPN reference station are processed by IONOLAB-TEC and IONOLAB-BIAS services.

IONOLAB-TEC is the state-of-art signal processing technique for GPS-TEC estimation for a single GPS station (Arikan et al., 2003, 2004, 2007; Nayir et al., 2007). IONOLAB-TEC provides accurate, reliable, and robust GPS-TEC estimation for any high-latitude, midlatitude, or equatorial GPS station for both quiet and distributed days with the same reliability and accuracy. IONOLAB-TEC combines data from all the GPS satellites that are above 100 elevation angle (horizon limit) of the GPS station with a temporal resolution of 30 s. The method calculates VTEC per satellite and combines them using a weighting function based on satellite positions which reduces the contamination caused by multipath effects (Arikan et al., 2008). The receiver differential code bias (DCB) is established using IONOLAB-BIAS method (Nayir et al., 2007). IONOLAB-TEC is available both as an online user-friendly service and as an executable file at http://www.ionolab.org (Sezen et al., 2013).

To illustrate, daily variability of TEC measurements in regions of seismic activity, reference stations with code names BOLG (44.5°N, 11.36°E) in Bologna, Italy, and IGMI (43.8°N, 11.21°E) in Firenze, Italy, is chosen. Since BOLG and IGMI reference stations are close to each other, it is possible to visualize interstation variability of daily TEC measurements. To illustrate, daily variability of TEC measurements in regions of negligible seismic activity, reference station with code name CAKO (46.39°N, 16.44°E) in Cakovec, Croatia, is chosen.

State of the ionosphere changes due to geomagnetic activity (Mandea & Korte, 2010). It is important to distinguish between extreme behavior in ionospheric state due to geomagnetic storms and seismic activity (Gulyaeva & Arikan, 2017; Gulyaeva et al., 2016, 2017a). Ionosphere-based parameters are widely used to assess the state of the ionosphere in terms of geomagnetic activities and geomagnetic storms. Geomagnetic activities of ionosphere can be investigated by monitoring A_p (planetary magnetic activity) index, *AE*

(auroral electrojet) index, and solar radiation related parameter sunspot number, *SSN*. Geomagnetic storms in ionosphere can also be investigated by monitoring K_p (global geomagnetic storm) index and *Dst* (disturbance storm time) index (Mandea & Korte, 2010; Rostoker, 1972; SILSO World Data Center, 2005). Daily A_p, *AE*, K_p, and *Dst* indices and *SSN* can be accessed via National Aeronautics and Space Administration (NASA) Goddard Space Flight Center OMNIWeb service (NASA OMNIWeb, SPDF Goddard Space Flight Center, n.d.). To illustrate daily variability of TEC, TEC data at reference station CAKO corresponding to two different dates of 6 and 8 January 2016 are shown in Figure 3.

EQ date, time, location, and magnitude data are obtained from the Advanced National Seismic System's comprehensive EQ catalog (ANSS ComCat) ANSS Comprehensive Catalog (ComCat), U.S. Geological Survey (n.d, ANSS Comprehensive Catalog (ComCat), U.S. Geological Survey (n.d).

3. Machine Learning-Based Earthquake Precursor Detection Technique: EQ-PD

Unlike sensors built to detect seismic activity during or after an EQ, GNSS networks can detect EQ precursors in ionospheric TEC, *foF*2, and *hmF*2 that are preseismic in nature. In this study, a machine learning-based ionospheric EQ precursor detection technique, EQ-PD, is proposed to detect EQ precursors using data provided by a GNSS network.

Due to inherent stochastic nature of the ionization in the atmosphere, TEC data show significant spatial and temporal variations. The detection of seismic activity related anomalies in the ionosphere requires detection of local variations beyond the expected range of variation in the TEC data.

The proposed EQ-PD technique of the ionospheric EQ precursors consists of five sequential stages. First, local TEC variations, (TEC_Δ), are estimated based on variation between the measured TEC at a GNSS station and its spatiotemporal estimate generated from the measurements of the neighboring GNSS stations. Second, by using a set of variation metrics, anomaly detection signals are generated as presented in section 3.2. In the third step, generated anomaly detection signals are thresholded for declaring anomaly decisions by using adaptively chosen spatial and spatiotemporal anomaly detection thresholds. These thresholds are generated with respect to past TEC variation behaviors of anomaly detection signals. Also, spatiotemporal anomaly detection thresholds depend on the geomagnetic states of the ionosphere, which are obtained by clustering geomagnetic indices and anomaly detection thresholds that are generated as described in section 3.3. Fourth, an EQ precursor detection signal is formed by combining all the anomaly decisions obtained from thresholded anomaly detection signals as discussed in section 3.4. Finally, as detailed in section 3.5, EQ precursor detections are generated by using SVM classifiers on the EQ precursor detection signal. Figure 2 shows the flow diagram of proposed EQ-PD technique.

3.1. Spatiotemporal TEC Interpolation

Let $\mathbf{x}_{u;d}$ denote available TEC data at reference u on day d:

$$\mathbf{x}_{u;d} = [x_{u;d}(1) \ \dots \ x_{u;d}(n) \ \dots \ x_{u;d}(N_{u;d})]^T, \tag{1}$$

where $\mathbf{x}_{u;d}$ is a $(N_{u;d} \times 1)$ vector, $x_{u;d}(n)$ is the nth TEC measurement, and $N_{u;d}$ is the total number of TEC measurements on day d. By denoting the number of neighboring reference stations that are located within R_r km radius of the reference station u as $N_{u;R_r}$, a spatially linear estimate for the TEC at reference station can be obtained as (Deviren et al., 2013) follows:

$$\hat{\mathbf{x}}_{u;d;R_r} = \sum_{v=1}^{N_{u;R_r}} \alpha_{u;d;R_r}(v)\mathbf{x}_{v;d;R_r}, \tag{2}$$

where $\alpha_{u;d;R_r}(v)$ is weight of the vth neighboring station and $\mathbf{x}_{v;d;R_r}$, a $(N_{u;d} \times 1)$ vector, represents the corresponding TEC measurements. The weights vector $\boldsymbol{\alpha}_{u;d;R_r}$ with size $(N_{u;R_r} \times 1)$ is chosen as the unique minimizer of the total estimation error over the range of days $[d_i, d_s]$:

$$\boldsymbol{\alpha}_{u;d;R_r} = \underset{\alpha_{u;d;R_r}(v)}{\operatorname{argmin}} \sum_{d_i}^{d_s} \left\| \mathbf{x}_{u;d_n} - \sum_{v=1}^{N_{u;R_r}} \alpha_{u;d;R_r}(v)\mathbf{x}_{v;d_n;R_r} \right\|_2^2, \tag{3}$$

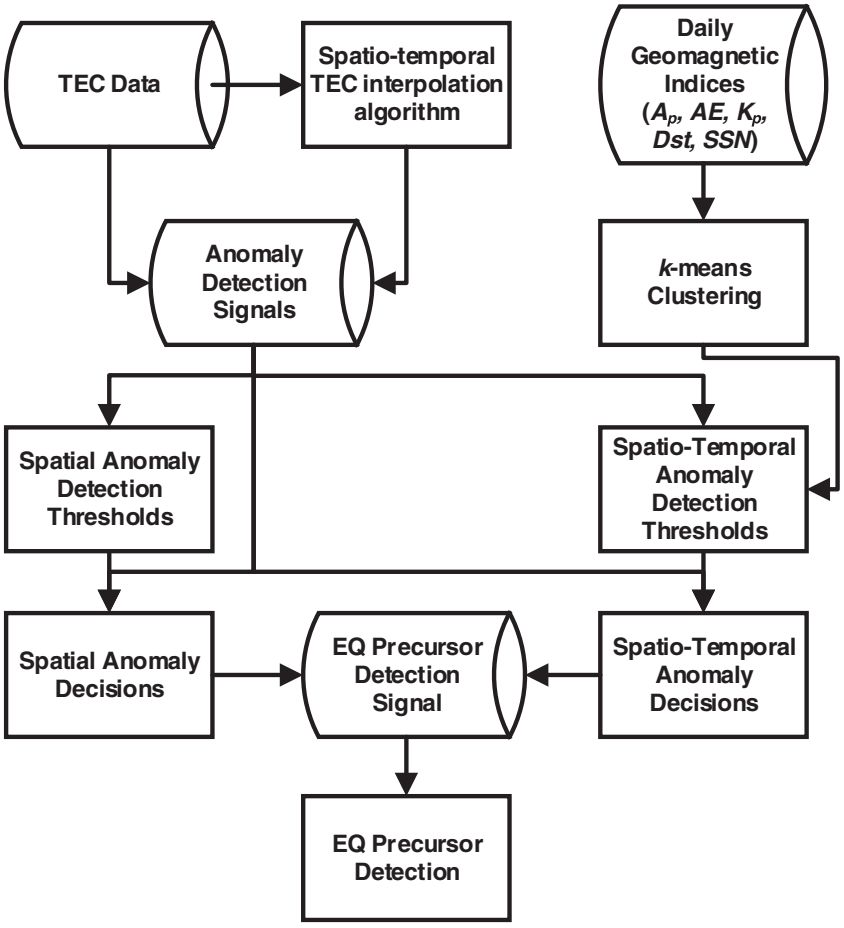

Figure 2. Flow diagram of the proposed EQ-PD technique. Detection signals are generated based on regional TEC data, and detection thresholds are adaptively chosen based on the geomagnetic indices.

which can be obtained in closed form as follows:

$$\boldsymbol{\alpha}_{u;d;R_r} = \left(\sum_{d_n=d_i}^{d_s} \mathbf{X}_{u;d_n;R_r}^T \mathbf{X}_{u;d_n;R_r} \right)^{-1} \left(\sum_{d_n=d_i}^{d_s} \mathbf{b}_{u;d_n;R_r} \right), \tag{4}$$

where $\mathbf{X}_{u;d_n;R_r}$ matrix $(N_{u;d} \times N_{u;R_r})$ is constructed by stacking TEC measurements of R_r km neighboring reference stations as

$$\mathbf{X}_{u;d_n;R_r} = [\mathbf{x}_{1;d_n;R_r} \ \cdots \ \mathbf{x}_{v;d_n;R_r} \ \cdots \ \mathbf{x}_{N_{u;R_r};d_n;R_r}], \tag{5}$$

and $\mathbf{b}_{u;d_n;R_r}$ vector $(N_{u;R_r} \times 1)$ is defined as follows:

$$\mathbf{b}_{u;d_n;R_r} = \mathbf{X}_{u;d_n;R_r}^T \mathbf{x}_{u;d_n}. \tag{6}$$

Since there exists a strong correlation between the magnetic activity of the Sun and the ionosphere, in this minimization, those days within $[d_i, d_s]$ are clustered with respect to their corresponding *SSN* and only those days that are in the same cluster are used (SILSO World Data Center, 2005).

By using the obtained weights $\boldsymbol{\alpha}_{u;d_n;R_r}$, $\hat{\mathbf{x}}_{u;d;R_r}$ can be computed and compared to $\mathbf{x}_{u;d}$ to detect the local TEC anomalies. Estimation performance of the spatiotemporal TEC interpolation technique on two different dates of 6 and 8 January 2016 at reference station CAKO is illustrated in Figure 3.

As shown in Figure 3, both TEC measurements and TEC estimates show high variability in daytime and relatively low variability at nighttime. Due to the strong correlation between strong solar radiation and TEC

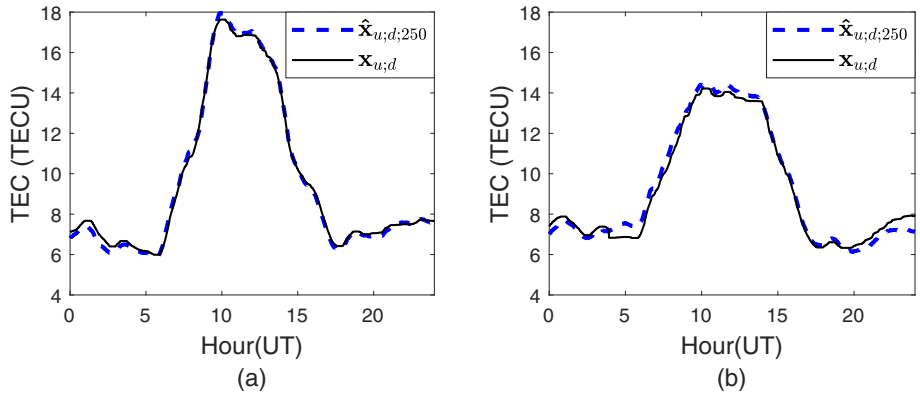

Figure 3. Daily TEC measurements $\mathbf{x}_{u;d}$ and their spatiotemporal estimates $\hat{\mathbf{x}}_{u;d;R_r}$ with $R_r = 250$ km at reference station CAKO on two different dates: (a) 6 January 2016 and (b) 8 January 2016.

variability of a reference station, u for the date d, $TECV_{u;d}$ can be partitioned with respect to variability sources and defined as follows:

$$TECV_{u;d} \simeq TECV_{sol} + TECV_{seis} + TECV_{oth}, \tag{7}$$

where $TECV_{sol}$ is variation in TEC triggered by strong solar radiation, $TECV_{seis}$ is variation in TEC triggered due to possible seismic activities during EQ preparation process, and $TECV_{oth}$ is variations on TEC due to other reasons. Since during the daytime $TECV_{sol}$ reaches to higher values and increases dynamic range, it dominates $TECV_{seis}$ and creates significant challenges in detection of $TECV_{seis}$. Thus, nighttime TEC variations with relatively weaker $TECV_{sol}$ components may serve as a more reliable detection signal for the presence of $TECV_{seis}$ components. Effect of strong solar radiation on TEC measurements can be reduced by applying a simple TEC measurement window as presented in Figure 4a. Proposed TEC measurement window suppresses the effect of solar radiation on a TEC measurement from 5 a.m. to 7 p.m. UT which corresponds to the night hours in the chosen region around Italy. Figure 4 illustrates two nighttime TEC measurements, $\mathbf{x}^{\text{night}}_{u;d}$ obtained with application of TEC measurement window on the TEC measurements presented in Figure 3. Spatiotemporal TEC estimates of the nighttime TEC measurements, $\hat{\mathbf{x}}^{\text{night}}_{u;d;R_r}$, are computed with spatiotemporal TEC interpolation technique and also presented in Figure 4.

3.2. Generation of Whole-Day and Nighttime Anomaly Detection Signals

The symmetric Kullback-Leibler divergence (SKLD) serves as a robust metric to measure the variation between the measurements available at a reference station u and its spatially interpolated estimate from the neighboring stations in the region (Arikan et al., 2016; Karatay et al., 2009; Necat Deviren et al., 2014; Turel & Arikan, 2010):

$$TEC^{\text{w-day}}_{\Delta} = SKLD(\mathbf{P}_{u;d}; \hat{\mathbf{P}}_{u;d;R_r}) = KLD(\hat{\mathbf{P}}_{u;d;R_r}|\mathbf{P}_{u;d}) + KLD(\mathbf{P}_{u;d}|\hat{\mathbf{P}}_{u;d;R_r}), \tag{8}$$

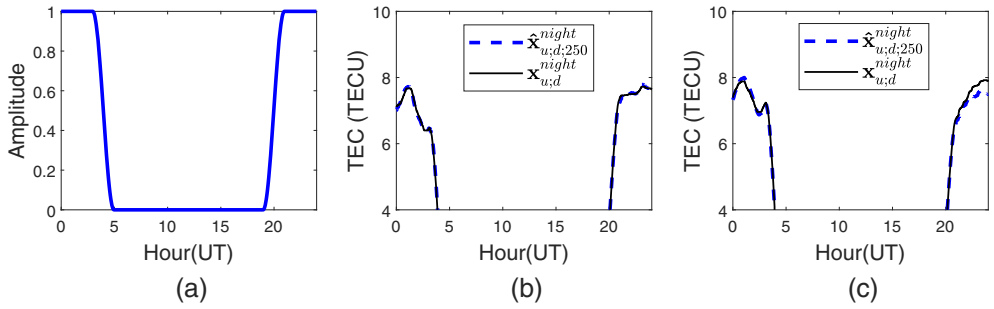

Figure 4. (a) TEC measurement window for generating the nighttime TEC measurements. Daily nighttime TEC measurements $\mathbf{x}^{\text{night}}_{u;d}$ and their spatiotemporal estimates $\hat{\mathbf{x}}^{\text{night}}_{u;d;R_r}$ with $R_r = 250$ km at reference station CAKO on two different dates: (a) 6 January 2016 and (b) 8 January 2016.

where $TEC_{\Delta}^{\text{w-day}}$ is the whole-day TEC variation and KLD is the Kullback-Leibler divergence given by

$$KLD(\hat{\mathbf{P}}_{u;d;R_r}|\mathbf{P}_{u;d}) = \left(\sum_{n=1}^{N_{u;d}} \hat{P}_{u;d;R_r}(n) \log\left(\frac{\hat{P}_{u;d;R_r}(n)}{P_{u;d}(n)}\right)\right), \tag{9}$$

$$KLD(\mathbf{P}_{u;d}|\hat{\mathbf{P}}_{u;d;R_r}) = \left(\sum_{n=1}^{N_{u;d}} P_{u;d}(n) \log\left(\frac{P_{u;d}(n)}{\hat{P}_{u;d;R_r}(n)}\right)\right), \tag{10}$$

where $\mathbf{P}_{u;d}$ and $\hat{\mathbf{P}}_{u;d;R_r}$ are the normalized $\mathbf{x}_{u;d}$ and $\hat{\mathbf{x}}_{u;d;R_r}$ as

$$\mathbf{P}_{u;d} = \frac{\mathbf{x}_{u;d}}{\|\mathbf{x}_{u;d}\|_1} = \mathbf{x}_{u;d}\left(\sum_{n=1}^{N_{u;d}} x_{u;d}(n)\right)^{-1}, \tag{11}$$

$$\hat{\mathbf{P}}_{u;d;R_r} = \frac{\hat{\mathbf{x}}_{u;d;R_r}}{\|\hat{\mathbf{x}}_{u;d;R_r}\|_1} = \hat{\mathbf{x}}_{u;d;R_r}\left(\sum_{n=1}^{N_{u;d}} \hat{x}_{u;d;R_r}(n)\right)^{-1}. \tag{12}$$

Due to the normalizations in (11) and (12), the obtained $\mathbf{P}_{u;d}$ and $\hat{\mathbf{P}}_{u;d;R_r}$ vectors contain unitless variables. Similarly, nighttime local TEC variation, $TEC_{\Delta}^{\text{night}}$, can be obtained by calculating the SKLD between $\mathbf{x}_{u;d}^{\text{night}}$ and its estimate $\hat{\mathbf{x}}_{u;d;R_r}^{\text{night}}$ as detailed below:

$$TEC_{\Delta}^{\text{night}} = SKLD(\mathbf{P}_{u;d}^{\text{night}}; \hat{\mathbf{P}}_{u;d;R_r}^{\text{night}}) = KLD(\hat{\mathbf{P}}_{u;d;R_r}^{\text{night}}|\mathbf{P}_{u;d}^{\text{night}}) + KLD(\mathbf{P}_{u;d}^{\text{night}}|\hat{\mathbf{P}}_{u;d;R_r}^{\text{night}}), \tag{13}$$

where $\mathbf{P}_{u;d}^{\text{night}}$ and $\hat{\mathbf{P}}_{u;d;R_r}^{\text{night}}$ are the normalized $\mathbf{x}_{u;d}^{\text{night}}$ and $\hat{\mathbf{x}}_{u;d;R_r}^{\text{night}}$ as

$$\mathbf{P}_{u;d}^{\text{night}} = \frac{\mathbf{x}_{u;d}^{\text{night}}}{\|\mathbf{x}_{u;d}^{\text{night}}\|_1} = \mathbf{x}_{u;d}^{\text{night}}\left(\sum_{n=1}^{N_{u;d}} x_{u;d}^{\text{night}}(n)\right)^{-1}, \tag{14}$$

$$\hat{\mathbf{P}}_{u;d;R_r}^{\text{night}} = \frac{\hat{\mathbf{x}}_{u;d;R_r}^{\text{night}}}{\|\hat{\mathbf{x}}_{u;d;R_r}^{\text{night}}\|_1} = \hat{\mathbf{x}}_{u;d;R_r}^{\text{night}}\left(\sum_{n=1}^{N_{u;d}} \hat{x}_{u;d;R_r}^{\text{night}}(n)\right)^{-1}. \tag{15}$$

Detection of anomalous behaviors in $TEC_{\Delta}^{\text{w-day}}$ and $TEC_{\Delta}^{\text{night}}$ can be performed on whole-day and nighttime anomaly detection signals which can be constructed by 2-D interpolation of all calculated $TEC_{\Delta}^{\text{w-day}}$ and $TEC_{\Delta}^{\text{night}}$ for every EPN reference station in the region of choice (Stanislawska et al., 2002; Stein, 1999). Figure 5 illustrates generated whole-day anomaly detection signals at two different dates: 6 and 8 January 2016. As seen in Figure 5, the generated whole-day anomaly detection signals vary in both space and time.

3.3. Adaptive Generation of the Anomaly Detection Thresholds

Detection of beyond the expected anomalies on the generated anomaly detection signals in near real time requires retrospective analysis on the past anomaly detection signals. Hence, days of past anomaly detection signals will be divided into two nonoverlapping time intervals named as training set and validation set. Present and future anomaly detection signal days will be reserved as test set and will not contribute to the retrospective analysis of anomaly detection signals. Near-real-time detection of the anomalies in the observed $TEC_{\Delta}^{\text{w-day}}$ and $TEC_{\Delta}^{\text{night}}$ defined in (8) and (13) can be accomplished by using thresholds that are determined based on retrospective analysis. Retrospective analysis of anomaly detection signals consists of two sequential stages. First, statistical behaviors of anomaly detection signals will be determined during the training set of dates. Thereafter, anomaly detection thresholds will be generated based on the obtained statistical behaviors in the previous stage.

Since TEC varies only in time and space, detection thresholds can be generated in three different ways: constant in time but varying in space (spatial thresholds) or constant in space and varying in time (temporal thresholds) or varying in both space and time (spatiotemporal thresholds). In this study, spatial and spatiotemporal anomaly detection thresholds are generated with respect to whole-day and nighttime anomaly detection signals. Training days of anomaly detection signals are divided into two classes: days with seismic activity and days with no-seismic activity. Anomaly detection thresholds are generated with the retrospective analysis based on the statistics in the days with no-seismic activity in the training set.

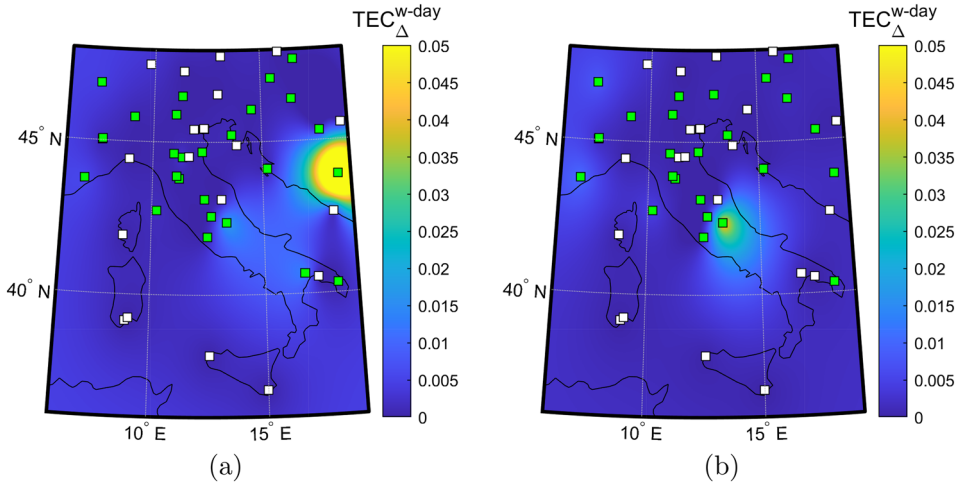

Figure 5. Whole-day anomaly detection signals at two different dates: (a) 6 January 2016 and (b) 8 January 2016. Available stations are marked with green squares, and unavailable stations are marked with white squares.

3.3.1. Spatial Anomaly Detection Thresholds

Statistical behaviors of anomaly detection signals on the training set of dates can be modeled by estimating a negative Pareto cumulative distribution for the local TEC variations, TEC_Δ at each EPN reference station in the region of choice on the days with no seismic activity (Aban et al., 2006).

Since the TEC measurement and their generated estimates are positive valued, upper truncated Pareto distributions provide appropriate statistical characterization on dense TEC measurements. For an upper-truncated Pareto random variable Z its tail probability is given by the following three parameter model in ρ, γ, and ϑ (Aban et al., 2006; Zaninetti & Ferraro, 2008):

$$P(Z > z) = \frac{\gamma^\rho (z^{-\rho} - \vartheta^{-\rho})}{1 - \left(\frac{\gamma}{\vartheta}\right)^\rho}\ ,\ 0 < \gamma \le z \le \vartheta < \infty\ . \tag{16}$$

Maximum likelihood (ML) estimates for the distribution parameters are as follows:

$$\hat{\gamma}\ =\ \min\left(TEC_{\Delta;u}(1)\ ,\ \ldots\ ,\ TEC_{\Delta;u}(N_{ns})\right)\ , \tag{17}$$

$$\hat{\vartheta}\ =\ \max\left(TEC_{\Delta;u}(1)\ ,\ \ldots\ ,\ TEC_{\Delta;u}(N_{ns})\right)\ , \tag{18}$$

and $\hat{\rho}$ is obtained as the solution to

$$\frac{N_{ns}}{\hat{\rho}} + \frac{N_{ns}\left(\frac{\hat{\gamma}}{\hat{\vartheta}}\right)^{\hat{\rho}} \log\left(\frac{\hat{\gamma}}{\hat{\vartheta}}\right)}{1 - N_{ns}\left(\frac{\hat{\gamma}}{\hat{\vartheta}}\right)^{\hat{\rho}}}\ =\ \sum_{i=1}^{N_{ns}} \left[\log(TEC_{\Delta;u}(i)) - \log(\hat{\gamma})\right]\ , \tag{19}$$

where N_{ns} is number of days with no seismic activity and $TEC_{\Delta;u}(i)$, $0 \le i \le n$ is the time series of the local TEC variations of reference station u (Aban et al., 2006; Zaninetti & Ferraro, 2008). Once these parameters are obtained, the conditional probability of a TEC_Δ that is observed in any given EPN reference station in the region can be estimated for seismically inactive days in between 2005 and 2016. Figure 6a illustrates the estimated upper-truncated Pareto negative cumulative distributions of $TEC_\Delta^{\text{w-day}}$ by using the seismically inactive days for the two EPN reference stations BOLG and IGMI in Italy. As shown in Figure 6a, estimated cumulative distributions show spatial variation that a given $TEC_\Delta^{\text{w-day}}$ tail probability of 0.2 corresponds to threshold levels of 0.001 and 0.004 at Reference Stations IGMI and BOLG, respectively. Similarly, it is also possible to estimate Pareto distributions of nighttime local TEC variations, $TEC_\Delta^{\text{night}}$. Figure 6b illustrates the estimated upper-truncated Pareto negative cumulative distributions of $TEC_\Delta^{\text{w-day}}$ and $TEC_\Delta^{\text{night}}$ by using the seismically inactive days for BOLG reference station in Italy. As shown in Figure 6b, estimated cumulative distributions differs from each other for nighttime and whole day local TEC variation data.

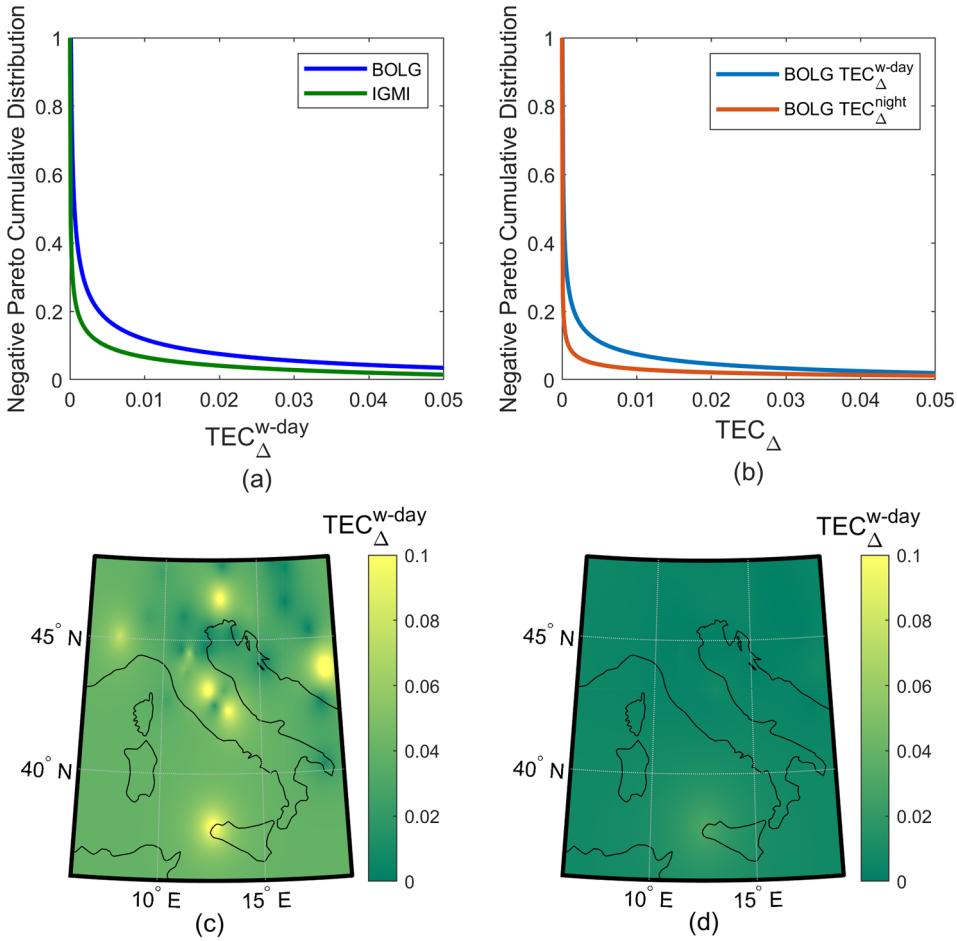

Figure 6. (a) Estimated upper-truncated Pareto negative cumulative distributions for the days in between 2005 and 2016 for two EPN Reference Stations BOLG and IGMI. (b) Estimated upper-truncated Pareto negative cumulative distributions of $TEC_\Delta^{w\text{-}day}$ and TEC_Δ^{night} for BOLG reference station. Two spatial anomaly detection thresholds with their chosen $TEC_\Delta^{w\text{-}day}$ tail probabilities (c) 0.01 and (d) 0.1.

Spatial anomaly detection thresholds can be obtained by choosing an appropriate TEC_Δ tail probability and finding the corresponding TEC_Δ on the estimated upper-truncated Pareto negative cumulative distribution for each of the EPN reference station in the region. Figure 6 illustrates two whole-day spatial anomaly detection thresholds generated by choosing $TEC_\Delta^{w\text{-}day}$ tail probabilities of 0.01 and 0.1, respectively. As shown in Figure 6, spatial anomaly detection thresholds shows significant variation in space and choosing a smaller TEC_Δ tail probability results with higher TEC_Δ threshold for each reference station.

Note that the generated spatial anomaly detection thresholds adapt to the anomaly detection signals spatially for the days with no seismic activity. Generated spatial anomaly detection thresholds estimate expected TEC_Δ statistically for each station individually with respect to their quiet time TEC_Δ distributions. Hence, it is possible to generate these thresholds in a spatially adaptive way.

3.3.2. Spatiotemporal Anomaly Detection Thresholds

By using training set of dates, spatiotemporal anomaly detection thresholds can be generated by clustering each day based on its ionospheric variability and obtaining a spatial anomaly detection threshold for each cluster of dates and for each EPN reference station in the region.

k-means clustering is an unsupervised machine learning technique that does not require labeling on the provided data. By using *k*-means clustering technique on the training data that consists of average of daily geomagnetic parameters (A_p, AE, K_p, and Dst indices and SSN), daily state of the ionosphere can be classified into *k* different geomagnetic activity clusters. To demonstrate *k*-means clustering technique implementation on geomagnetic activity classification of ionosphere, daily geomagnetic parameters are archived for days in

Table 1
Three Different Cluster Means for the Days in Between 2005 and 2016 and Mean Value of Each Geomagnetic Parameter

k-means cluster	K_p	SSN	*Dst* (nT)	A_p (nT)	*AE* (nT)
1	3.19481	51.8314	−25.8080	21.7639	350.6083
2	1.12953	20.9920	−5.4621	4.8475	93.5610
3	1.40000	112.5771	−5.8790	6.2715	127.5974
Mean	1.58834	50.9175	−9.3896	8.3969	150.7634

between 2005 and 2016. Archived data normalized with respect to its minimum and the maximum values of the geomagnetic indices. k-means clustering technique is implemented for different k values ranging between 1 to 20. Among the clustering results, k value of the clustering result that has the highest mean silhouette score is chosen (Kaufman & Rousseeuw, 2009; Rousseeuw, 1987). Table 1 presents the cluster means generated for $k = 3$ and the mean value of each geomagnetic parameter. Note that some of the chosen daily geomagnetic parameters such as *Dst* and *AE* indices are provisional or not final. Therefore, these indices may contain artificial noise and baseline shifts (Mandea & Korte, 2010). However, effect of these variations on the generated cluster means is insignificant when the archived data duration is in the order of years and it is safe to implement k-means clustering technique. Additionally, EQ-PD technique utilizes SVM that performs robust detection of EQ precursors even on those days which are misclustered due to a possible error on the provisional indices.

As exemplified in Table 1, each cluster represents a different state of the ionosphere based on its geomagnetic activity. When the geomagnetic activities represented by these clusters are compared with each other, Cluster-1 represents days with high geomagnetic activity, Cluster-3 represents days with moderate geomagnetic activity, and Cluster-2 represents days with low geomagnetic activity in the ionosphere. It is possible to classify a day of interest with respect to the smallest distance between its geomagnetic parameters and the generated cluster means.

Spatiotemporal anomaly detection thresholds can be obtained by estimating k different upper-truncated Pareto negative cumulative distributions representing the k different geomagnetic activity states of ionosphere for seismically inactive days and choosing an appropriate TEC_Δ tail probability for finding the corresponding TEC_Δ threshold on the estimated negative cumulative distributions of each EPN reference station in the region of choice. Figure 7a illustrates the estimated $k = 3$ different upper-truncated Pareto negative cumulative distributions of $TEC_\Delta^{\text{w-day}}$ by using seismically inactive days in between 2005 and 2016 and the cluster means presented in Table 1 for the BOLG reference station, Figure 7b illustrates the EQs that have taken place in the region of choice with magnitudes greater than 4 in Richter scale for the same time period. Figure 7 also presents a spatiotemporal anomaly detection threshold obtained from the set of days that belongs to Cluster-1 and Cluster-3 in Table 1 by choosing TEC_Δ tail probability as 0.01. As presented in Figure 7, EQ epicenters and generated spatiotemporal anomaly detection thresholds are highly correlated.

Note that the generated spatiotemporal anomaly detection thresholds adapt to the anomaly detection signals both spatially and temporally for the days with no seismic activity. Generated spatial anomaly detection thresholds estimate expected TEC_Δ statistically for each station and for each day with respect to the multiple quiet time TEC_Δ distributions that are generated for each station, individually. Hence, it is possible to generate these thresholds in both spatially and temporally adaptive way.

3.4. Generation of EQ Precursor Detection Signal

Proposed EQ-PD technique built on the hypothesis that presence of strong seismic activities results in detectable anomalies in TEC_Δ. Therefore, statistical behavior of TEC_Δ anomalies during the no-seismic activity days is investigated to choose proper thresholds on TEC_Δ and investigate their detection performance on the test data. Scope of this investigation can be expanded by increasing the diversity of the generated anomaly detection thresholds.

There are several studies that report significant TEC disturbances 9 days prior to EQs (Arikan et al., 2016; Karatay et al., 2009, 2010; Necat Deviren et al., 2014). A comprehensive study of Le et al. (2011) revealed several facts that TEC disturbances are mostly concentrated on 1 to 9 days prior to EQs and it's 1.8 to 3.6

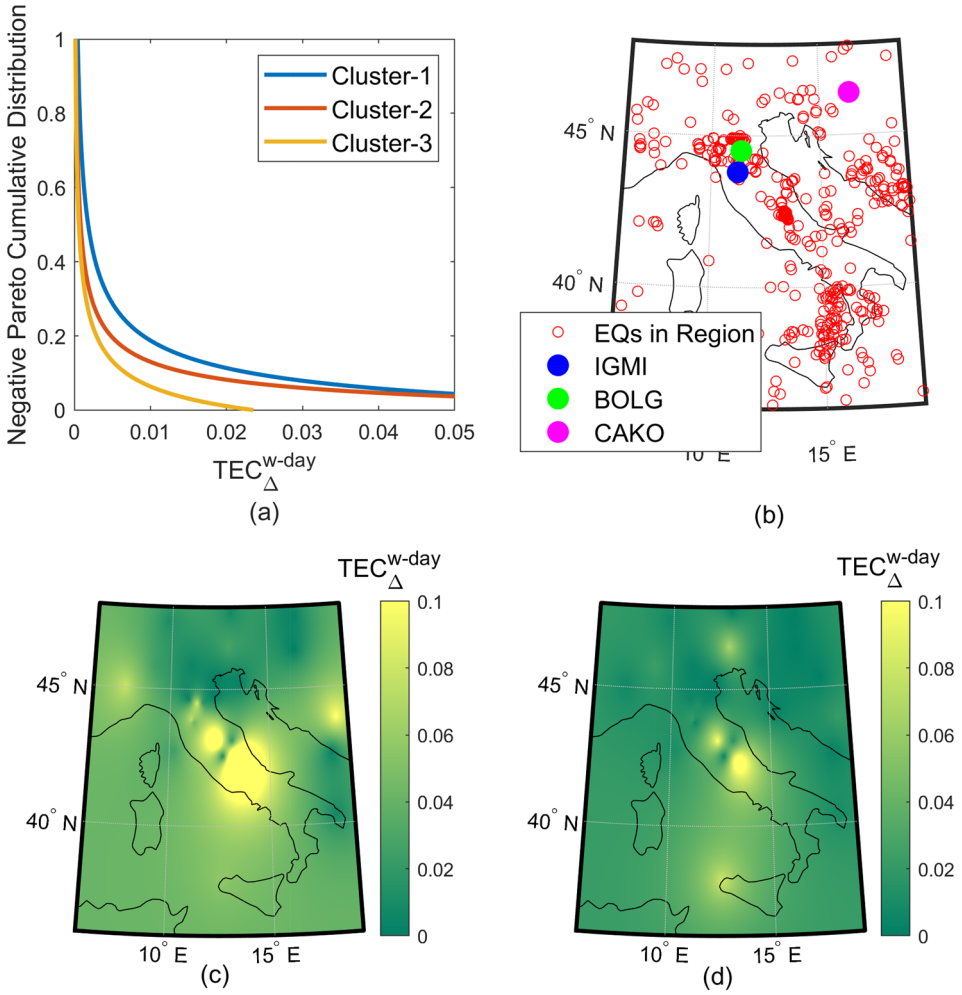

Figure 7. (a) Estimated upper-truncated Pareto negative cumulative distributions for the clustered days in between 2005 and 2016 for BOLG reference station. (b) Locations of IGMI, BOLG, and CAKO reference stations that are used for data visualization with EQs that have taken place in the region of choice during days in between 2005 and 2016. A spatiotemporal anomaly detection threshold is obtained for (c) Cluster-1 days and (d) Cluster-3 days with chosen $TEC_\Delta^{w\text{-}day}$ tail probability of 0.01.

times more likely to observe TEC anomalies 1 to 9 days prior to strong EQs than observing these anomalies during seismically inactive or quiet days. Since significant TEC disturbances are observed 9 days prior to an EQ, these days including with the EQ day are selected for the seismic activity days class. All the remaining days are included to the no-seismic activity days class.

For a given TEC_Δ tail probability, an anomaly detection threshold can be generated by finding the corresponding TEC_Δ on the estimated Pareto distributions for each reference station of interest. Since the estimated Pareto distributions are obtained from no-seismic activity days, they depict the TEC_Δ behavior of each reference station for the days with no-seismic activity. Therefore, each different TEC_Δ tail probability corresponds to a different probability of false alarm (P_{FA}) of TEC_Δ anomaly detections for the no-seismic activity days class. In this study, three different whole-day and three different nighttime anomaly detection signals are generated by changing the R_r estimation radius of spatiotemporal TEC interpolation technique as detailed in sections 3.1 and 3.2. Spatial and spatiotemporal anomaly detection thresholds are generated based on the statistics obtained from these six different anomaly detection signals for 23 different TEC_Δ tail probabilities for each anomaly detection threshold. Therefore, it is possible to generate $6 \times 2 \times 23$ statistically different anomaly decisions for a given day. Anomaly decisions are marked as 1's, if there is a threshold exceedance and 0's otherwise. All 276 different daily anomaly decisions are stacked to form the daily EQ precursor detection signal.

3.5. Generation of EQ Precursor Detections

SVM is one of the widely used machine learning techniques for binary classification (Burges, 1998; Platt, 1998). In this study, SVM with Gaussian kernel is chosen for the classification of EQ precursor detection signal into the classes of days with and without seismic activity. Class labels for the days with seismic activity and no-seismic activity are chosen as $y = 1$ and $y = 0$, respectively. Daily EQ precursor detections can be generated by obtaining decision boundary coefficients, $\boldsymbol{\theta}$ that minimize SVM cost function in (20). Main objective of the SVM is to minimize the following cost function subject to the decision boundary coefficients, $\boldsymbol{\theta}$ on the n different anomaly decisions of EQ precursor detection signal with m days of duration:

$$\min_{\theta} C_{reg} \sum_{d=1}^{m} y^{(d)} C_{1;h}(\boldsymbol{\theta}^T \mathbf{s}^{(d)}) + (1 - y^{(d)}) C_{0;h}(\boldsymbol{\theta}^T \mathbf{s}^{(d)}) + \frac{1}{2} \sum_{j=1}^{m} \theta_j^2, \tag{20}$$

where C_{reg} is the regularization constant and $C_{0;h}$ and $C_{1;h}$ are the hinge loss functions for the cost of classifying when $y = 0$ and $y = 1$, respectively, and defined as follows:

$$C_{0;h}(\boldsymbol{\theta}^T \mathbf{s}^{(d)}) = \max(0, (1 + \boldsymbol{\theta}^T \mathbf{s}^{(d)})), \tag{21}$$

$$C_{1;h}(\boldsymbol{\theta}^T \mathbf{s}^{(d)}) = \max(0, (1 - \boldsymbol{\theta}^T \mathbf{s}^{(d)})), \tag{22}$$

$\mathbf{s}^{(d)}$ is the dth day sample with m features obtained by using a Gaussian kernel with parameter σ on the EQ precursor detection signal as

$$\mathbf{s}^{(d)} = [\mathbf{s}_1^{(d)} \ \dots \ \mathbf{s}_r^{(d)} \ \dots \ \mathbf{s}_m^{(d)}]^T, \tag{23}$$

$$\mathbf{s}_r^{(d)} = \exp\left(-\frac{\|\mathbf{P}_E^{(d)} - \mathbf{P}_E^{(r)}\|^2}{2\sigma^2} \right) = \exp\left(-\frac{\sum_{j=1}^{n} (P_{E;j}^{(d)} - P_{E;j}^{(r)})^2}{2\sigma^2} \right), \tag{24}$$

where $P_{E;j}^{(d)}$ is the jth anomaly decision of the EQ precursor detection signal for the day d.

Before the training of SVM model, each sample $\mathbf{s}^{(d)}$ is multiplied by a weight, $w^{(d)}$, (initially $w_{init}^{(d)} = 1$ for dth sample) and weights are normalized to sum up to the value of the prior probability in the respective class, $p_D(0)$ for no seismic activity class with m_0 samples (Class 0) and $p_D(1)$ for seismic activity class with m_1 samples (Class 1):

$$\mathbf{p}_D = \begin{bmatrix} p_D(0) \\ p_D(1) \end{bmatrix} = \frac{1}{m_0 + m_1} \begin{bmatrix} m_0 \\ m_1 \end{bmatrix}. \tag{25}$$

Before the training of a cost-sensitive SVM model, prior probability vector $\mathbf{p}_D$ can be updated by multiplication with a cost matrix which is formed by cost of generating a false alarm $Cost_{FA}$ and cost of generating a misdetection $Cost_{MD}$ as in

$$\mathbf{p}_{DC} = \begin{bmatrix} 0 & Cost_{FA} \\ Cost_{MD} & 0 \end{bmatrix} \mathbf{p}_D. \tag{26}$$

Updated prior probability vector, $\mathbf{p}_{DC}$, can be used to generate dth sample weight, $w^{(d)}$, for $d \in$ Class 0 as in

$$w^{(d)} = \frac{w_{init}^{(d)}}{\sum_{\forall d \in Class\,0} w_{init}^{(d)}} p_{DC}(0), \tag{27}$$

and likewise $d \in$ Class 1 as in

$$w^{(d)} = \frac{w_{init}^{(d)}}{\sum_{\forall d \in Class\,1} w_{init}^{(d)}} p_{DC}(1). \tag{28}$$

Table 2
Training, Validation, and Test Set Information

	Date (D Month YYYY)		Number	Number
Data set	Start	End	of days	of EQs
Training	1 January 2014	1 April 2015	456	35
Validation	2 April 2015	23 December 2015	266	21
Test	24 December 2015	30 September 2016	282	24

Therefore, samples contribute to the solution of objective function in (20) with respect to their class and cost-sensitive learning will be performed. For $Cost_{MD}$ is set as 1, choosing $Cost_{FA} > 1$ will force the SVM classifier to reduce the number of generated false alarms.

SVM is a parametric model that C_{reg}, σ and $Cost_{FA}$ parameters affect the precursor detection performance of the technique. Therefore, optimal $\boldsymbol{\theta}$ must be obtained for different combinations of model parameters with training, for a chosen combination of model parameters, performance of these $\boldsymbol{\theta}$ must be validated with validation and $\boldsymbol{\theta}$ which obtained for the chosen model parameters should be tested.

In the following performance of the proposed EQ-PD technique will be demonstrated by applying cost-sensitive regularized SVM with a Gaussian kernel on the EQ precursor detection signals for the days in 2014 and 2016.

4. Performance of the Proposed EQ-PD Technique

Near-real-time implementation performance of the proposed TEC-based EQ-PD technique is evaluated by training, validation, and test sets that consist of distinct EQ precursor detection signals. The start and the end of dates, number of days, and number of EQs information of training, validation, and test sets are presented in Table 2. To illustrate near-real-time test performance of the proposed EQ-PD technique, precursor detections of the 24 EQs that had taken place in the region of choice shown in Table 3 in 2016 with magnitude greater than 4 in Richter scale are investigated (ANSS Comprehensive Catalog [ComCat], U.S. Geological Survey, n.d). During the performance evaluation, EQ with highest magnitude is chosen among EQs that had taken place at the same day for days with multiple EQs. Hence, 24 daily different EQs out of 43 EQs are investigated.

In this study, three different whole-day and three different nighttime TEC estimates are generated by choosing the estimation radius parameter R_r as 250, 300, and 350 km in section 3.1 for every reference station in the region during the days in between dates of 1 January 2005 and 30 September 2016. Thereafter, whole-day and nighttime anomaly detection signals are generated by measuring the whole-day $TEC_{\Delta}^{\text{w-day}}$ and nighttime $TEC_{\Delta}^{\text{night}}$ local TEC variations between the TEC measurements and their different estimates for every reference station in the region as detailed in section 3.2.

To obtain spatial statistical behaviors of these six different anomaly detection signals, upper-truncated Pareto distributions are estimated for the no-seismic activity days in between dates of 1 January 2005 and 1 April 2015 which is the final day of the training set. Similarly, spatiotemporal statistical behaviors of these six different anomaly detection signals are obtained by clustering the same no-seismic activity days into $k = 3$ different cluster of days and estimating a upper-truncated Pareto distribution for each different cluster as detailed in section 3.3. Note that the clustering of the geomagnetic parameters is implemented for the same time interval to obtain the cluster means to label the upcoming days of validation and test sets. Spatial and spatiotemporal anomaly detection thresholds are generated by choosing 23 different TEC_{Δ} tail probabilities on the estimated Pareto distributions for each anomaly detection threshold type.

Training set consists of a EQ precursor detection signal generated for the days in between the dates presented in Table 2. The signal is formed by applying $6 \times 2 \times 23 = 276$ different anomaly detection thresholds on the appropriate anomaly detection signals and obtaining 276 different daily anomaly decisions.Validation and test sets are formed by applying anomaly detection thresholds and cluster means that are obtained during the training on future anomaly detection signals with days in between the dates presented in Table 2.

To demonstrate EQ precursor detection performance of the proposed EQ-PD technique, multiple SVM models are trained on the training set to obtain decision boundary coefficients, $\boldsymbol{\theta}$, with different combinations

Table 3
The 24 Daily Different EQs That Have Taken Place in the Region of Choice

EQ # - Date	Time	EQ epicenter		*Mw*
(# - D Month YYYY)	(hh:mm:ss)	Latitude	Longitude	(Richter)
1 - 2 January 2016	12:36:28	36,4556°N	12,1175°E	4,3
2 - 6 January 2016	18:44:46	39,9354°N	15,5508°E	4,4
3 - 8 January 2016	13:07:42	42,9131°N	18,5153°E	4,6
4 - 13 January 2016	17:01:29	36,1699°N	14,7607°E	4,1
5 - 16 January 2016	18:55:11	41,5846°N	14,6515°E	4,4
6 - 17 January 2016	16:36:07	36,5834°N	12,8374°E	4,2
7 - 8 February 2016	15:35:43	37,023°N	14,8791°E	4,5
8 - 14 February 2016	14:51:29	43,0612°N	17,4307°E	4,4
9 - 4 April 2016	18:53:05	39,2273°N	15,4905°E	4,2
10 - 25 May 2016	22:10:28	36,81°N	15,79°E	4,1
11 - 30 May 2016	20:24:20	42,71°N	11,96°E	4,4
12 - 2 June 2016	10:49:12	36,5748°N	11,1289°E	4
13 - 23 June 2016	14:37:56	44,0943°N	9,9234°E	4,2
14 - 5 July 2016	05:54:38	37,5845°N	17,0005°E	4,5
15 - 30 July 2016	20:21:38	44,94°N	7,21°E	4,2
16 - 24 August 2016	01:36:32	42,723°N	13,1877°E	6,2
16 - 24 August 2016	01:56:01	42,6404°N	13,1986°E	4,6
16 - 24 August 2016	02:01:08	42,7856°N	13,1447°E	4,2
16 - 24 August 2016	02:05:55	42,6352°N	13,3022°E	4,1
16 - 24 August 2016	02:07:31	42,6243°N	13,1756°E	4,2
16 - 24 August 2016	02:19:44	42,6817°N	13,1667°E	4
16 - 24 August 2016	02:33:29	42,8413°N	13,1533°E	5,6
16 - 24 August 2016	02:51:27	42,7596°N	13,1812°E	4,1
16 - 24 August 2016	02:59:36	42,8104°N	13,1056°E	4,1
16 - 24 August 2016	03:40:11	42,6522°N	13,2549°E	4,2
16 - 24 August 2016	04:06:51	42,7959°N	13,0745°E	4,5
16 - 24 August 2016	11:50:31	42,8989°N	13,0834°E	4,6
16 - 24 August 2016	14:02:22	42,7989°N	13,1462°E	4,3
16 - 24 August 2016	17:46:11	42,7279°N	13,2082°E	4,3
16 - 24 August 2016	23:22:06	42,6589°N	13,1775°E	4,1
17 - 25 August 2016	03:17:16	42,7611°N	13,214°E	4,4
17 - 25 August 2016	04:51:41	42,6397°N	13,2478°E	4,2
17 - 25 August 2016	12:36:07	42,6654°N	13,1732°E	4,4
18 - 26 August 2016	04:28:25	42,6°N	13,29°E	4,8
19 - 27 August 2016	02:50:59	42,8608°N	13,2683°E	4,1
20 - 28 August 2016	13:07:34	42,6665°N	13,2488°E	4,3
20 - 28 August 2016	15:55:35	42,7975°N	13,1733°E	4,2
20 - 28 August 2016	16:42:02	42,8814°N	13,102°E	4,3
21 - 31 August 2016	18:12:52	42,8627°N	13,2213°E	4,1
22 - 3 September 2016	01:34:13	42,831°N	13,0956°E	4,4
22 - 3 September 2016	10:18:51	42,87°N	13,21°E	4,4
23 - 11 September 2016	18:39:02	42,68°N	13,28°E	4
24 - 19 September 2016	23:34:26	42,7166°N	13,1893°E	4,1

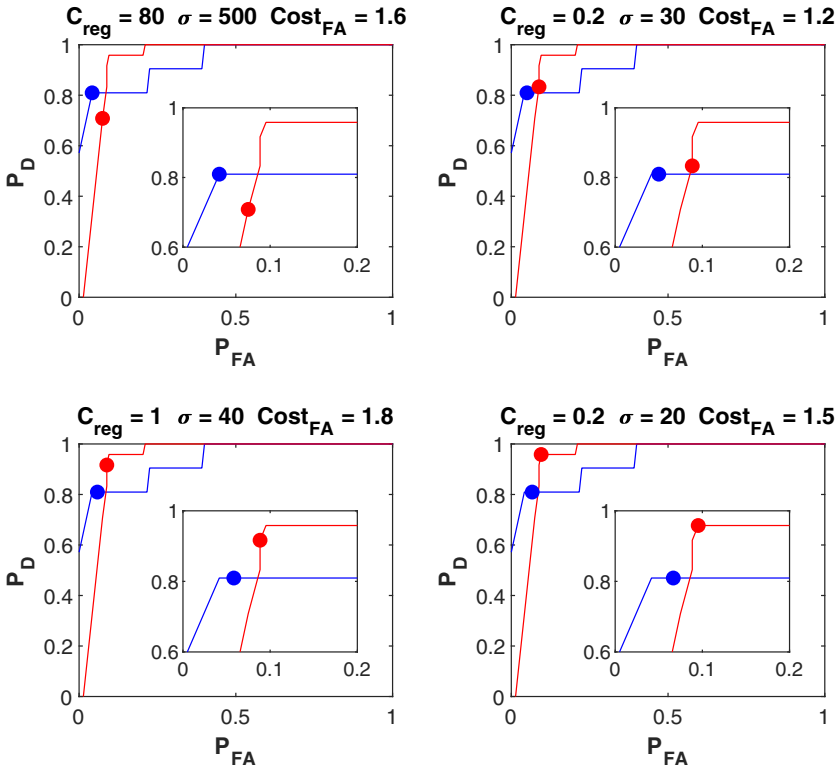

Figure 8. Validation (blue) and test (red) ROC curves with chosen model parameters and their corresponding validation (blue circle) and test (red circle) ROC points for Validation_1 and Test_1 (top left), Validation_2 and Test_2 (top right), Validation_3 and Test_3 (bottom left), and Validation_4 and Test_4 (bottom right).

of model parameters C_{reg}, σ, and $Cost_{FA}$. $Cost_{MD}$ parameter chosen as 1, C_{reg} parameter ranges between 0.01 and 10,000, σ parameter ranges between 0.001 and 1,000, and $Cost_{FA}$ parameter ranges between 1 and 2.1. Each trained SVM model is validated for different combinations of model parameters on the validation set. Among the validated SVM models, the model parameters that result with the highest probability of detection (P_D) for a given P_{FA} is chosen to obtain receiver operating characteristic (ROC) curve of validation set. ROC curve of test set is also obtained, when SVM models with chosen model parameters are applied on test set. Figure 8 illustrates validation and testing performances of the proposed technique on ROC space with four different model parameter sets and their corresponding validation and test performances. Additionally, number of detected EQ precursors, number of generated false alarms, and SVM training model parameters are given in Table 4. As expected, when number of generated false alarms are taken into consideration in Table 4, chosen model validation performances outperform the corresponding test performances.

For a given day, it is possible to obtain the reference station with the highest level of anomaly beyond the expected TEC_{Δ}. Therefore, every EQ precursor detection can be attributed to a reference station. EQ precursor detection distance of each EQ can be obtained by averaging the distances between chosen EQ epicenter and the reference station positions that contribute to the detection of the EQ precursor. Average EQ precursor detection distance can also be obtained by averaging the EQ precursor detection distances of all EQs. Average EQ precursor detection distances and precursor detection histograms of Test_1, Test_2, Test_3, and Test_4 ROC points that are presented in Figure 8 are visualized in Figure 9. As shown in Figure 9, the nearest average EQ precursor detection distance is observed in Test_1, while the farthest average EQ precursor detection distance is observed in Test_4. Also in Table 4, test performance results on average precursor detection distance for various EQ groups are presented.

During the SVM model parameter selection process, choosing a high C_{reg} parameter reduces the effect of θ coefficient regularization and causes overfitting of the model. Similarly, choosing a small σ parameter, causes greater variance on the features that are obtained with the Gaussian kernel and causes overfitting of the model. As exemplified in Table 4, neither overfitting nor underfitting is observed on the chosen model.

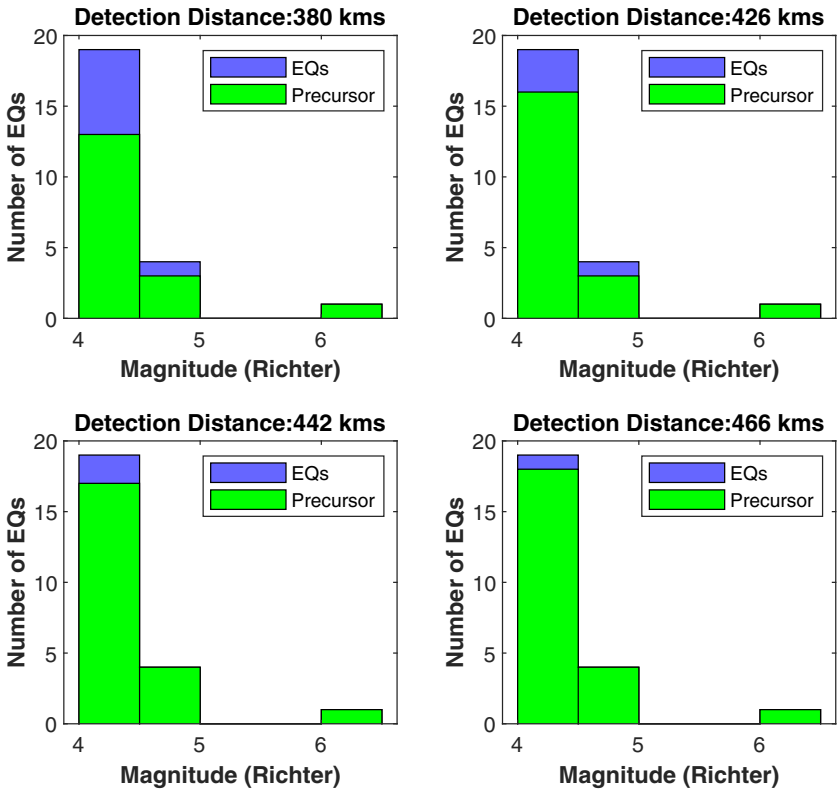

Figure 9. EQ precursor detection histograms and average precursor detection distances for Test_1 (top left), Test_2 (top right), Test_3 (bottom left), and Test_4 (bottom right) ROC points.

During the test, proposed EQ-PD technique generated 22 EQ precursor detections prior to 24 EQs and 13 false alarms in 147 days of no-seismic activity.

To compare EQ precursor detection performance of the proposed EQ-PD technique with random guessing, 100,000 Monte Carlo simulations are performed for each ROC point with their respective performance parameters. The following ROC performance parameters are defined for performance comparison purposes as: number of detected EQ precursors (TP_{EQ}), number of EQs ($N_{Pos,EQ}$), number of misdetected EQ precursors (FN_{EQ}), number of detected daily precursors (TP_{day}), number of seismically active days ($N_{Pos,day}$), number of misdetected daily precursors (FN_{day}), number of generated daily false alarms (FP_{day}), number of daily true negatives (TN_{day}), and number of possible false alarm days ($N_{Neg,day}$). Performance parameters of the chosen ROC points are presented in Table 4.

For each random guessing simulation, random and daily different $N_{Pos,EQ}$ EQ dates and $TP_{day} + FP_{day}$ precursory detection dates are generated, TP_{EQ} and FP_{day} parameters of random guessing simulation are calculated with respect to the time difference between EQ dates and precursory detection dates. Similarly, FN_{EQ} and TN_{day} parameters of the random guessing simulation are obtained by the equalities: $FN_{EQ} = N_{Pos,EQ} - TP_{EQ}$ and $TN_{day} = N_{Neg,day} - FP_{day}$. Performance of the proposed EQ-PD technique and the random guessing simulations is compared by using Matthews correlation coefficient (MCC) score defined as follows (Matthews, 1975):

$$MCC(TP, TN, FP, FN) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}. \quad (29)$$

MCC score gives an indication about how much better a given performance evaluation is than a random guessing performance. A $MCC = 1$ indicates perfect agreement between precursory detections and EQs, $MCC = 0$ is expected for precursory detections no better than random, and $MCC = -1$ indicates total disagreement between precursory detections and EQs.

Table 4
Performance Parameters of the Chosen ROC Points

ROC point		Validation$_1$	Test$_1$	Validation$_2$	Test$_2$	Validation$_3$	Test$_3$	Validation$_4$	Test$_4$
$TP_{EQ}/N_{Pos,EQ}$		17/21	17/24	17/21	20/24	17/21	22/24	17/21	23/24
$FN_{EQ}/N_{Pos,EQ}$		4/21	7/24	4/21	4/24	4/21	2/24	4/21	1/24
$TP_{day}/N_{Pos,days}$		123/146	81/135	123/146	99/135	123/146	115/135	123/146	125/135
$FN_{day}/N_{Pos,days}$		23/146	54/135	23/146	36/135	23/146	20/135	23/146	10/135
$FP_{day}/N_{Neg,days}$		5/120	11/147	6/120	13/147	7/120	13/147	8/120	14/147
$TN_{day}/N_{Neg,days}$		115/120	136/147	114/120	134/147	113/120	134/147	112/120	133/147
$N_{Pos,days}+N_{Neg}$		266	282	266	282	266	282	266	282
Mean (μ_{RG}) MCC score (RG)		0.28476	0.36973	0.28279	0.33400	0.28092	0.30479	0.27894	0.28388
Std (σ_{RG}) MCC score (RG)		0.06053	0.07059	0.06039	0.06597	0.06014	0.06332	0.06008	0.06171
$MCC_R(TP_{EQ},FN_{EQ},TN_{day},FP_{day})$		0.75335	0.59466	0.73187	0.65564	0.71163	0.71304	0.69248	0.72806
$Z=(MCC_R-\mu_{RG})/\sigma_{RG}$		7.7410	3.1859	7.4362	4.8751	7.1613	6.4471	6.8825	7.1969
Training	C_{reg}	80	—	0.2	—	1	—	0.2	—
Model	σ	500	—	30	—	40	—	20	—
Parameters	$Cost_{FA}$	1.6	—	1.2	—	1.8	—	1.5	—
Average precursor	M<4.3	—	389	—	495	—	508	—	545
detection distance	4.3≤M<4.6	—	415	—	402	—	431	—	443
(km)	4.6≤M	—	280	—	280	—	280	—	280

Note. For four different model parameter sets: Number of detected EQ precursors (TP_{EQ}), number of EQs ($N_{Pos,EQ}$), number of misdetected EQ precursors (FN_{EQ}), number of detected daily precursors (TP_{day}), number of seismic activity days ($N_{Pos,day}$), number of misdetected daily precursors (FN_{day}), number of generated daily false alarms (FP_{day}), number of daily true negatives (TN_{day}), number of possible false alarm days ($N_{Neg,days}$), MCC score mean of 100,000 random guessing experiments (μ_{RG}), MCC score standard deviation of 100,000 random guessing experiments (σ_{RG}), MCC score of the related ROC point (MCC_R), Z score of MCC_R with respect to μ_{RG} and σ_{RG}, validated SVM training model parameters, and average precursor detection distance for the 11 weak EQs with magnitude smaller than 4.3, 10 moderate EQs with magnitudes in between 4.3 and 4.6, and 3 strong EQs with magnitudes greater than 4.6 in Richter scale.

MCC scores of the related validation and test ROC points are presented in the tenth row of Table 4 and named as MCC_R. Mean and standard deviation values of 100,000 Monte Carlo simulations are defined as μ_{RG} and σ_{RG} and presented in the eighth and ninth rows of Table 4, respectively. Z score of the MCC_R with respect to μ_{RG} and σ_{RG} is a statistical indicator of performance. As shown in the eleventh row of Table 4, all of the MCC z scores are higher than 3 that performance of the proposed EQ-PD technique outperforms random guessing simulations for all chosen ROC points.

5. Conclusions and Future Research

Studies show the fact that the ionosphere is not only affected by position and radiation of the Sun or the geomagnetic activities but also affected by the seismic activities in the Earth's crust and surface.

In this study, a near-real-time EQ-PD technique based on detection of local ionospheric anomalies is presented and implemented by using ionospheric data obtained from the EPN and NASA Goddard Space Flight Center OMNIWeb service. The proposed EQ-PD technique performs local detection of anomalies as a binary hypothesis testing on signals that are generated by using the SKLD. In the proposed EQ-PD technique, the null hypothesis is defined as the locally generated SKLD-based signals in the period of 10 consecutive days ending with the day of the decision do not contain any anomaly due to an EQ precursor. An SVM classifier is used for the binary classification of the EQ precursors on the observed signals. Adaptive operation performance of the proposed EQ-PD technique has been evaluated with validation for the 21 EQs occurred in a chosen region covering Italy with magnitude higher than 4 in Richter scale during 2015 and tested for the 24 EQs occurred in the region with magnitude higher than 4 in Richter scale during 2016. It is observed that the proposed EQ-PD technique is able to detect 17 out of 21 EQ precursors while generating 7 false alarms during the validation and 22 out of 24 EQ precursors while generating 13 false alarms during the test. Furthermore, MCC analysis on the performance parameters of the proposed EQ-PD technique is performed by comparing these performances with random guessing simulations. It is observed that statistical performance of the EQ-PD technique is better than the random guessing in all simulated cases.

In conclusion, the obtained results suggest that EQ-PD technique has a significant potential for EQ precursor detection in near real time by monitoring local TEC variations in sufficiently dense GPS networks with respect to the state of the ionosphere. We believe that the EQ precursor detections of the EQ-PD technique can serve as distinctive features for an EQ prediction or forecasting technique that will be developed in the future.

In addition, alternative cost-sensitive supervised machine learning techniques will be investigated for further improvement on the performance of the proposed EQ-PD technique. Finally, the performance of the proposed EQ-PD technique should be compared with competitive or intelligent guessing techniques rather than completely random guessing.

Data Availability Statement

IONOLAB-TEC service and generated TEC data are available at the https://www.ionolab.org/index.php?page=index&language=en website. Raw ionospheric data and EUREF Permanent GNSS Network (EPN) reference station locations can be accessed via the https://www.epncb.oma.be/_networkdata/data_access/dailyandhourly/datacentres.php website. Daily A_p, AE, K_p, Dst indices, and SSN can be accessed via NASA Goddard Space Flight Center OMNIWeb service available at the https://omniweb.gsfc.nasa.gov/form/dx1.html website. EQ date, time, location and magnitude data are obtained from the Advanced National Seismic System's comprehensive earthquake catalog (ANSS ComCat) available at the https://earthquake.usgs.gov/earthquakes/search/ website.

Acknowledgments
We would like to thank anonymous reviewers for their constructive comments on the earlier versions of the manuscript. Based on their suggestions, we expanded statistical characterization of the proposed EQ-PD technique providing a strong basis for our conclusions. This study is supported by TUBITAK 114E541, 115E915, and Joint TUBITAK 114E092 and AS CR 14/001 projects. All equations and simulation parameters that are used during simulations are presented in the manuscript. The total electron content estimations were generated using IONOLAB-TEC service. The authors thank IONOLAB, EUREF, and NASA for providing ionospheric data.

References

Aban, I. B., Meerschaert, M. M., & Panorska, A. K. (2006). Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, *101*(473), 270–277.

Akhoondzadeh, M. (2016). Decision tree, bagging and random forest methods detect TEC seismo-ionospheric anomalies around the time of the Chile, (Mw = 8.8) earthquake of 27 february 2010. *Advances in Space Research*, *57*(12), 2464–2469.

Akyol, A. A. (2013). Investigation on the reliability of earthquake prediction based on ionospheric electron content variation (Master's Thesis), Bilkent University. http://www.thesis.bilkent.edu.tr/0007074.pdf

Akyol, A. A., Arikan, O., Arikan, F., & Deviren, M. N. (2013). Investigation on the reliability of earthquake prediction based on ionospheric electron content variation. In *Proceedings of the 16th International Conference on Information Fusion* (pp. 1658–1663). IEEE. https://ieeexplore.ieee.org/document/6641201

Allegri, L., Bella, F., Della Monica, G., Ermini, A., Improta, S., Sgrigna, V., & Biagi, P. F. (1983). Radon and rilt anomalies detected before the Irpinia (south Italy) earthquake of November 23, 1980 at great distances from the epicenter. *Geophysical Research Letters*, *10*(4), 269–272.

ANSS Comprehensive Catalog (ComCat), U.S. Geological Survey (n.d). https://earthquake.usgs.gov/earthquakes/search/. (Accessed:2017-04-17)

Arikan, F., Arikan, O., & Erol, C. B. (2007). Regularized estimation of TEC from GPS data for certain midlatitude stations and comparison with the IRI model. *Advances in Space Research*, *39*(5), 867–874.

Arikan, F., Deviren, M. N., Lenk, O., Sezen, U., & Arikan, O. (2012). Observed ionospheric effects of 23 October 2011 van, Turkey earthquake. *Geomatics, Natural Hazards and Risk*, *3*(1), 1–8.

Arikan, F., Erol, C. B., & Arikan, O. (2003). Regularized estimation of vertical total electron content from Global Positioning System data. *Journal of Geophysical Research*, *108*(A12), 1469. https://doi.org/10.1029/2002JA009605

Arikan, F., Erol, C. B., & Arikan, O. (2004). Regularized estimation of vertical total electron content from GPS data for a desired time period. *Radio Science*, *39*, RS6012. https://doi.org/10.1029/2004RS003061

Arikan, F., Nayir, H., Sezen, U., & Arikan, O. (2008). Estimation of single station interfrequency receiver bias using GPS-TEC. *Radio Science*, *43*, RS4004. https://doi.org/10.1029/2007RS003785

Arikan, F., Shukurov, S., Tuna, H., Arikan, O., & Gulyaeva, T. L. (2016). Performance of GPS slant total electron content and IRI-Plas-STEC for days with ionospheric disturbance. *Geodesy and Geodynamics*, *7*(1), 1–10.

Asencio-Cortés, G., Martínez-Álvarez, F., Morales-Esteban, A., & Reyes, J. (2016). A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. *Knowledge-Based Systems*, *101*, 15–30.

Asencio-Cortés, G., Martínez-Álvarez, F., Morales-Esteban, A., Reyes, J., & Troncoso, A. (2015). Improving earthquake prediction with principal component analysis: Application to Chile. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 393–404). Springer. https://www.doi.org/10.1007/978-3-319-19644-2_33

Asim, K. M., Idris, A., Iqbal, T., & Martínez-Álvarez, F. (2018). Earthquake prediction model using support vector regressor and hybrid neural networks. *PloS one*, *13*(7), e0199004.

Asteriadis, G., & Livieratos, E. (1989). Pre-seismic responses of underground water level and temperature concerning a 4.8 magnitude earthquake in Greece on October 20, 1988. *Tectonophysics*, *170*(1–2), 165–169.

Bendick, R., & Bilham, R. (2017). Do weak global stresses synchronize earthquakes? *Geophysical Research Letters*, *44*, 8320–8327. https://doi.org/10.1002/2017GL074934

Boucouvalas, A. C., Gkasios, M., Tselikas, N. T., & Drakatos, G. (2015). Modified-Fibonacci-Dual-Lucas method for earthquake prediction. In *Proc. of SPIE* (Vol. 9535, pp. 95351A–1).

Bruyninx, C., Habrich, H., Söhne, W., Kenyeres, A., Stangl, G., & Völksen, C. (2012). Enhancement of the EUREF permanent network services and products, *Geodesy for planet Earth* (pp. 27–34). Springer.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167.

Davidenko, D. V., & Pulinets, S. A. (2019). Deterministic variability of the ionosphere on the eve of strong ($M \geq 6$) earthquakes in the regions of Greece and Italy according to long-term measurements data. *Geomagnetism and Aeronomy*, *59*(4), 493–508.
Davies, K., & Hartmann, G. K. (1997). Studying the ionosphere with the Global Positioning System. *Radio Science*, *32*(4), 1695–1703.
Devi, M., Medhi, A., Sarma, A. J. D. S., & Barbara, A. K. (2013). Growth and inhibition of equatorial anomaly prior to an earthquake (EQ): Case studies with total electron content (TEC) data for major EQs of Japan 2011 and Indonesia 2012. *Positioning*, *2013*.
Deviren, M. N., Arikan, F., & Arikan, O. (2013). Spatio-temporal interpolation of total electron content using a GPS network. *Radio Science*, *48*, 302–309. https://doi.org/10.1002/rds.20036
FAQs, Can you predict earthquakes?, U.S. Geological Survey (n.d.) https://www.usgs.gov/faqs/can-you-predict-earthquakes. (Accessed:2018-04-03)
Fidani, C. (2010). The earthquake lights (EQL) of the 6 April 2009 Aquila earthquake, in Central Italy. *Natural Hazards and Earth System Sciences*, *10*(5), 967–978.
Geller, R. J. (1997). Earthquake prediction: A critical review. *Geophysical Journal International*, *131*, 425–450.
Grant, R. A., & Halliday, T. (2010). Predicting the unpredictable; Evidence of pre-seismic anticipatory behaviour in the common toad. *Journal of Zoology*, *281*(4), 263–271.
Gulyaeva, T., & Arikan, F. (2017). Statistical discrimination of global post-seismic ionosphere effects under geomagnetic quiet and storm conditions. *Geomatics, Natural Hazards and Risk*, *8*(2), 509–524.
Gulyaeva, T. L., Arikan, F., & Stanislawska, I. (2017a). Earthquake aftereffects in the equatorial ionization anomaly region under geomagnetic quiet and storm conditions. *Advances in Space Research*, *60*(2), 406–418.
Gulyaeva, T. L., Arikan, F., & Stanislawska, I. (2017b). Persistent long-term (1944–2015) ionosphere-magnetosphere associations at the area of intense seismic activity and beyond. *Advances in Space Research*, *59*(4), 1033–1040.
Gulyaeva, T. L., Arikan, F., Stanislawska, I., & Poustovalova, L. V. (2016). Global distribution of zones of enhanced risk for the ionospheric weather. *Journal of Geography, Environment and Earth Science International*, *4*(1), 1–13.
Hartmann, J., & Levy, J. K. (2005). Hydrogeological and gasgeochemical earthquake precursors—A review for application. *Natural Hazards*, *34*(3), 279–304.
Ibrahim, M. A., Park, J., & Athens, N. (2018). Earthquake warning system: Detecting earthquake precursor signals using deep neural networks. Technical Report CS 230. http://cs230.stanford.edu/projects_spring_2018/reports/8291012.pdf
IONOLAB, Ionospheric Research Laboratory (n.d). https://www.ionolab.org/
Kannan, S. (2014). Innovative mathematical model for earthquake prediction. *Engineering Failure Analysis*, *41*, 89–95.
Karatay, S., Arikan, F., & Arikan, O. (2009). Investigation of hourly and daily patterns for lithosphere-ionosphere coupling before strong earthquakes. In *Recent Advances in Space Technologies, 2009. RAST'09. 4th International Conference on* (pp. 670–674). IEEE.
Karatay, S., Arikan, F., & Arikan, O. (2010). Investigation of total electron content variability due to seismic and geomagnetic disturbances in the ionosphere. *Radio Science*, *45*, RS5012. https://doi.org/10.1029/2009RS004313
Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). Hoboken, NJ, USA: John Wiley & Sons. https://doi.org/10.1002/9780470316801
Kon, S., Nishihashi, M., & Hattori, K. (2011). Ionospheric anomalies possibly associated with $M \geq 6.0$ earthquakes in the Japan area during 1998–2010: Case studies and statistical study. *Journal of Asian Earth Sciences*, *41*(4–5), 410–420.
Kouris, S. S., & Fotiadis, D. N. (2002). Ionospheric variability: A comparative statistical study. *Advances in Space Research*, *29*(6), 977–985.
Kouris, S. S., Polimeris, K. V., & Cander, L. R. (2006). Specifications of TEC variability. *Advances in Space Research*, *37*(5), 983–1004.
Last, M., Rabinowitz, N., & Leonard, G. (2016). Predicting the maximum earthquake magnitude from seismic data in israel and its neighboring countries. *PloS one*, *11*(1), e0146101.
Le, H., Liu, J.-Y., & Liu, L. (2011). A statistical analysis of ionospheric anomalies before 736 *M*6.0+ earthquakes during 2002–2010. *Journal of Geophysical Research*, *116*, A02303. https://doi.org/10.1029/2010JA015781
Li, W., Guo, J., Yue, J., Yang, Y., Li, Z., & Lu, D. (2016). Contrastive research of ionospheric precursor anomalies between Calbuco volcanic eruption on April 23 and Nepal earthquake on April 25, 2015. *Advances in Space Research*, *57*(10), 2141–2153.
Liu, J. Y., Chen, C. H., Chen, Y. I., Yang, W. H., Oyama, K. I., & Kuo, K. W. (2010). A statistical study of ionospheric earthquake precursors monitored by using equatorial ionization anomaly of GPS TEC in Taiwan during 2001–2007. *Journal of Asian Earth Sciences*, *39*(1), 76–80.
Liu, J. y., Chen, Y. I., Chuo, Y. J., & Chen, C.-S. (2006). A statistical investigation of preearthquake ionospheric anomaly. *Journal of Geophysical Research*, *111*, A05304. https://doi.org/10.1029/2005JA011333
Liu, J.-Y., Chen, Y.-I., Jhuang, H.-K., & Lin, Y.-H. (2004). Ionospheric *foF*2 and TEC anomalous days associated with $M >= 5.0$ earthquakes in Taiwan during 1997–1999. *Terrestrial Atmospheric and Oceanic Sciences*, *15*(3), 371–384.
Liu, J. Y., Chen, Y. I., Pulinets, S. A., Tsai, Y. B., & Chuo, Y. J. (2000). Seismo-ionospheric signatures prior to $M \geq 6.0$ Taiwan earthquakes. *Geophysical Research Letters*, *27*(19), 3113–3116.
Liu, J. Y., Chuo, Y. J., Shan, S. J., Tsai, Y. B., Chen, Y. I., Pulinets, S. A., Yu, S. B., et al. (2004). Pre-earthquake ionospheric anomalies registered by continuous GPS TEC measurements. In *Annales Geophysicae* (Vol. 22, pp. 1585–1593).
Mahmoudi, J., Arjomand, M. A., Rezaei, M., & Mohammadi, M. H. (2016). Predicting the earthquake magnitude using the multilayer perceptron neural network with two hidden layers. *Civil Engineering Journal*, *2*(1), 1–12.
Mandea, M., & Korte, M. (Eds.) (2010). *Geomagnetic observations and models* (Vol. 5). NY, USA: Springer Science & Business Media. https://doi.org/10.1007/978-90-481-9858-0
Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442–451.
Moustra, M., Avraamides, M., & Christodoulou, C. (2011). Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. *Expert Systems with Applications*, *38*(12), 15,032–15,039.
Namgaladze, A. A., Zolotov, O. V., Zakharenkova, I. E., Shagimuratov, I. I., & Martynenko, O. V. (2009). Ionospheric total electron content variations observed before earthquakes: Possible physical mechanism and modeling. arXiv preprint arXiv:0905.3313. https://arxiv.org/ftp/arxiv/papers/0905/0905.3313.pdf
NASA OMNIWeb, SPDF Goddard Space Flight Center (n.d.) https://omniweb.gsfc.nasa.gov/form/dx1.html. (Accessed:2017-04-17)
Nayir, H., Arikan, F., Arikan, O., & Erol, C. B. (2007). Total electron content estimation with Reg-Est. *Journal of Geophysical Research*, *112*, A11313. https://doi.org/10.1029/2007JA012459
Necat Deviren, M., Gulyaeva, T., Sezen, U., Arikan, F., & Arikan, O. (2014). Detection of seismic precursors using distance metrics between GPS-TEC and IRI-Plas. In *40th COSPAR Scientific Assembly* (Vol. 40).

Ouzounov, D., Pulinets, S., Romanov, A., Romanov, A., Tsybulya, K., Davidenko, D., et al. (2011). Atmosphere-ionosphere response to the M9 Tohoku earthquake revealed by multi-instrument space-borne and ground observations: Preliminary results. *Earthquake Science*, *24*(6), 557–564. https://doi.org/10.1007/s11589-011-0817-z

Oyama, K.-I., Devi, M., Ryu, K., Chen, C. H., Liu, J. Y., Liu, H., et al. (2016). Modifications of the ionosphere prior to large earthquakes: Report from the ionosphere precursor study group. *Geoscience Letters*, *3*(1), 6.

Panakkat, A., & Adeli, H. (2009). Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. *Computer-Aided Civil and Infrastructure Engineering*, *24*(4), 280–292.

Papazachos, B. C., & Papaioannou, C. A. (1993). Long-term earthquake prediction in the Aegean area based on a time and magnitude predictable model. *Pure and Applied Geophysics*, *140*(4), 593–612.

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-14.pdf

Plotkin, V. V. (1999). GPS detection of ionospheric perturbation before the 13 February 2001, El Salvador earthquake. *Natural Hazards and Earth System Science*, *3*(3/4), 249–253.

Pulinets, S. A. (1998). Seismic activity as a source of the ionospheric variability. *Advances in Space Research*, *22*(6), 903–906.

Pulinets, S. (2004). Ionospheric precursors of earthquakes; Recent advances in theory and practical applications. *Terrestrial Atmospheric and Oceanic Sciences*, *15*(3), 413–436.

Pulinets, S. A., Boyarchuk, K. A., Lomonosov, A. M., Khegai, V. V., & Lyu, I. Y. (2002). Ionospheric precursors to earthquakes: A preliminary analysis of the *foF*2 critical frequencies at Chung-Li ground-based station for vertical sounding of the ionosphere (Taiwan island). *Geomagnetism and Aeronomy*, *42*(4), 508–513.

Pulinets, S. A., & Davidenko, D. V. (2018). The nocturnal positive ionospheric anomaly of electron density as a short-term earthquake precursor and the possible physical mechanism of its formation. *Geomagnetism and Aeronomy*, *58*(4), 559–570.

Pulinets, S. A., Gaivoronska, T. B., Contreras, A. L., Ciraolo, L., et al. (2004). Correlation analysis technique revealing ionospheric precursors of earthquakes. *Natural Hazards and Earth System Science*, *4*(5/6), 697–702.

Pulinets, S. A., Kotsarenko, A. N., Ciraolo, L., & Pulinets, I. A. (2007). Special case of ionospheric day-to-day variability associated with earthquake preparation. *Advances in Space Research*, *39*(5), 970–977.

Pulinets, S. A., Leyva-Contreras, A., Bisiacchi-Giraldi, G., & Ciraolo, C. (2005). Total electron content variations in the ionosphere before the Colima, Mexico, earthquake of 21 January 2003. *Geofisica Internacional*, *44*(4).

Pulinets, S., & Ouzounov, D. (2011). Lithosphere-Atmosphere-Ionosphere Coupling (LAIC) model—An unified concept for earthquake precursors validation. *Journal of Asian Earth Sciences*, *41*(4–5), 371–382.

Pulinets, S. A., Ouzounov, D., Karelin, A. V., Boyarchuk, K. A., & Pokhmelnykh, L. A. (2006). The physical nature of thermal anomalies observed before strong earthquakes. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(4), 143–153.

Rostoker, G. (1972). Geomagnetic indices. *Reviews of Geophysics*, *10*(4), 935–950.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

SILSO World Data Center (2005). The International Sunspot Number. International Sunspot Number Monthly Bulletin and online catalogue.

Sayin, I., Arikan, F., & Arikan, O. (2008). Regional TEC mapping with random field priors and kriging. *Radio Science*, *43*, RS5012. https://doi.org/10.1029/2007RS003786

Sezen, U., Arikan, F., Arikan, O., Ugurlu, O., & Sadeghimorad, A. (2013). Online, automatic, near-real time estimation of GPS-TEC: IONOLAB-TEC. *Space Weather*, *11*, 297–305. https://doi.org/10.1002/swe.20054

Stanislawska, I., Juchnikowski, G., Cander, L. R., Ciraolo, L., Bradley, P. A., Zbyszynski, Z., & Swiatek, A. (2002). The kriging method of TEC instantaneous mapping. *Advances in Space Research*, *29*(6), 945–948.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. NY, USA: Springer. https://doi.org/10.1007/978-1-4612-1494-6

Trigunait, A., Parrot, M., Pulinets, S., Li, F., et al. (2004). Variations of the ionospheric electron density during the Bhuj seismic event. In *Annales Geophysicae* (Vol. 22, pp. 4123–4131).

Tronin, A. A., Biagi, P. F., Molchanov, O. A., Khatkevich, Y. M., & Gordeev, E. I. (2004). Temperature variations related to earthquakes from simultaneous observation at the ground stations and by satellites in Kamchatka area. *Physics and Chemistry of the Earth, Parts A/B/C*, *29*(4), 501–506.

Tronin, A. A., Hayakawa, M., & Molchanov, O. A. (2002). Thermal IR satellite data application for earthquake research in Japan and China. *Journal of Geodynamics*, *33*(4), 519–534.

Turel, N., & Arikan, F. (2010). Probability density function estimation for characterizing hourly variability of ionospheric total electron content. *Radio Science*, *45*, RS6016. https://doi.org/10.1029/2009RS004345

Wang, Q., Guo, Y., Yu, L., & Li, P. (2017). Earthquake prediction based on spatio-temporal data mining: An LSTM network approach. *IEEE Transactions on Emerging Topics in Computing*.

Wyss, M. (1997). Cannot earthquakes be predicted? *Science*, *278*(5337), 487–490.

Wyss, M., Aceves, R., Park, S., Geller, R., Jackson, D., Kagan, Y., & Mulargia, F. (1997). Cannot Earthquakes be Predicted? *Science*, *278*(5337), 487–490. Retrieved November 9, 2020, from http://www.jstor.org/stable/2894929

Yadav, K. S., Karia, S. P., & Pathak, K. N. (2015). Anomalous variation in GPS based TEC prior to the 5 earthquakes in 2009 and 2010. *Positioning*, *6*(04), 96.

Zaninetti, L., & Ferraro, M. (2008). On the truncated Pareto distribution with applications. *Central European Journal of Physics*, *6*(1), 1–6.

Zhang, M. L., Shi, J. K., Wang, X., & Radicella, S. M. (2004). Ionospheric variability at low latitude station: Hainan, China. *Advances in Space Research*, *34*(9), 1860–1868.

Zolotov, O. V., Namgaladze, A. A., & Prokhorov, B. E. (2013). Specific features of ionospheric total electron content variations in the periods of preparation of the earthquakes on March 11, 2011 (Japan) and October 23, 2011 (Turkey). *Russian Journal of Physical Chemistry B*, *7*(5), 599–605.