



# Support vector machines for TEC seismo-ionospheric anomalies detection

M. Akhoondzadeh

Remote Sensing Division, Surveying and Geomatics Engineering Department, University College of Engineering, University of Tehran, Tehran, Iran

Correspondence to: M. Akhoondzadeh (makhonz@ut.ac.ir)

Received: 3 June 2012 – Revised: 12 January 2013 – Accepted: 14 January 2013 – Published: 6 February 2013

**Abstract.** Using time series prediction methods, it is possible to pursue the behaviors of earthquake precursors in the future and to announce early warnings when the differences between the predicted value and the observed value exceed the predefined threshold value. Support Vector Machines (SVMs) are widely used due to their many advantages for classification and regression tasks. This study is concerned with investigating the Total Electron Content (TEC) time series by using a SVM to detect seismo-ionospheric anomalous variations induced by the three powerful earthquakes of Tohoku (11 March 2011), Haiti (12 January 2010) and Samoa (29 September 2009). The duration of TEC time series dataset is 49, 46 and 71 days, for Tohoku, Haiti and Samoa earthquakes, respectively, with each at time resolution of 2 h. In the case of Tohoku earthquake, the results show that the difference between the predicted value obtained from the SVM method and the observed value reaches the maximum value (i.e., 129.31 TECU) at earthquake time in a period of high geomagnetic activities. The SVM method detected a considerable number of anomalous occurrences 1 and 2 days prior to the Haiti earthquake and also 1 and 5 days before the Samoa earthquake in a period of low geomagnetic activities. In order to show that the method is acting sensibly with regard to the results extracted during nonevent and event TEC data, i.e., to perform some null-hypothesis tests in which the methods would also be calibrated, the same period of data from the previous year of the Samoa earthquake date has been taken into the account. Further to this, in this study, the detected TEC anomalies using the SVM method were compared to the previous results (Akhoondzadeh and Saradjian, 2011; Akhoondzadeh, 2012) obtained from the mean, median, wavelet and Kalman filter methods. The SVM detected anomalies are similar to those detected using the previous

methods. It can be concluded that SVM can be a suitable learning method to detect the novelty changes of a nonlinear time series such as variations of earthquake precursors.

**Keywords.** Ionosphere (Ionospheric irregularities)

## 1 Introduction

Due to the nonsystematic behavior and intrinsic complexity of earthquake anomalies that is related to the complication of its physical mechanism, earthquake prediction has become a difficult yet challenging task. Appropriate detection of the novelty phenomena in a nonlinear time series, such as seismo-ionospheric precursors, could lead to reduction of the uncertainty in the estimation of earthquake parameters (i.e., time, location and magnitude). Since the variations of an earthquake precursor depend on many nonseismic parameters and cannot be easily acknowledged using the earthquake physical mechanism, the mathematical methods implementing pattern recognition independent from the physical models could be developed.

Support Vector Machines (SVMs) are widely used due to their many advantages for classification and regression tasks. In previous studies it has been indicated that in contrast to classical and intelligent methods such as Autoregressive Moving Average (ARMA) and Artificial Neural Network (ANN), the SVMs can make models in a much more efficient way, using less data than the other methods (Thissen et al., 2003). It can be shown that, if the difference between the predicted value using the SVM method and the observed value exceeds the predefined threshold value, then the observed unusual precursor value in the absence of nonseismic effective parameters could be regarded as earthquake

anomaly. Liu et al. (2004) statistically described the temporal parameters of the seismo-ionospheric precursors detected during 1–5 days prior to the earthquakes using TEC (Total Electron Content) data for 20 major earthquakes in Taiwan. Xu et al. (2011) reported the existence of the unusual changes in latent heat flux and TEC 3 days prior to the powerful Tohoku earthquake of 11 March 2011. Klimenko et al. (2011) suggested that such great TEC enhancement observed 3 days prior to the Wenchuan earthquake could be explained by combined action of seismogenic vertical electric field and internal gravity waves (IGWs) generated by the solar terminator. After removing the influence of solar radiation origin in GIM (Global Ionospheric Map) TEC, the analysis results of He et al. (2012) show that the TEC around the forthcoming epicenter and its conjugate were significantly enhanced in the afternoon period of 8 March 2011, 3 days before the Tohoku earthquake.

### 1.1 TEC precursor

The effects of the preseismic activity on the ionosphere can be investigated using the ionospheric electron density variations. These anomalies usually happen in the D-layer, E-layer and F-layer and may be observed 1 to 10 days prior to the earthquake, and continue a few days after it (Hayakawa and Molchanov, 2002; Pulnits and Boyarchuk, 2004; Akhoondzadeh, 2011).

Currently, thousands of GPS receivers are used to monitor the Earth's surface deformations. TEC data retrieved from GPS measurements have made a considerable contribution to the understanding of seismo-ionospheric variations (Liu et al., 2004). In this study TEC variations have been analyzed using GIM data provided by the NASA Jet Propulsion Laboratory (JPL). The GIM is constructed from a  $5^\circ \times 2.5^\circ$  (longitude, latitude) grid with a time resolution of 2 h.

Hypotheses exist to explain the seismic electromagnetic mechanism based on geophysical and geochemical processes (Pulnits and Boyarchuk, 2004). Preseismic electric field and its polarity cause the electrons in the F-layer to penetrate to lower layers and therefore to create an anomaly in the ionospheric parameters. The thin layer of particles created before earthquakes due to ions radiation from the earth has a main role in transferring the electric field into the atmosphere and then to the ionosphere. The vertical electric field on the ground surface is transformed into an electric field perpendicular to the geomagnetic field lines. This zonal component leads to plasma density anomalies, which are observed prior to the earthquake (Namgaladze et al., 2009; Pulnits, 2009).

### 1.2 Geomagnetic indices

The ionospheric parameters are affected by solar geophysical conditions and geomagnetic storms especially in the equatorial and polar regions. Also, auroral activity has an important role in the mid-latitude ionosphere perturbations.

The ionosphere current and equatorial storm-time ring current in periods of solar–terrestrial disturbances produce significant geomagnetic field disturbances. The measured such parameters may display variations even in absence of seismic activity making it difficult to separate preseismic ionospheric phenomena from the ionospheric disturbances due to the solar–terrestrial activities. Therefore, to discriminate the seismo-ionospheric perturbations from geomagnetic disturbances, the geomagnetic and solar indices (i.e., Dst, Kp and F10.7) accessed through NOAA (<http://spider.ngdc.noaa.gov>) have been checked. The Kp index monitors the planetary activity on a worldwide scale, while the Dst index records the equatorial ring current variations (Mayaud, 1980).

## 2 Methodology

Classical models such as autoregressive moving average (ARMA) and intelligent methods such as artificial neural networks (ANNs) are well known methods for time series prediction. The ARMA method is easily and quickly implemented, and it efficiently acts for linear solutions. ANNs are a class of intelligent systems that can discover patterns with a few a priori assumptions and learn any complex functional relationship from the data to model a phenomenon. But these methods do not lead to unique solutions due to differences in their initial parameters, including the number of pattern inputs, lag values, the number of hidden layers, and their number of neurons.

SVMs have been applied for classification and regression tasks but their principles can be extended feasibly to the time series prediction. In linear learning machines such as SVMs, a linear function ( $f(x) = wx + b$ ) is used to solve the regression problem. The best line is defined to be that line which minimizes the following cost function (CF) (Thissen et al., 2003; Muller et al., 1997):

$$CF = 0.5 \|w\|^2 + C \sum_{i=1}^N L^\varepsilon(x_i, y_i, f)$$

$$\text{Subject to: } \begin{aligned} y_i - wx_i - b &\leq \varepsilon + \xi_i; \\ wx_i + b - y_i &\leq \varepsilon + \xi_i^*; \\ \xi_i \xi_i^* &\geq 0 \end{aligned} \quad (1)$$

where  $w$  is a weight decay which is used to regularize weight sizes and penalizes large weights. Since the large weights increase the variance and therefore decrease the efficiency of the SVM, using this regularization the weights meet the smaller values. The second part of the cost function is a penalty function which penalizes errors larger than  $\pm\varepsilon$  using a so called  $\varepsilon$ -insensitive loss function  $L^\varepsilon$  for each of the  $N$  training points. The positive constant  $C$  indicates the toleration value of  $\varepsilon$  deviations. Errors exceeding the  $\pm\varepsilon$  are depicted with the so-called slack variables  $\xi$  and  $\xi^*$ , respectively. The last parts of the Eq. (1) are constraints showing

the predictive errors between the predictions ( $w x_i + b$ ) and true values ( $y_i$ ).

The minimization of Eq. (1) can be done by applying Lagrangian theory. Based on this theory the weight vector is equal to the linear combination of the training data:

$$w = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i, \tag{2}$$

where  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers that are related to a specific training point. The asterisk again indicates difference above and below the regression line. Using introducing of this equation into the linear function,  $f(x) = w x + b$ , the following solution is derived for an unknown data point  $x$  (Thissen et al., 2003):

$$\begin{aligned} f(x) &= \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b \\ &= \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \\ K(x_i, x) &= \langle \phi(x_i), \phi(x) \rangle, \end{aligned} \tag{3}$$

where  $K$  is the so-called kernel function.

Burges (1998) has shown that the solution of this equation is global and unique because the cost function is strictly convex. Since some Lagrange multipliers are zero, not all training points contribute to the solution. In other words, if these training points with zero Lagrange multipliers are not introduced, the same solutions are obtained. Training points with nonzero Lagrange multipliers are called support vectors and represent the shape of the solution (Thissen et al., 2003).

This method can be extended by mapping the data  $x_i$  into a high-dimensional feature space via a nonlinear mapping  $\phi(x_i)$ . Without full knowledge of  $\phi(x_i)$ , using the kernel function, the data can be mapped into a feature. The most used kernel functions are the Gaussian RBF kernel:

$$K(x_i, x) = \exp(-\|x - x_i\|^2 / 2\sigma^2). \tag{4}$$

The values of  $\varepsilon$  and  $C$  and also the type of the kernel and its parameters are tuned by the user, and the aforesaid parameters can be chosen based on an optimized solution. If time series prediction is seen as autoregression in time, then a regression method can be used for this task (Thissen et al., 2003).

A time series is a set of time-ordered observations  $x_t$ , each one being recorded at a specific time  $t$ . To start prediction process by SVMs, one input object ( $x_i$ ) to the SVM is a time series of consecutive measurements:  $x_i = \{x(t_i), x(t_i - s), \dots, x(t_i - \tau s)\}$ , where  $s$  is the sampling time step (i.e., 1 day) and  $\tau$  determines the time window and thus the number of elements of the input vector. The output of the regression,  $y_i$ , is equal to  $x(t_i + h)$ , where  $h$  is the prediction horizon. The parameter of  $h$  is selected as 3. When performing time series prediction, the input window becomes an additional tunable parameter (Thissen et al., 2003).

To implement the SVM method, training, validating and testing data were initially set respectively to 40 %, 20 % and 40 % of all TEC data. The parameter of  $\tau$  is selected as 2. The input patterns in the SVM method are

$$\begin{aligned} x_4 &= f(x_1, x_2, x_3) \\ x_5 &= f(x_2, x_3, x_4) \\ &\vdots \\ x_N &= f(x_{N-3}, x_{N-2}, x_{N-1}). \end{aligned} \tag{5}$$

At each step, using the training data, the SVM method is implemented and then the predictive error is minimized during the validation of data. The prediction error (PE) can be written as

$$PE = \sum_{i=4}^N (x_i - \hat{x}_i), \tag{6}$$

where  $x_i$  and  $\hat{x}_i$  are the actual value and the output from the SVM method, respectively.

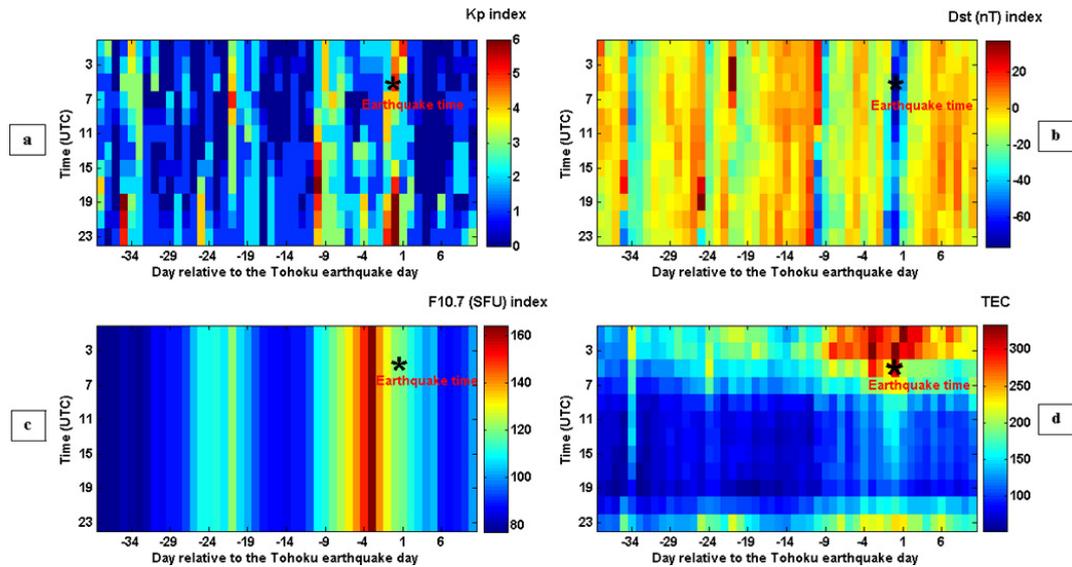
Finally, the TEC value is predicted and then is compared to the true value in the testing set. Increasing the size of the training set does not lead to a decrease of the predictive error. The training set is used to determine the best model settings while the test set is used to determine the final predictive errors for each prediction horizon (Thissen et al., 2003). At the next step, the size of the validating set increases by the order of 1 and also the size of the testing set decreases by the order of 1. The SVM algorithm is executed and the TEC value is estimated again for the next time. The process is repeated until the all of TEC values are predicted. In the case of the testing process, if the value of  $DX_i$  (i.e., the difference between the actual value  $X_i$  and the predicted value  $\hat{X}_i$ ) is outside the predefined bounds  $\mu \pm 1.5 \times \sigma$ , ( $\mu$  and  $\sigma$  are the mean and the standard deviation of  $DX_i$  values) the anomaly is detected.

### 3 Implementation

The implementation has been performed on three earthquakes as case studies (Table 1). The first case study is an earthquake which occurred near the northeast coast of Honshu in Japan with a magnitude of  $M_w = 9.0$  on 11 March 2011 at 14:46:23 LT (UTC = LT - 09:00). The second one is about an earthquake that happened in Haiti with a magnitude of  $M_w = 7.0$  on 12 January 2010 at 16:53:10 LT (UTC = LT + 05:00). The third case study is about a strong earthquake of magnitude  $M_w = 8.1$  that occurred in Samoa Island on 29 September 2009 at 06:48:11 LT (UTC = LT + 11:00).

#### 3.1 Tohoku earthquake

Figure 1a illustrates the variations of Kp geomagnetic index during the period of 1 February to 21 March 2011. An



**Fig. 1.** (a), (b) and (c) show, respectively, the variations of Kp, Dst and F10.7 geomagnetic and solar indices during the period of 1 February to 21 March 2011. (d) TEC variations. An asterisk indicates the earthquake time. The x-axis represents the days relative to the Tohoku earthquake day. The y-axis represents the universal time coordinate.

**Table 1.** List of the earthquakes selected in this study (reported by <http://earthquake.usgs.gov/>).

Area	Date	Time (UTC)	Geographic latitude, longitude	Magnitude ( $M_w$ )	Focal depth (km)
Samoa Islands	29 September 2009	17:48:10.99	15.49° S, 172.10° W	8.1	18.0
Haiti	12 January 2010	21:53:10	18.457° N, 72.533° W	7.0	13.0
Tohoku	11 March 2011	05:46:23	38.322° N, 142.369° E	9.0	29.0

asterisk indicates the earthquake time. The x-axis represents the days relative to the earthquake day. The y-axis represents the universal time coordinate. The high geomagnetic activities are seen 35 and 21 days before the earthquake onset. The high Kp values between 14:00 and 22:00 UTC, 10 days before the earthquake and also the moderate values of this index from 7 to 9 days before the event can be interpreted as high geomagnetic activities. The Kp value reaches the values of 4, 4.5 and 5 between 18:00 and 24:00 UTC 1 day before the earthquake, and increases to the maximum value of 6.0 from 12:00 to 18:00 UTC after the main shock. These unusual variations of Kp index from 1 day before to 1 day after the earthquake can hide pre and postseismic ionospheric anomalies.

Figure 1b shows the variations of Dst geomagnetic index during the period of 1 February to 21 March 2011. The unusual Dst values observed from 7 to 10 days and also 32 to 35 days before the earthquake indicate the high geomagnetic activities. It can be concluded that the detected perturbations on TEC variations during these periods might not be related to seismic activity. The Dst value exceeds the lower boundary value (i.e.,  $-20$  nT) at 08:00 UTC 1 day before the earthquake, and then gradually decreases and reaches the

minimum value of  $-76.5$  (nT) at earthquake time. The unusual variations of Dst values are seen up to 1 day after the earthquake. This study examines the TEC variations during the period of the Tohoku earthquake to find the preseismic anomalies in low geomagnetic activities ( $Kp < 2.5$  and  $Dst > -20$  nt).

Figure 1c shows the variations of solar radio flux (F10.7) during the period of 1 February to 21 March 2011. F10.7 is often expressed in SFU or solar flux units. The F10.7 value gradually increases from about 14 days before the earthquake and reaches the maximum value of 164.30 (SFU) on 8 March 2011, which is 3 days before the event. High levels of sunspot activity lead to improved signal propagation on higher frequency bands, although they also increase the levels of solar noise and ionospheric disturbances. After four years without any X-flares, the sun produced two of the powerful blasts in less than one month: one taking place on 19 February 2011, and the other taking place on 9 March 2011. 9 March ended with a powerful solar flare. Earth-orbiting satellites detected an X1.5-class explosion from behemoth sunspot 1166 around 23:23 UTC. In addition, on 10 March 2011 around 06:30 UTC, a coronal mass ejection (CME) struck the magnetic field of the Earth. This

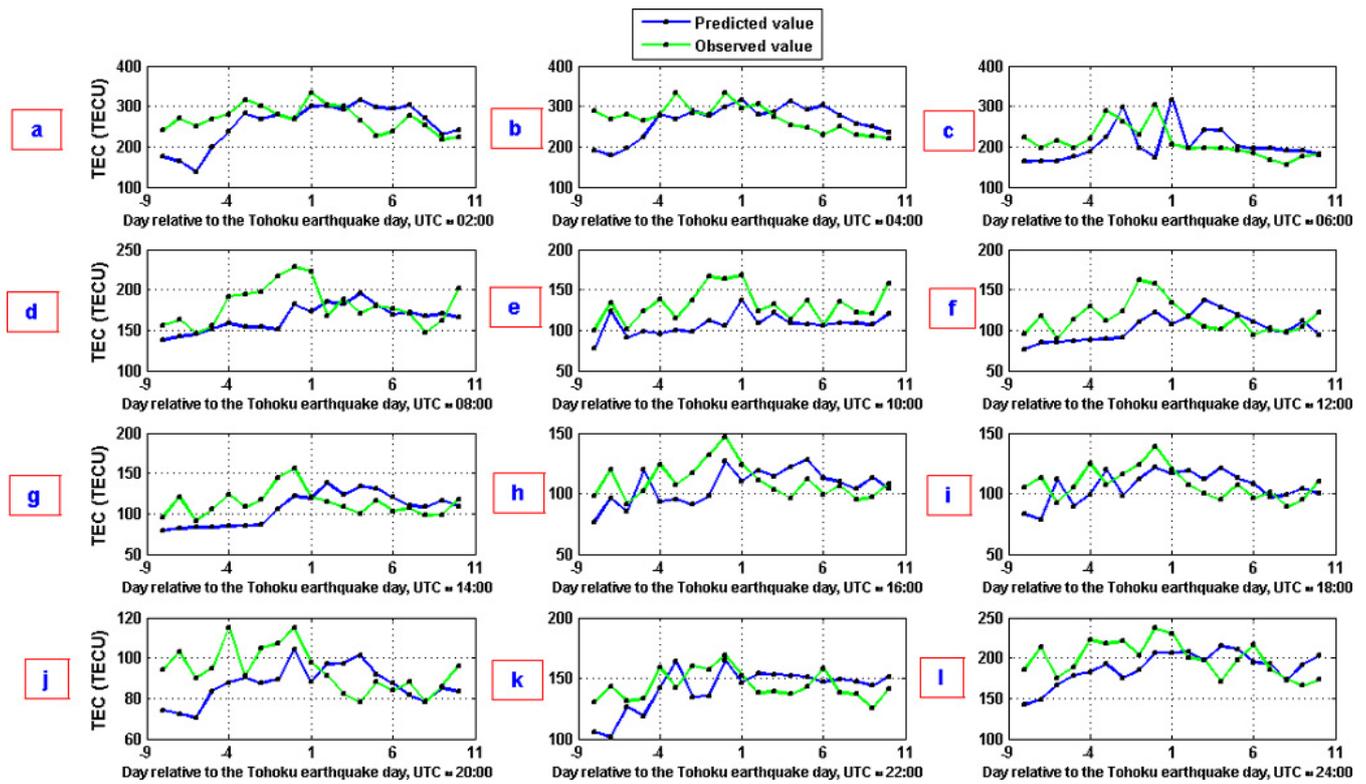


Fig. 2. Variations of the observed (green curve) and predicted (blue curve) TEC values on days selected as testing set at different universal times. The x-axis represents the day relative to the Tohoku earthquake day.

was a result of an M3 flare that took place on 7 March 2011, which released the fastest CME since September 2005 (<http://spaceweather.com/>).

Figure 1d shows TEC variations derived from GPS stations close to the epicenter. By visual inspection (without performing analysis), unusual TEC values are clearly seen around the earthquake day, especially between 9 days before to 8 days after the earthquake.

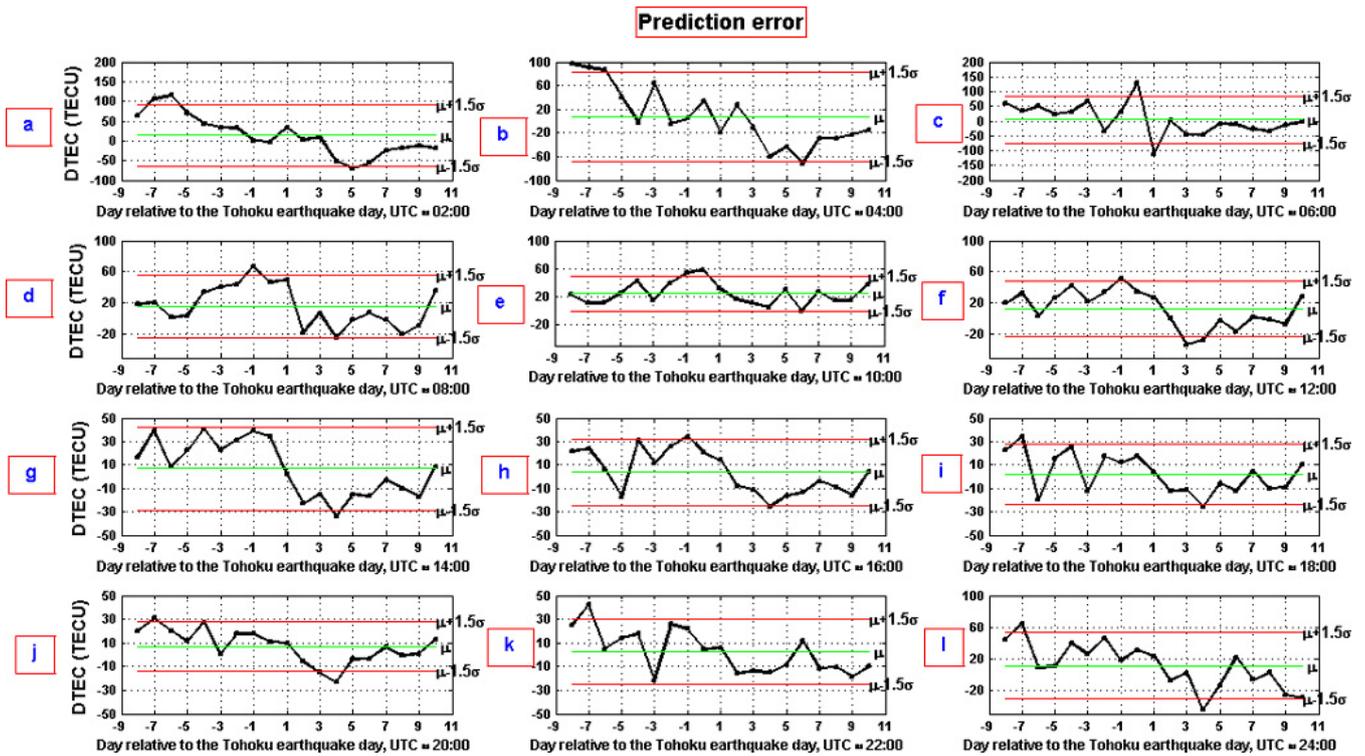
To implement the SVM method, training, validating and testing data were initially set respectively to 40 %, 20 % and 40 % of all TEC data. Green and blue curves in Figs. 2a through l represent the observed and the predicted TEC values using SVM, respectively, during the days selected as the testing set. It can be seen that the differences between these two values reach the noticeable values 6 days before the earthquake at 02:00 UTC, 1 day prior to the earthquake at 08:00, 12:00 and 16:00 UTC and also on earthquake day at 06:00, 08:00, 10:00, 12:00 and 14:00 UTC.

Figures 3a through l represent the differences between the observed and the predicted TEC values during the testing data. Figures 3c, d, e, f and h clearly show the unusual values 1 day prior to earthquake and also on earthquake day.

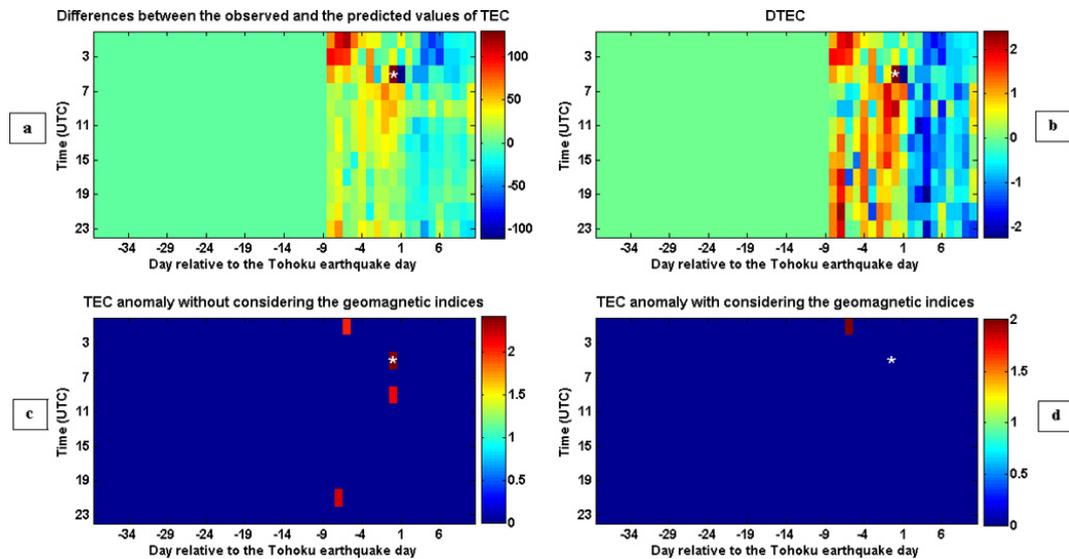
Figures 3c and e indicate that the differences values exceeded the upper bound on earthquake date at 06:00 and 10:00 UTC, respectively. The difference value exceeds the higher bound ( $\mu + 1.5 \times \sigma$ ) on earthquake time (i.e.,

06:00 UTC) with the value of 43 % of the higher bound. In other words, on these hours, the SVM was not able to predict the TEC values based on the model deduced from the training data. Therefore, the observed values at mentioned hours could be considered as earthquake anomaly.

Figure 4a represents the differences between the observed and the predicted TEC values during the days selected as testing set. The high difference value at earthquake time is detected by visual inspection of the Fig. 4a. Figure 4b shows the DTEC values obtained from  $Dx = \frac{x-\mu}{\sigma}$ , where  $x$ ,  $\mu$ ,  $\sigma$  and  $Dx$  are the parameter value, mean value, standard deviation and differential of  $x$ , respectively. According to this, if the absolute value of  $Dx$  were greater than  $k$ , ( $|Dx| > k$ ), the behavior of the relevant parameter ( $x$ ) is regarded as anomalous. In Fig. 4c, anomalous TEC values are only depicted at times when  $|DTEC| > 2.0$ . The DTEC value exceeds the upper bound with the value of 13.0 % and 12.37 %, 6 days before the earthquake at 03:00 UTC, and also the day of the earthquake at 10:00 UTC, respectively. It had also reached to its maximum value (i.e., 42.28 %) at 06:00 UTC on earthquake time (Fig. 4c). Then, to distinguish pre-earthquake anomalies from the other anomalies related to the geomagnetic activities, the four conditions of  $|DTEC| > 2.0$ ,  $Kp < 2.5$ ,  $Dst > -20$  nt and  $F10.7 < 100$  (SFU) are jointly used, using AND operator to construct the anomaly map (Fig. 4d). It is seen that except the anomaly detected 6 days before



**Fig. 3.** Variations of the differences between the observed and the predicted values of TEC on days selected as testing set at different universal times. The red horizontal lines indicate the upper and lower bounds ( $\mu \pm 1.5 \times \sigma$ ). The green horizontal line indicates the mean value ( $\mu$ ). The x-axis represents the day relative to the Tohoku earthquake day.



**Fig. 4.** (a) Differences between the observed and the predicted values of TEC. (b) DTEC variations. (c) Detected anomalies without considering the geomagnetic indices. (d) Detected anomalies with considering the geomagnetic indices.

the earthquake at 03:00 UTC, the other detected anomalies in Fig. 4c have been masked by high geomagnetic activities.

### 3.2 Haiti earthquake

Figure 5a shows the variations of Kp geomagnetic index during the period of 1 December 2009 to 16 January 2010, which is around the Haiti earthquake date. Figure 5b shows

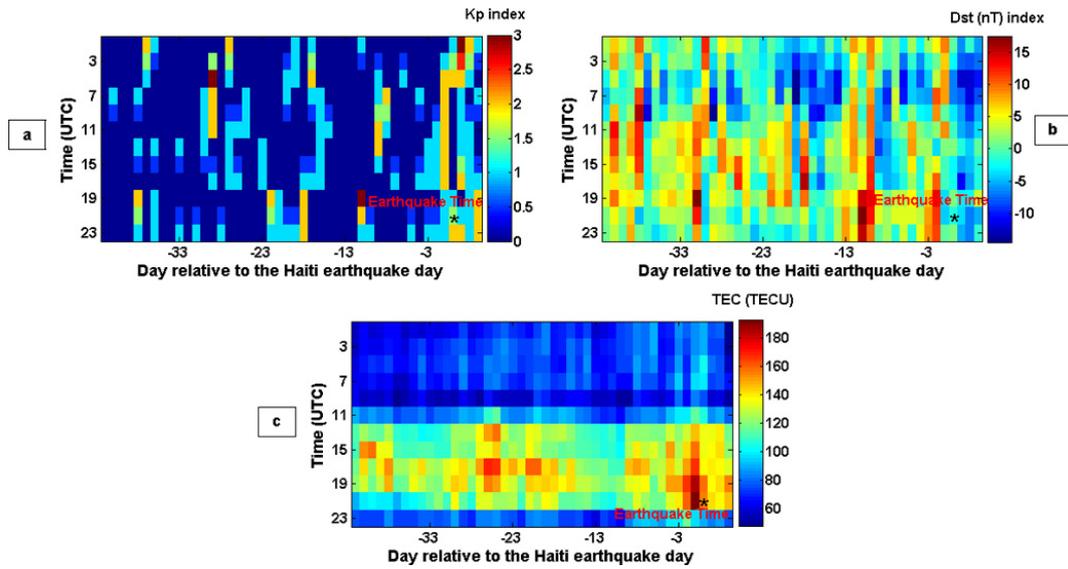


Fig. 5. Data of the Haiti earthquake showing variations of (a) Kp geomagnetic index, (b) Dst geomagnetic index and (c) TEC.

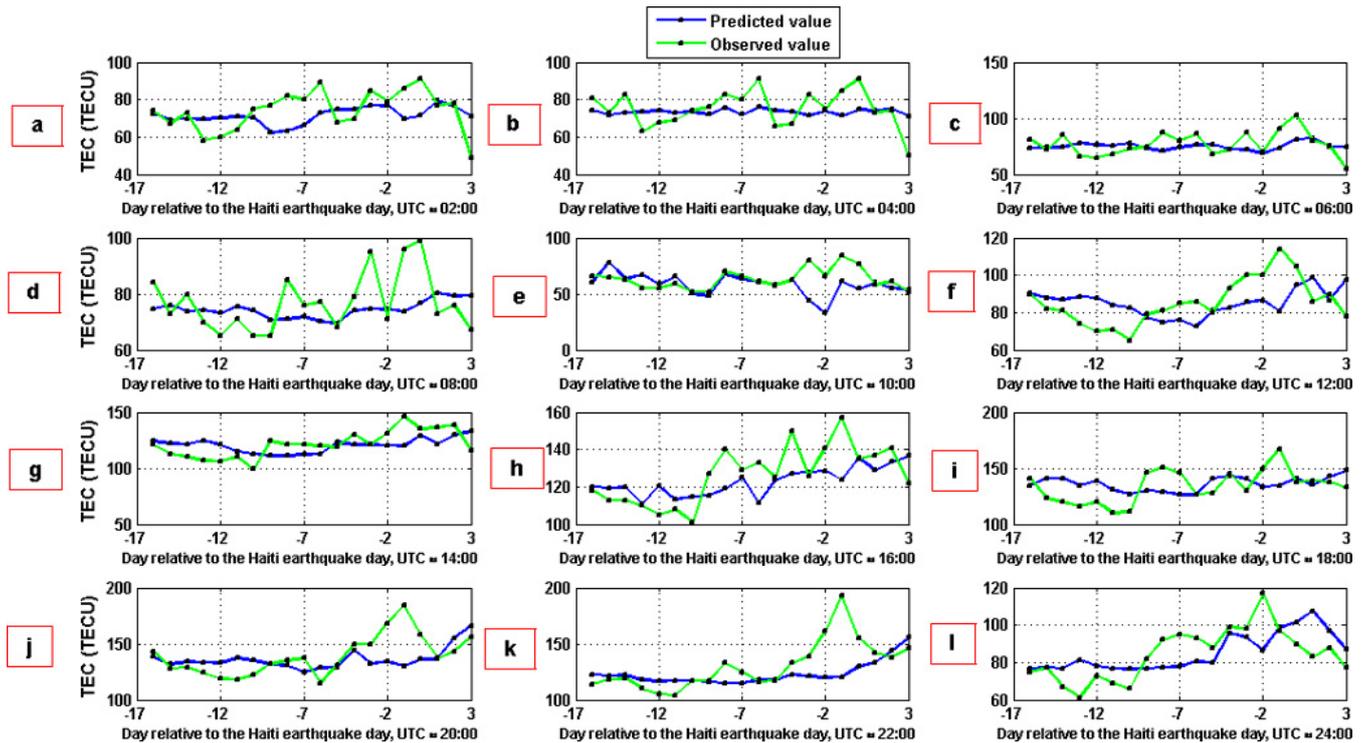


Fig. 6. Variations of the observed (green curve) and predicted (blue curve) TEC values on days selected as testing set at different universal times. The x-axis represents the day relative to the Haiti earthquake day.

variations of Dst geomagnetic index also in the same period. The unusual Kp and Dst values are observed from 10 to 11 days before the earthquake. Figure 5a and b illustrate quiet geomagnetic conditions around the earthquake day. Figure 5c shows TEC variations derived from GPS stations close to

the epicenter. By visual inspection, unusual TEC values are clearly seen around the earthquake day.

When implementing the SVM method, it can be seen that the differences between the observed and predicted TEC values during the testing set reach the unusual values 1, 2 and 3 days before the earthquake (Figs. 6 and 7). The high

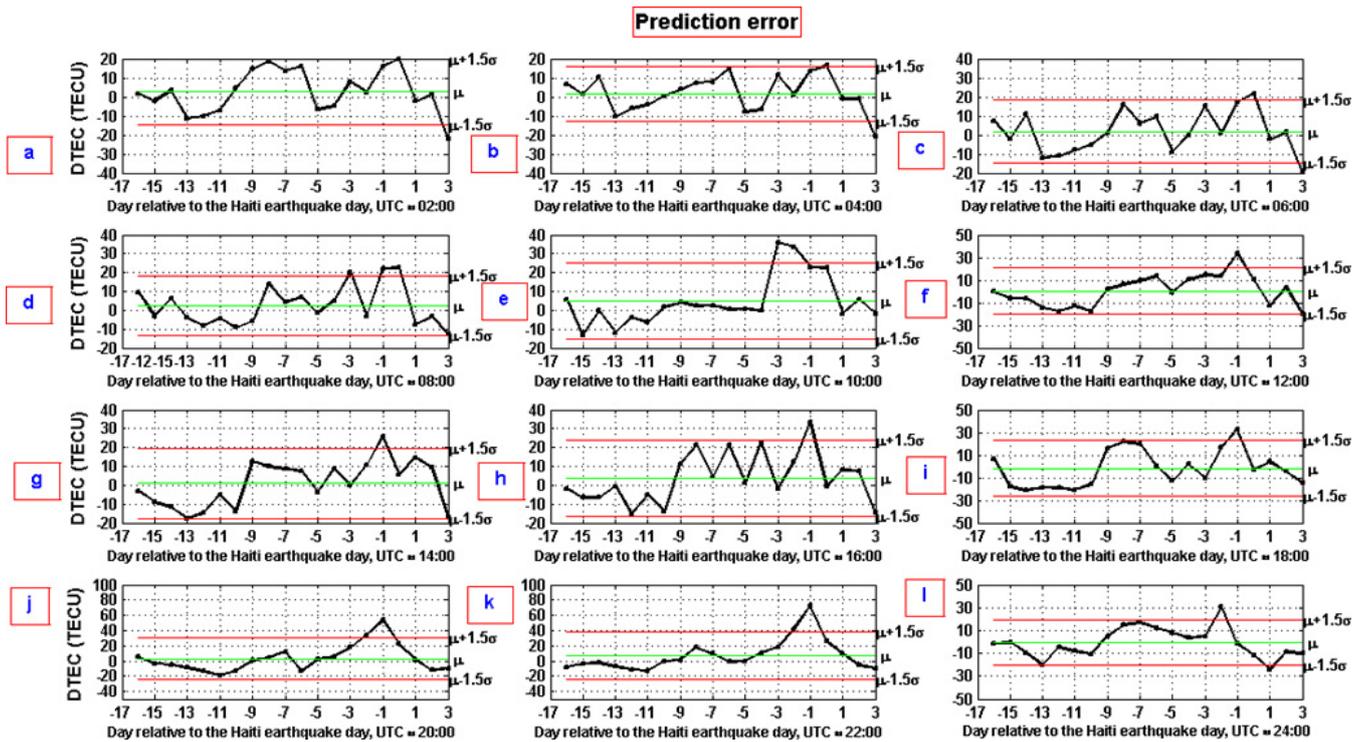


Fig. 7. Variations of the differences between the observed and the predicted values of TEC on days selected as testing set at different universal times. The x-axis represents the day relative to the Haiti earthquake day.

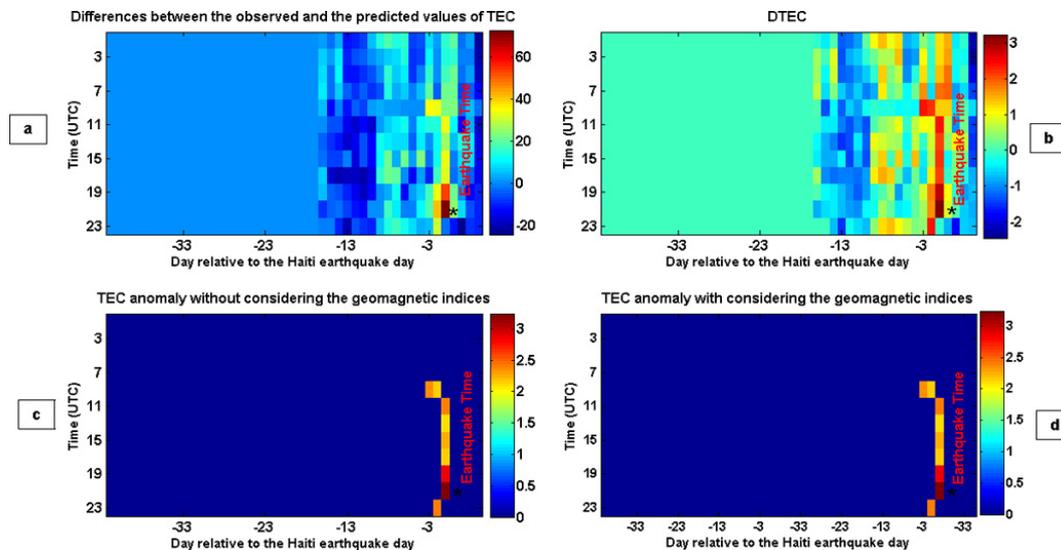


Fig. 8. (a) Differences between the observed and the predicted values of TEC. (b) DTEC variations. (c) Detected anomalies without considering the geomagnetic indices. (d) Detected anomalies with considering the geomagnetic indices.

difference values are detected at 20:00 and 22:00 UTC 1 day before the event (Fig. 8a and b). In Fig. 8c anomalous TEC values are only depicted at times when  $|DTEC| > 1.5$ . The TEC value exceeds the higher bound ( $\mu \pm 1.5 \times \sigma$ ) 1 day prior to the earthquake at 20:00 UTC with the value of 44 % of the higher bound. It had also reached to its maximum value (i.e.,

1.62 times the higher bound) 1 day before the earthquake at 22:00 UTC (Fig. 8c). The DTEC value also exceeds the upper bound with the value of 18.5 % 3 days before the earthquake at 10:00 UTC (Fig. 8c). Due to the quiet geomagnetic conditions around the earthquake date, the Fig. 8d represents the anomalies same as those seen in Fig. 8c.

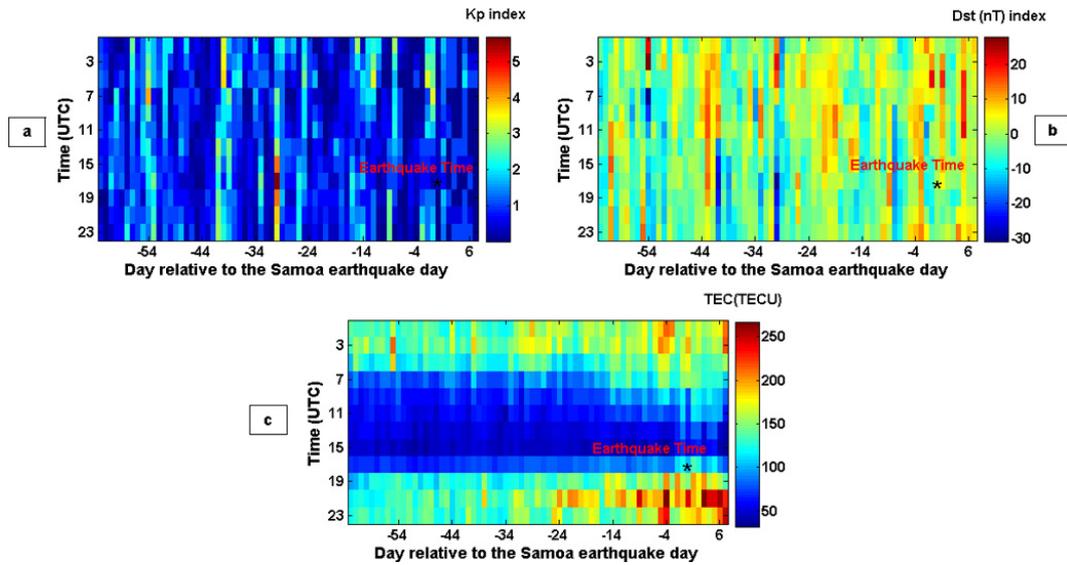


Fig. 9. Data of the Samoa earthquake showing variations of (a) Kp geomagnetic index, (b) Dst geomagnetic index and (c) TEC.

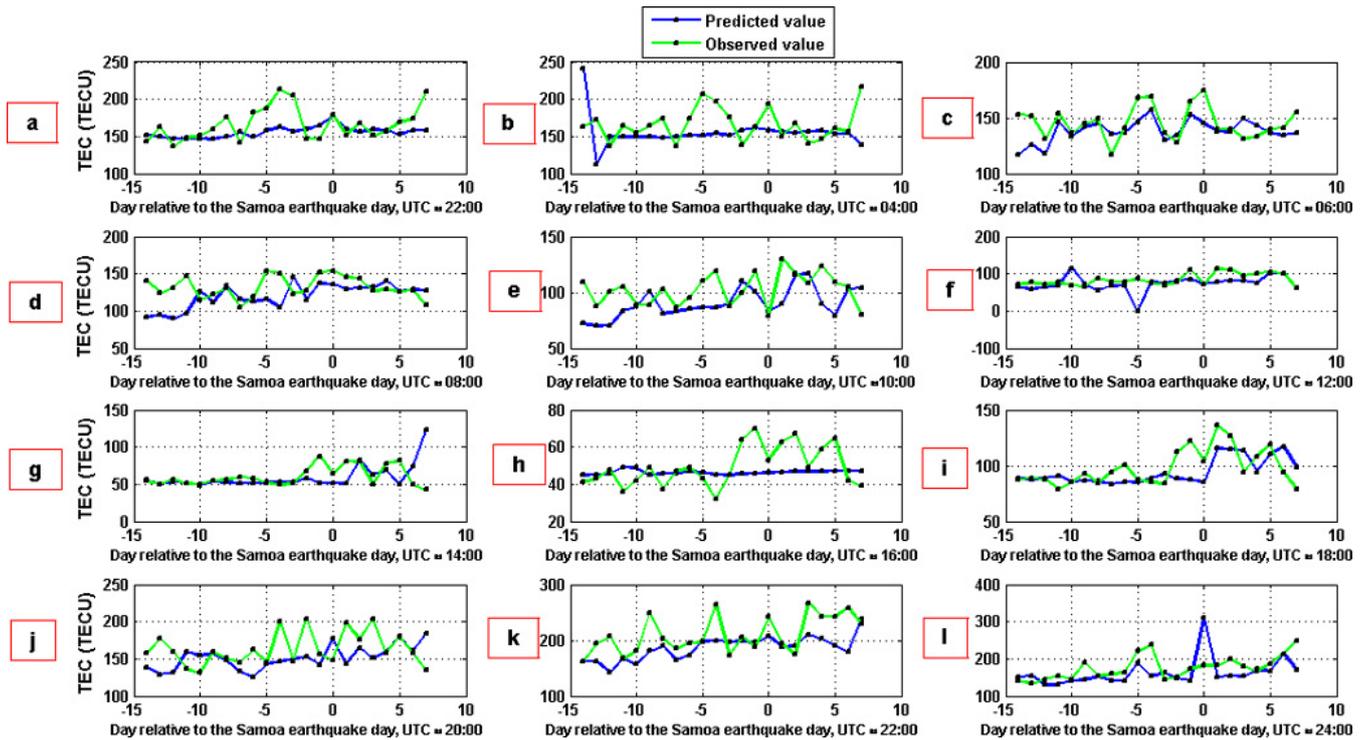


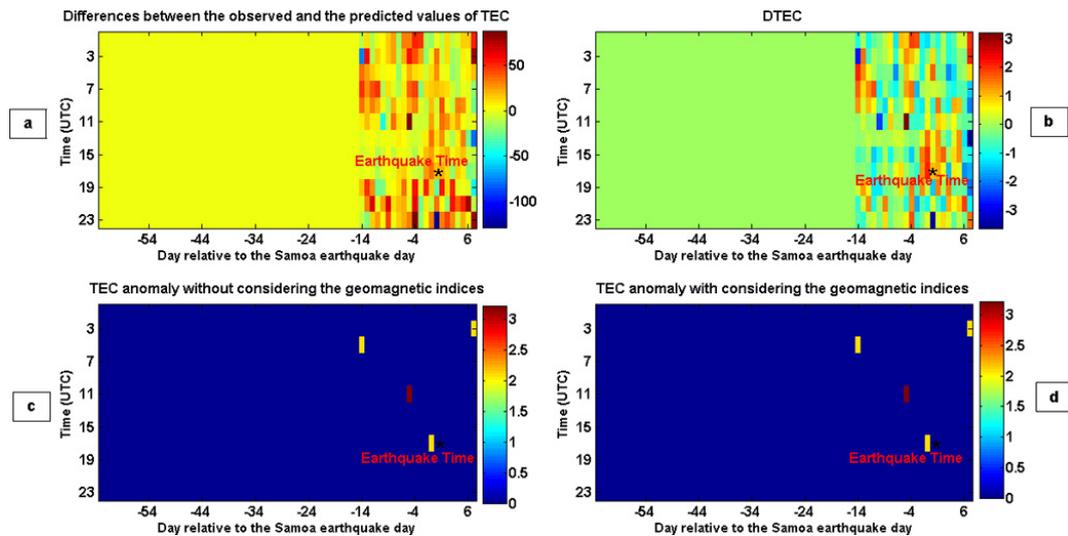
Fig. 10. Variations of the observed (green curve) and predicted (blue curve) TEC values on days selected as testing set at different universal times. The x-axis represents the day relative to the Samoa earthquake day.

### 3.3 Samoa earthquake

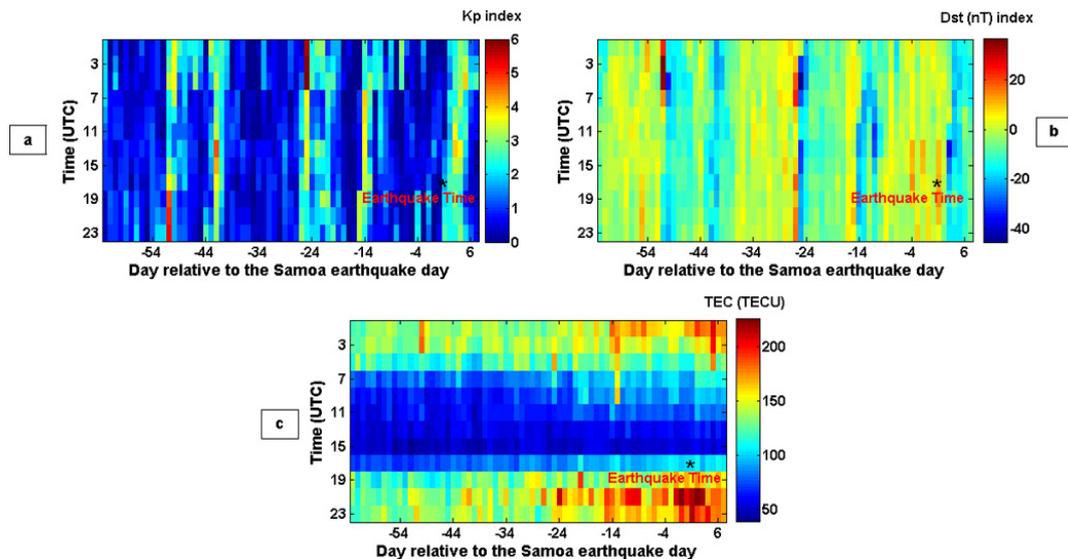
Figure 9c illustrates the TEC values from 28 July to 7 October 2009 (i.e., around the Samoa earthquake date) measured using GPS ground stations close to the epicenter. By visual inspection of TEC variations, perturbations can be seen near to the earthquake date. But to distinguish earthquake pertur-

bations from solar and magnetic disturbances, variations of Dst and Kp indices during the same time interval were incorporated. Figures 9a and b illustrate, respectively, the variations of Kp and Dst indices that were relatively quiet during days prior to the earthquake.

Using SVM method, unusual behaviors are seen in TEC variations when the difference between the observed and



**Fig. 11.** (a) Differences between the observed and the predicted values of TEC. (b) DTEC variations. (c) Detected anomalies without considering the geomagnetic indices. (d) Detected anomalies with considering the geomagnetic indices.



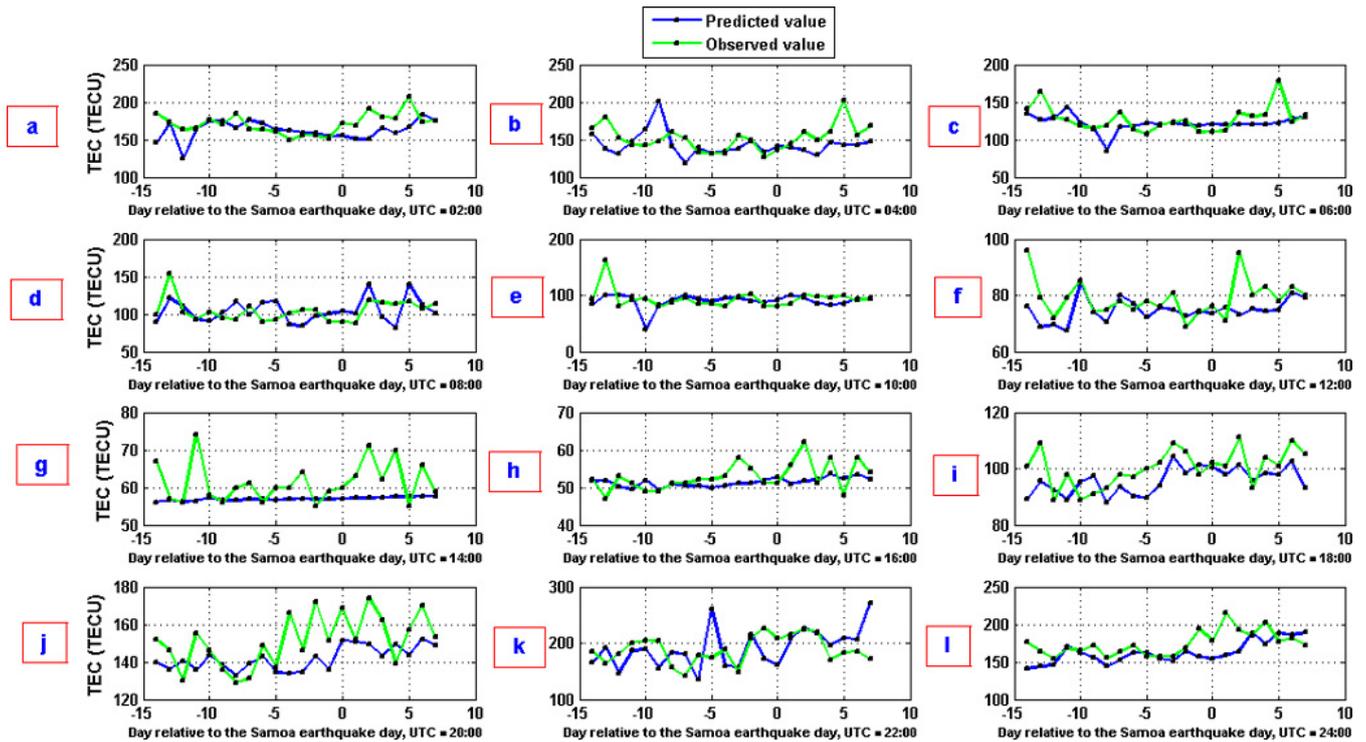
**Fig. 12.** Data of the one year before the Samoa earthquake showing variations of (a) Kp geomagnetic index, (b) Dst geomagnetic index and (c) TEC.

the predicted TEC values reaches maximum value (i.e., 87.7 TECU) at 12:00 UTC 5 days before the earthquake and exceeds the higher bound of the order of 215 % (Figs. 10f and 11a). Figures 11b and c illustrate that the DTEC changes have exceeded the maximum threshold at 18:00 UTC 1 day before the earthquake by the order of 41 %. Then, anomalous TEC is only depicted during quiet geomagnetic conditions, when  $|DTEC| > 1.5$ ,  $Kp < 2.5$  and  $Dst > -20$  (nT). Figure 11d indicates that the mentioned prominent anomalies have occurred during a period of low geomagnetic activities.

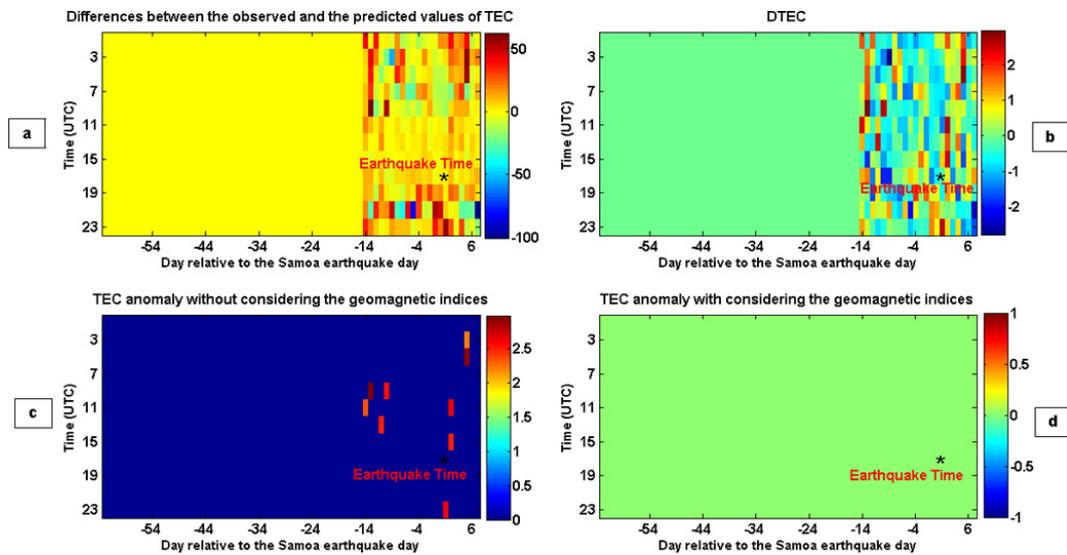
To perform null-hypothesis tests, the same period of data from the previous year in the case of Samoa Islands has

been considered. Figure 12c illustrates the TEC values from 28 July to 7 October 2008 (i.e., about 1 year before the Samoa earthquake). To isolate earthquake perturbations from solar and magnetic disturbances, variations of Dst and Kp indices during the same time interval were incorporated (Figs. 12a and b).

Figure 13 represents the results of SVM analysis for about one year before the Samoa earthquake (29 September 2009) from 28 July to 7 October 2008. Figures 13 and 14 (a, b and c) represent the anomalous TEC variations during 10 to 15 days before the earthquake, but Fig. 14d indicates that the



**Fig. 13.** Variations of the observed (green curve) and predicted (blue curve) TEC values on days selected as testing set at different universal times. The x-axis represents the day relative to the Samoa earthquake day but for the one year before the Samoa earthquake.

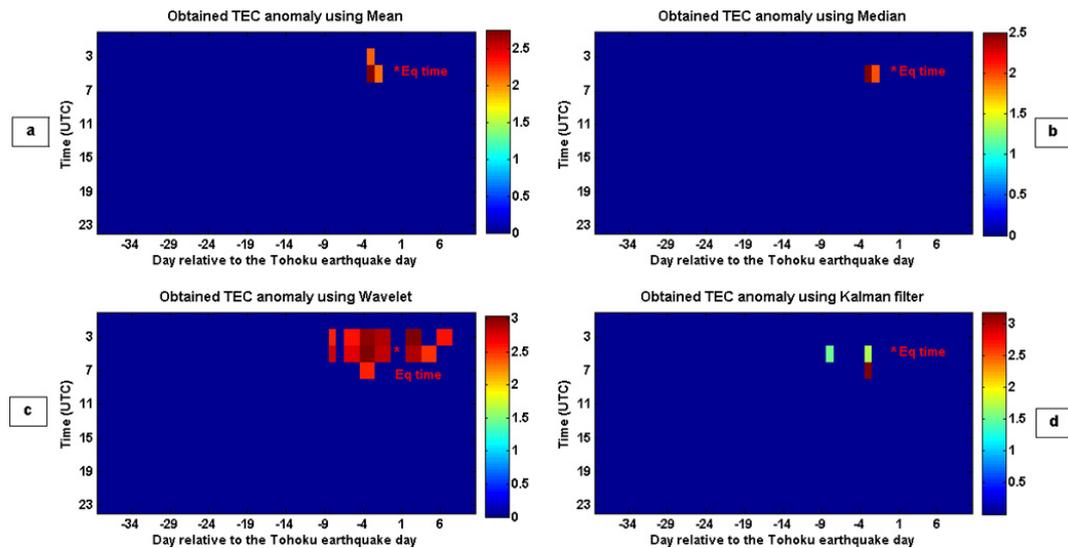


**Fig. 14.** (a) Differences between the observed and the predicted values of TEC. (b) DTEC variations. (c) Detected anomalies without considering the geomagnetic indices. (d) Detected anomalies with considering the geomagnetic indices.

observed anomalies are most likely related to the high geomagnetic activities.

#### 4 Discussions

Already, concerning Tohoku, Haiti and Samoa earthquakes, four anomaly detection methods including mean, median, wavelet and Kalman filter have been evaluated by the author



**Fig. 15.** Result of analysis for the Tohoku earthquake showing detected TEC anomaly using methods (a) mean, (b) median, (c) wavelet transform and (d) Kalman filter during quiet geomagnetic conditions. The x-axis represents the day relative to the Tohoku earthquake day.

to detect the anomalous TEC variations (Akhoondzadeh and Saradjian, 2011; Akhoondzadeh, 2012).

Figure 15a shows detected TEC anomalies using mean method based on:  $|DTEC| > 2.0$ ,  $Kp < 2.5$  and  $Dst > -20$  nt around the Tohoku earthquake date. The TEC value exceeds the higher bound ( $\mu + 2.0 \times \sigma$ ) 3 days prior to the Tohoku earthquake at 04:00 UTC with the value of 7.26% of the higher bound. It had also been reached to its maximum value (i.e., 37.81 %) at 06:00 UTC on the same date. After executing the median method, Fig. 15b shows the TEC anomaly map during quiet geomagnetic conditions. In other words, anomalous TEC values are only depicted at times when  $|DTEC| > 1.5$ ,  $Kp < 2.5$  and  $Dst > -20$  (nT). Figure 15b illustrates an increase (67.29 %) in TEC is clearly observed at 06:00 UTC 3 days before the earthquake. Variations of TEC values clearly exceed the higher bound  $M + 1.5 \times IQR$  (Where  $M$  and  $IQR$  are median value and interquartile range, respectively) of the order of 32.07 % at 06:00 LT 2 days before earthquake. By applying Daubechies 1-D wavelet transform, Figure 15c shows TEC anomalies detected using wavelet transformation from 8 days before to 7 days after the earthquake. The peak of anomaly reaches up to 25.80 % above the threshold value at earthquake time. Figure 15d shows the differences between the predicted TEC values using Kalman filter method and observed TEC values during the 14 days before to 10 days after the earthquake. Based on quiet geomagnetic conditions ( $Kp < 2.5$  and  $Dst > -20$  nt), an unusual increase of TEC (112.33 % above the threshold value) can be seen at 08:00 UTC 3 days before the earthquake (Fig. 15d). Figure 15d also illustrates an increase of 15.75 % from the normal state 3 days before the earthquake at 06:00 UTC (Akhoondzadeh, 2012).

The characteristics of the detected anomalies prior to the Haiti and Samoa earthquakes using the median, wavelet and Kalman filter methods can be found in Tables 2 and 3, respectively (Akhoondzadeh and Saradjian, 2011). It can be seen that there is good agreement between the anomalies detected using the previous implemented methods and SVM. In the case of the Haiti earthquake, each of the four methods detected anomalous TEC variations from 1 to 3 days prior to the earthquake. Also, by applying the four mentioned methods, the prominent unusual variations of TEC are seen 1 and 5 days before the Samoa earthquake.

## 5 Conclusions

Anomaly detection is extremely important for earthquake parameters estimation. In this paper an application of SVM in the earthquake precursor's domain has been developed. Concerning the powerful Tohoku earthquake of 11 March 2011, the results show that the difference between the predicted value obtained from the SVM method and the observed value reaches the maximum value at earthquake time during a period of high geomagnetic activities. The SVM method detected a considerable number of anomalous occurrences 1 and 2 days prior to the Haiti earthquake and also 1 and 5 days before the Samoa earthquake in a period of low geomagnetic activities. The SVM method could be a powerful tool in modeling complex phenomena, such as earthquake precursor time series, that we may not know what the underlying data generating process is. Appropriate selection of the applied parameters such as the number of lagged observations and inner parameters of the kernel function is another challenging task of the SVM modeling. In the domain of earthquake anomaly detection, SVMs could be a good candidate

**Table 2.** Detected anomalies for the Haiti earthquake (12 January 2010) using the median/interquartile, wavelet and Kalman filter methods. Day is relative to the earthquake day. Value calculated by  $p = \pm 100 \times (|Dx| - k) / k$  (Akhoondzadeh and Saradjian, 2011).

Median/interquartile			Wavelet			Kalman filter		
Day	Time	Value %	Day	Time	Value %	Day	Time	Value %
3	5	40	4	17	7	8	5	13
3	19	8	4	19	2	8	7	31
2	17	7	3	15	4	5	17	4
2	19	114	3	17	15	4	9	4
1	3	39	3	19	17	4	13	56
1	5	67	2	17	7	4	17	23
1	7	90	2	19	9	3	7	31
1	15	49				2	15	9
1	17	87				2	17	17
1	19	3				2	17	9
0	1	67				1	3	12
0	3	55				1	5	41
0	5	20				1	7	31
0	7	30				1	9	12
						1	11	5
						1	17	4
						1	19	59

**Table 3.** Detected anomalies for the Samoa earthquake (29 September 2009) using the median/interquartile, wavelet and Kalman filter methods (Akhoondzadeh and Saradjian, 2011).

Median/interquartile			Wavelet			Kalman filter		
Day	Time	Value %	Day	Time	Value %	Day	Time	Value %
4	9	8	8	13	13	7	17	64
4	13	15	6	19	1	5	15	171
2	3	7	5	13	15	4	11	48
2	5	36	5	19	1	4	13	49
2	9	12	4	13	15	2	9	57
1	1	25	1	13	8	2	11	53
1	3	140	0	13	8	1	3	30
1	5	84				1	5	12
1	7	10				1	17	32

because (1) it can model nonlinear relations in an efficient and stable way, (2) a limited set of training points contribute to the solution, and (3) the SVM is trained as a convex optimization problem resulting in a global solution which yields unique solutions. These advantages stem from the specific formulation of a convex objective function with constraints which is solved using Lagrange multipliers and have the characteristics (Thissen et al., 2003). But the training time in SVMs can be large for data sets containing many objects. The best time of tuning can be selected based on an optimization method such as genetic algorithms. In this study the detected TEC anomalies using the SVM method were compared to the previous results (Akhoondzadeh and Saradjian, 2011; Akhoondzadeh, 2012) obtained from mean, median, wavelet and Kalman filter methods. The detected anomalies

using the SVM are similar to those detected using the previous methods. It should be noted that the training phase of the SVM was not feasible with the large data set and therefore in some cases, the SVM performed slightly worse than the neural network (Thissen et al., 2003). Mattera and Haykin (1999) have shown that, in contrast to other models, SVMs perform efficiently due to the use of a nonlinear kernel combined with the use of an  $\epsilon$ -insensitive band which to some extent reduces the effect of the noise.

*Acknowledgements.* The author would like to acknowledge the NASA Jet Propulsion Laboratory for the TEC data and the NOAA for the geomagnetic indices.

Topical Editor K. Kauristie thanks two anonymous referees for their help in evaluating this paper.

## References

- Akhoondzadeh, M.: Comparative study of the earthquake precursors obtained from satellite data. PhD thesis, University of Tehran, Surveying and Geomatics Engineering Department, Remote Sensing Division, 2011.
- Akhoondzadeh, M.: Anomalous TEC variations associated with the powerful Tohoku earthquake of 11 March 2011, *Nat. Hazards Earth Syst. Sci.*, 12, 1453–1462, doi:10.5194/nhess-12-1453-2012, 2012.
- Akhoondzadeh, M. and Saradjian, M. R.: TEC variation analysis concerning Haiti (January 12, 2010) and Samoa (September 29, 2009) earthquakes, *Adv. Space Res.*, 47, 94–104, doi:10.1016/j.asr.2010.07.024, 2011.
- Burges, C. J. C.: A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2, 121–167, 1998.
- Hayakawa, M. and Molchanov, O. A.: *Seismo- Electromagnetics: Lithosphere-Atmosphere-Ionosphere Coupling*, Terra Scientific Publishing Co. Tokyo, pp. 1–477, 2002.
- He, L., Wu, L., Pulinets, S., Liu, S., and Yang, F.: A nonlinear background removal method for seismo-ionospheric anomaly analysis under a complex solar activity scenario: A case study of the M9.0 Tohoku earthquake, *Adv. Space Res.*, 50, 211–220, 2012.
- Klimenko, M., Klimenko, V., Zakharenkova, I., Pulinets, S., Zhao, B., and Tsidilina, M.: Formation mechanism of great positive TEC disturbances prior to Wenchuan earthquake on May 12, 2008, *Adv. Space Res.*, 48, 488–499, 2011.
- Liu, J. Y., Chuo, Y. J., Shan, S. J., Tsai, Y. B., Chen, Y. I., Pulinets, S. A., and Yu, S. B.: Pre-earthquake ionospheric anomalies registered by continuous GPS TEC measurements, *Ann. Geophys.*, 22, 1585–1593, doi:10.5194/angeo-22-1585-2004, 2004.
- Mattera, D. and Haykin, S.: Support vector machines for dynamic reconstruction of a chaotic system, *Advances in Kernel Methods Support Vector Learning*, MIT Press, Cambridge, 243–254, 1999.
- Mayaud, P. N.: Derivation, Meaning and use of geomagnetic indices, *Geophy.*, 22, American Geo. Union, Washington, D.C., 1980.
- Muller, K. R., Smola, A., Ratsch, G., Scholkopf, B., Kohlmorgen, J., and Vapnik, V.: Predicting time series with support vector machines, *Proceedings of ICANN '97*, Springer LNCS 1327, Berlin, 999–1004, 1997.
- Namgaladze, A. A., Klimenko, M. V., Klimenko, V. V., and Zakharenkova, I. E.: Physical mechanism and mathematical modelling of earthquake ionospheric precursors registered in Total Electron Content, *Geomagnetism and Aeronomy*, 49, 252–262, 2009.
- Pulinets, S. A.: Physical mechanism of the vertical electric field generation over active tectonic faults, *Adv. Space Res.*, 44, 767–773, 2009.
- Pulinets, S. A. and Boyarchuk, K. A.: *Ionospheric Precursors of Earthquakes*, Springer, Berlin, 2004.
- Thissen, U., Brakel, R. V., de Weijer, A. P., Melssen, W. J., and Buydens, L. M. C.: Using support vector machines for time series prediction, *Chemometrics and Intelligent Laboratory Systems*, 69, 35–49, 2003.
- Xu, T., Chen, Z., Li, C., Wu, J., Hu, Y., and Wu, Z.: GPS total electron content and surface latent heat flux variations before the 11 March 2011 M9.0 Sendai earthquake, *Adv. Space Res.*, 48, 1311–1317, 2011.