**Natural Hazards and Earth System Sciences**

# A comparison of classical and intelligent methods to detect potential thermal anomalies before the 11 August 2012 Varzeghan, Iran, earthquake ($M_{\mathrm{w}} = 6.4$)

**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)

**Abstract.** In this paper, a number of classical and intelligent methods, including interquartile, autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and support vector machine (SVM), have been proposed to quantify potential thermal anomalies around the time of the 11 August 2012 Varzeghan, Iran, earthquake ($M_{\mathrm{w}} = 6.4$). The duration of the data set, which is comprised of Aqua-MODIS land surface temperature (LST) night-time snapshot images, is 62 days. In order to quantify variations of LST data obtained from satellite images, the air temperature (AT) data derived from the meteorological station close to the earthquake epicenter has been taken into account. For the models examined here, results indicate the following: (i) ARIMA models, which are the most widely used in the time series community for short-term forecasting, are quickly and easily implemented, and can efficiently act through linear solutions. (ii) A multilayer perceptron (MLP) feed-forward neural network can be a suitable non-parametric method to detect the anomalous changes of a non-linear time series such as variations of LST. (iii) Since SVMs are often used due to their many advantages for classification and regression tasks, it can be shown that, if the difference between the predicted value using the SVM method and the observed value exceeds the pre-defined threshold value, then the observed value could be regarded as an anomaly. (iv) ANN and SVM methods could be powerful tools in modeling complex phenomena such as earthquake precursor time series where we may not know what the underlying data generating process is. There is good agreement in the results obtained from the different methods for quantifying potential anomalies in a given LST time series. This paper indicates that the detection of the potential thermal anomalies derive credibility from the overall efficiencies and potentialities of the four integrated methods.

## 1 Introduction

Land surface temperature (LST) anomalies may arise due to different factors such as climatological conditions, forest fire, geothermal phenomena, earth's warming, and so on. Some researchers have professed that observed pre-seismic thermal anomalies are precursory activities and others have not. A pre-seismic thermal anomaly is an abnormal increase in LST that may be observed around 1–10 days prior to an earthquake, with increases of temperature on the order of 3–12 °C or more, and usually disappears a few days after the event (Qiang et al., 1991). Blackett et al. (2011) found no convincing evidence of LST precursors to the 2001 Gujarat earthquake, and urged care in the use of approaches aimed at identifying such seismic thermal anomalies. This idea that the occurrence of strong earthquakes may be accompanied by thermal anomalies originated in Russia, China and Japan (Tronin, 1996). In 1980, Russian researchers detected some short-lived thermal infrared anomalies from satellite images prior to an earthquake in central Asia (Tronin, 1996). Then, other researchers reported a number of observations on thermal anomalies before strong earthquakes (Qiang et al., 1991, 1999; Tronin et al., 2000; Ouzounov and Freund, 2004; Pulinets et al., 2006; Akhoondzadeh, 2011).
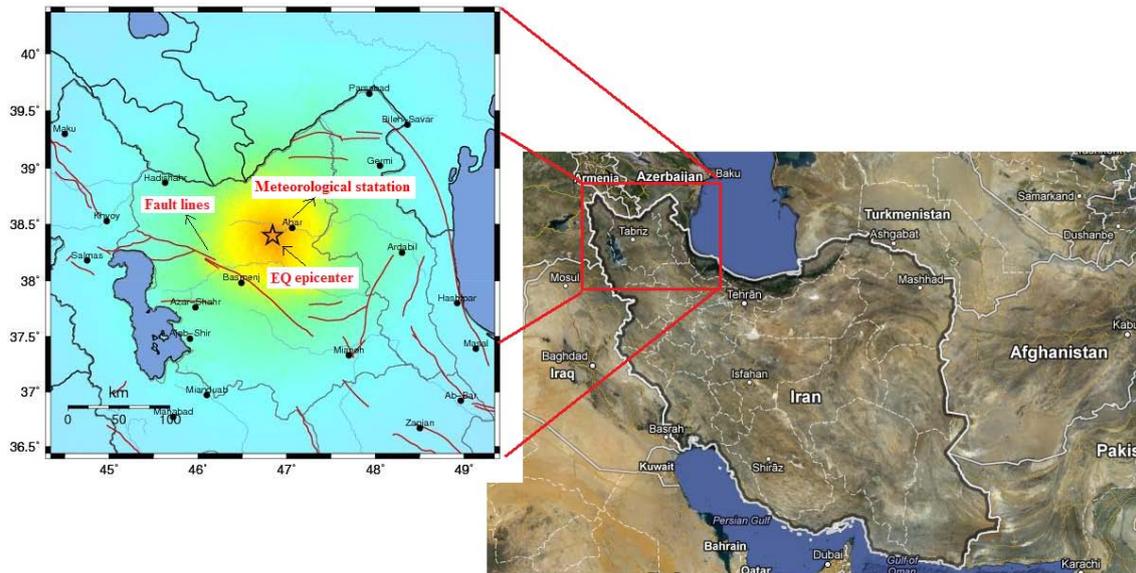
**Fig. 1.** The geographic location of the Varzeghan earthquake epicenter in NW Iran. A black circle close to the epicenter represents the Ahar meteorological station. The locations of the major faults are noted as red graphic overlays. (Figure from www.iiees.ac.ir/).

Proper detection of anomalous phenomena in a non-linear time series could be done using the non-parametric mathematical methods implementing pattern recognition independent of physical processes. In this paper several classical and intelligent techniques, including interquartile, autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and support vector machine (SVM), have been compared for potentially pulling out the LST anomalies, but without stating that LST anomalies are always followed by earthquakes (and indeed, sometimes they are not).

It should be noted that the potential thermal anomalies may have origins other than earthquakes. In the case of earthquakes, emission of gases such as methane, carbon dioxide and hydrogen due to the opening and closure of micropores upon induced stresses and also the changes of ground water regime have been offered as a possible cause for generation of thermal anomalies (Qiang et al., 1991; Tronin, 1996). Freund et al. (2007, 2009) have described using a laboratory experiment that stimulated infrared (IR) emission due to hole–hole recombination, and its follow-on effects may help us to understand the enhanced IR emission seen in night-time satellite images of the land surface before some major earthquakes.

A number of stationary and sun-synchronous remote sensing satellites can sense the radiation coming up from the earth in thermal bands and might provide useful information prior the earthquakes. Due to their suitable temporal, spatial and spectral resolutions, thermal infrared bands of NOAA-AVHRR, Terra-MODIS, and Meteosat-5 data have been used in most of the recent studies connected with thermal anomaly. In this study, LST variations close to the studied earthquake epicenter have been analyzed using the LST night-time images of Aqua-MODIS provided by NASA (http://modis.gsfc.nasa.gov/data). These data are generated on a daily basis at a temperature resolution of 0.02 °C. Each pixel of a LST image covers an area of $1 \times 1\,km^2$ on the ground.

In order to quantify the variations of LST data obtained from satellite images, this study applied the air temperature (AT) data downloaded from the website: (http://www.wunderground.com/). These data have been collected by the Ahar meteorological station (38.4° N, 47.1° E) close to the studied earthquake epicenter (Fig. 1).

The selected case study is an earthquake which occurred in NW Iran (Fig. 1), 60 km NE Tabriz, Ahar region (38.33° N, 46.83° E), with a magnitude of $M_w = 6.4$ on 11 August 2012 at 16:53:18 LT (UTC = LT−04:30). This earthquake caused 306 people to be killed, 3037 to be injured and 30 000 to become homeless.

## 2 Methodology

Daily variations of the land surface temperature depend on season, geographic location, climatological conditions and other unknown parameters. The unknown variations preclude the possibility of using methods based on normal distribution of data for anomaly detection.

ARIMA, ANN and SVM methods could be powerful tools in modeling complex phenomena such as LST time series, even though we may not know what the underlying data generating process is.

ARIMA models are the most popular class of models for forecasting of a time series where the future value of a

variable is assumed to be a linear function of several past observations and random errors.

Artificial neural networks are a class of intelligent systems that can discover patterns with a few a priori assumptions and can learn any complex functional relationship from the data to model a phenomenon. Neural networks are suitable tools to capture the autocorrelation structure in a time series even if the underlying law governing the series is unknown or too complex to describe (Faraway and Chatfield, 1995). A large number of successful applications have shown that neural networks can be a very effective tool in modeling and forecasting of non-linear time series. The most popular and successful model is the feed-forward multilayer perceptron (MLP) network. In a MLP, neurons are grouped in layers and only forward connections are included.

In order to detect an anomaly, the total available data are split into a training set and a test set. The training set is used for construction of the neural network, whereas the test set is used for measuring the predictive error of the model. In order to determine the best network configuration, the effective parameters, which influence the value of predictive error, including the number of pattern input, lag value: the number of hidden layers and their number of neurons, the activation functions and the learning algorithm, have been obtained via an iterative process to assess the minimum predictive error when the training process was implemented.

To start the prediction process, $N$ observations, $y_1, y_2,\ldots, y_N$, are selected as the training set and the remaining ones, $y_{N+1}, y_{N+2},\ldots, y_{N+m}$, are considered as the test set. The number of input nodes corresponds to the number of lagged observations used to discover the underlying pattern in a time series. Different input nodes can affect either the learning or predictive capability of the network. In this study a network with three nodes in input layer, two nodes in hidden layer and one node in output layer has been proposed. In other words, each four observations in the training set constitute a pattern vector, which three of them are input values and the last one is output value.

The training patterns in the proposed network are

$$X_4 = f(X_1, X_2, X_3) \tag{1}$$
$$X_5 = f(X_2, X_3, X_4)$$
$$\ldots$$
$$X_N = f(X_{N-3}, X_{N-2}, X_{N-1})$$

The training process is executed to find the optimized connection weights such that the prediction error (PE) is minimized. The PE equation can be written as

$$\mathrm{PE} = \sum_{i=4}^{N} (X_i - \hat{X}_i), \tag{2}$$

where, $\hat{X}_i$ is the output from the network.

The testing patterns are

$$X_{N+4} = f(X_{N+1}, X_{N+2}, X_{N+3}) \tag{3}$$
$$X_{N+5} = f(X_{N+2}, X_{N+3}, X_{N+4})$$
$$\ldots$$
$$X_{N+m} = f(X_{N+m-3}, X_{N+m-2}, X_{N+m-1})$$

In the case of the testing process, if the difference value $DX_i$ between the actual value $X_i$ and the predicted value $\hat{X}_i$ is outside the pre-defined bounds ($M \pm IQR$), the anomaly is detected.

SVMs have been applied for classification and regression tasks, but their principles can be extended feasibly to the time series forecasting (Cao and Tay, 2001; Muller et al., 2007).

To start prediction process by SVMs, one input object ($x_i$) to the SVM is a time series of consecutive measurements: $x_i = \{x(t_i), x(t_i - s),\ldots, x(t_i - \tau s)\}$, 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 30 %, 20 % and 50 % of all LST data. The parameter of $\tau$ is selected as 2.

At each step, using the training data, the SVM method is implemented and then the predictive error (Eq. 2) is minimized during the validation of data, where $X_i$ and $\hat{X}_i$ in Eq. (2) are the observed value and the output from the SVM method, respectively.

Finally, the LST value is predicted and then 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, while the size of the testing set decreases. The SVM algorithm is executed, and again the LST value is estimated at the next time. The process is repeated until the all of the LST 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 pre-defined bounds, $M \pm IQR$, the anomaly is detected.

## 3 Implementation

The blue line in Fig. 2a illustrates the time series of AT data close to the epicenter during the period of 1 July to 31 August 2012. It should be mentioned that August is usually a dry period with little or no cloud cover and rain. The mean of AT data during the period of 1 July to 31 August from 2007 to
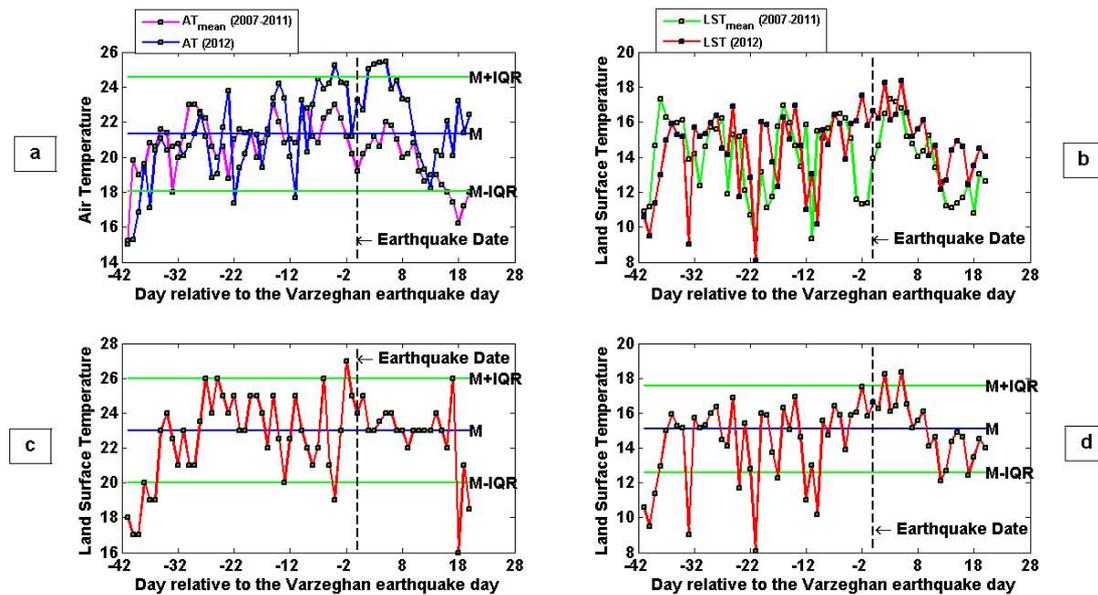
**Fig. 2. (a)** Results of air temperature data analysis for the Varzeghan earthquake (11 August 2012) from 1 July to 31 August 2012. The blue line with open squares represents the AT variations in 2012. The mean of AT during the same period from 2007 to 2011 is represented as a magenta line with open squares. **(b)** Same as **(a)** but for land surface temperature data. **(c)** The variations of the averaged LST of pixels along the fault line close to the epicenter from 01 July to 31 August 2012. **(d)** LST variations of epicenter obtained from night-time MODIS satellite images. The x-axis represents the day relative to the earthquake day. Median, and higher and lower bounds ($M \pm IQR$) are seen as blue and green horizontal lines, respectively. The earthquake day is represented as a vertical dotted line.

2011 is shown as a magenta line (Fig. 2a). The x-axis represents the days relative to the earthquake day. The vertical dotted line shows the earthquake date. Median, and higher and lower bounds ($M \pm IQR$) are seen as blue and green horizontal lines, respectively. The interquartile method has been used to find the signal fluctuations beyond the lower and higher bounds. When implementing the interquartile method, the AT value exceeds the higher bound ($M + IQR$) 4 days before the earthquake and reaches its maximum value of 33.33 % of the higher bound (Fig. 2a). These unusual variations are also seen 2, 3, 4 and 5 days after the event. It can be seen that the difference values between the AT in 2012 (blue line) and the averaged AT during the 5 yr preceding the earthquake year (magenta line) reach maximum values from 8 days before to 9 days after the earthquake.

The red line in Fig. 2b illustrates the time series of LST data averaged over a $5 \times 5$ pixel area centered on the earthquake epicenter during the period of 1 July to 31 August 2012. The mean LST data during the period of 1 July to 31 August from 2007 to 2011 years is shown as a green line. It can be seen that the difference values between the LST in 2012 and the averaged LST during the same period from 2007 to 2011 are noticeable in period of 7 to 13 August 2012.

A description of the robust satellite-based thermal infrared (RST) method and its implementation can be found in the paper of Filizzola et al. (2004). Using the RST approach, the

following index is computed:

$$\otimes_V (r,t) = \frac{V(r,t) - \mu_V(r)}{\sigma_V(r)}, \tag{4}$$

where $r$ represents location coordinates on a satellite image, $t$ is the acquisition time of image, and $V(r,t)$ is the value of a variable $V$ (i.e. LST) at location $r$ and at acquisition time $t$. $\mu_V(r)$ and $\sigma_V(r)$ are the time average value and the standard deviation of LST at location $r$, respectively.

Using the RST method, the LST anomalies maps were obtained (Fig. 3). According to Eq. (4), if the absolute value of $\otimes_V (r,t)$ would be greater than $k$ ($|\otimes_V (r,t)| > k$), the behaviour of the relevant parameter ($V$) is regarded as anomalous. Therefore, in Fig. 3 all the pixel values greater than 1.5 ($|\otimes_V (r,t)| > 1.5$) could be considered as anomalies. We see a positive temperature excursion with a maximum of $\otimes_V (r,t) = 2.25$ ($\sim 8\,°C$) compared to the background LST from 6 to 9 August 2012 for the area close to the epicenter. They develop rather rapidly five days before the earthquake and disappear within a few days after the event. In each panels of Fig. 3, a red line illustrates the nearest fault line to the epicenter. The average of LST values of pixels along the fault line for each image were computed, and then the variations of the mentioned LST values during the period of 1 July to 31 August 2012 were obtained (Fig. 2c). After applying the interquartile method, variations of LST values exceed the upper bound ($M + IQR$) on the order of 21.12 %, 2 days before event.
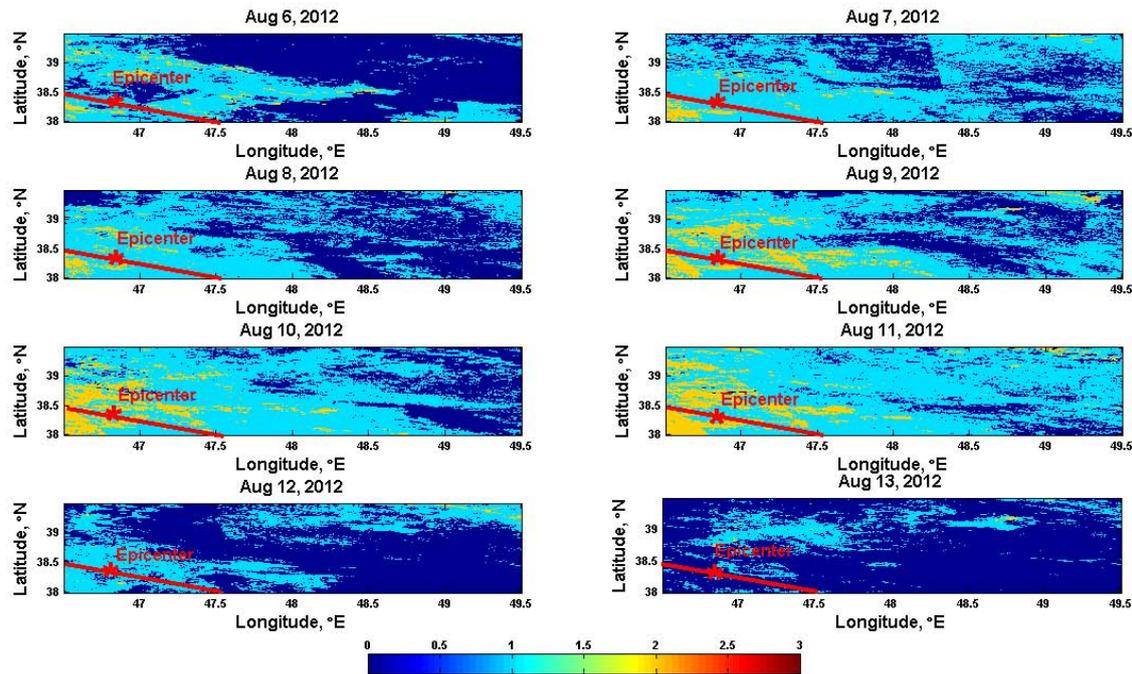
**Fig. 3.** Night-time land surface temperature images obtained from the RST (robust satellite-based thermal infrared) method during the period of 6 to 13 August 2012. Red asterisk and curve represent the epicenter and fault line, respectively.

Figure 2d represents the variations of LST data close to the Varzeghan earthquake epicenter during the period of 1 July to 31 August 2012. After applying the interquartile method, variations of LST exceed the upper bound ($M + IQR$) on the order of 0.02 % 2 days before the event. An unusual increase in LST values is also observed with the value of 26.10 % and 30.92 % of the higher bound 2 and 5 days after the event, respectively. It should be noted that aftershocks with magnitudes of $M_{\mathrm{w}} = 4.4$ and 4.6 occurred 1 day and 2 days after the earthquake.

To implement the ARIMA method, training data were set to 50 % of all data. In the other words, using the ARIMA method, the best fitted model to 50 % of data was selected and then applied to predict the future value during the test data set. In the case of the testing process, if the difference value $DX_i$ between the actual value $X_i$ and the predicted value $\hat{X}_i$ is outside the pre-defined bounds, $M \pm IQR$, the anomaly is detected. Red and green lines in Fig. 4a represent the observed and the predicted LST values using the ARIMA method, respectively, during the days selected as testing set. It can be seen that the predicted values vary around the mean value of the test set. Figure 4b represents the differences between the observed and the predicted LST values using the ARIMA method for the testing data. The results illustrate that the prediction error increases during the test set and reaches the maximum value 2 days prior to earthquake. The prediction error exceeds the higher bound ($M + IQR$) with the value of 14 %. The $DX$ value also exceeds the higher bound 2 days after the earthquake with the value of 26.76 % of the higher bound. It should be noted that since the predicted values obtained from ARIMA model move around the mean of data, therefore it can not be a very good candidate for strictly non-linear time series forecasting.

To implement the neural network method, training and testing data were set respectively to 40 % and 60 % of all data. Red and green lines in Fig. 4c represent the observed and the predicted LST values using the MLP neural network, respectively, during the days selected as the testing set. Figure 4d represents the differences between the observed and the predicted LST values during the testing data. This figure clearly shows the unusual differences 2 days before and 2 days after the earthquake. The value of the prediction error exceeds the higher bound ($M + IQR$) 2 days prior to the earthquake with the value of 12.20 % of the higher bound. 2 days after the earthquake, the $DX$ value passes the higher bound with the value of 23.95 %. In other words, on these days, the MLP neural network was unable to predict the LST values based on the model deduced from the training data. Therefore, the observed values on these mentioned days could be considered as anomalies.

When implementing the SVM method, training, validating and testing data were initially set respectively to 35 %, 15 % and 50 % of all LST data. Red and green lines in Fig. 4e represent the observed and the predicted LST values using SVM during the days selected as the testing set. It can be seen that the differences between these two values reach noticeable values 3, 4 and 6 days before and 3 days after the earthquake. Figure 4f illustrates the differences between the observed and
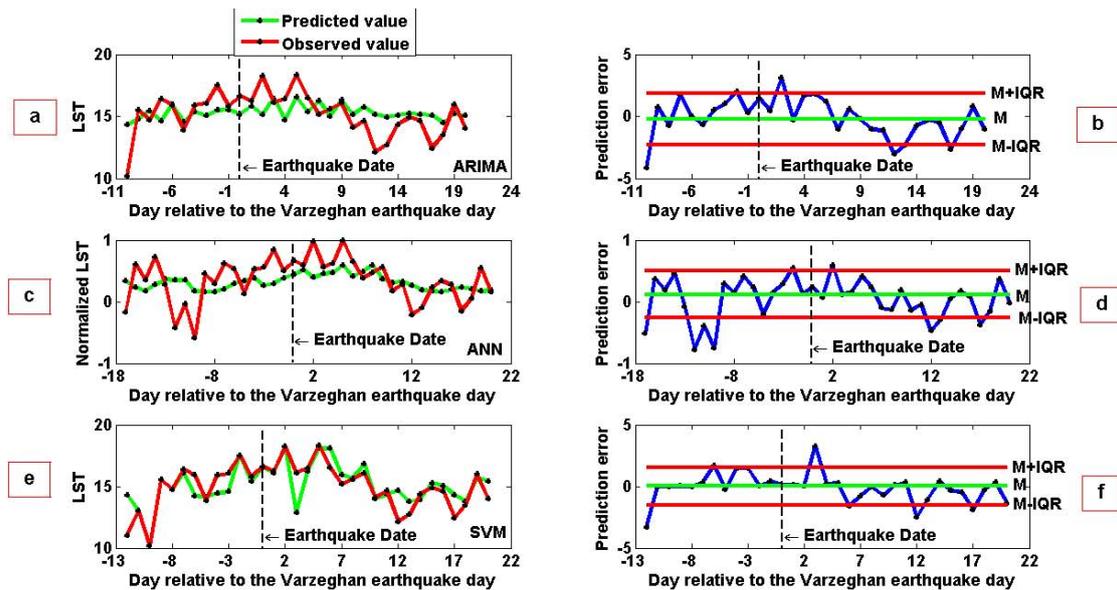
**Fig. 4. (a)**, **(c)** and **(e)** represent the variations of the observed (red line with open squares) and the predicted (green line with open squares) LST values obtained from ARIMA, ANN and SVM methods, respectively, during days selected as testing set. **(b)**, **(d)** and **(f)** represent the variations of the differences between the observed and the predicted values of LST obtained from ARIMA, ANN and SVM methods on days selected as testing set. The earthquake day is represented as a vertical dotted line. The red horizontal lines indicate the upper and lower bounds ($M \pm IQR$). The green horizontal line indicates the median value ($M$). The x-axis represents the day relative to the earthquake day.

the predicted LST values for the testing data. This figure indicates that the difference values reach the higher bound 3 days after the earthquake. The value of prediction error exceeds the upper bound with the value of 108.3 % 3 days after the main event. In other words, on these days, the SVM was unable to predict the LST values based on the model deduced from the training data. Therefore, the observed values at mentioned days could be considered as anomalies.

## 4 Discussion and conclusions

In this study, night-time LST time series have been analyzed to locate relevant anomalous variations prior to the Varzeghan (12 August 2012) earthquake. The detected anomalies in LST variations derived from MODIS satellite data were quantified using the observed anomalies in AT variations obtained from the meteorological station close to the earthquake epicenter.

It is seen that each method has to some extent different lead times for the anomalies even when using the same data source of LST values. These differences in the results might be related to the intrinsic differences of each method. The interquartile method detects any unusual variations falling outside the predefined bounds. Therefore, this method may exclude non-seismic signals. A traditional method such as ARIMA is easily and quickly implemented, but it efficiently acts for linear solutions. In other words, traditional methods assume that the future value of a time series is linearly re-

lated to the past observations. Therefore, they can not be a very suitable tool to model a non-linear time series such as LST variations. Contrary to these classical models, neural networks are able to identify the patterns from noisy data and hence give better forecasts (Zhang, 2001). The ANN method appreciably detects any unpredictable values of the studied time series. The efficiency of the ANN method in proper detection of thermal anomalies depends on the type of activation function, duration of training data and also related parameters such as number of neurons and hidden layers in the training process. Therefore, this method does not lead to unique solutions due to differences in its initial parameters. The mentioned parameters play a major role in determining the non-linear autocorrelation structure of the non-linear time series.

Since SVMs are non-parametric methods and do not require any assumptions about the underlying model, they could be powerful tools in modeling complex phenomena, such as LST time series, for which 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. For anomaly detection, SVMs could be a good candidate because (1) they can model non-linear 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 (Thissen et al., 2003). But the training time in SVMs can be large for data sets containing many objects. 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. Mattera and Haykin (1999) have shown that in contrast to other models, SVMs perform efficiently due to the use of a non-linear kernel which to some extent reduces the effect of the noise.

It can be concluded that the proposed method in this paper gains some credibility from the overall capabilities of the four integrated methods. This can be done by voting among the mentioned methods and selecting the anomalies detected in the majority of the methods while ignoring the minor ones obtained only by some methods.

Edited by: B. D. Malamud
Reviewed by: F. Freund and one anonymous referee

## 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, 1–157, 2011.

Blackett, M., Wooster, M. J., and Malamud, B. D.: Exploring land surface temperature earthquake precursors: A focus on the Gujarat (India) earthquake of 2001, Geophys. Res. Lett., 38, L15303, doi:10.1029/2011GL048282, 2011.

Cao, L. J. and Tay, F. E. H.: Financial forecasting using support vector machines, Neural Computing & Applications, 10, 84–192, 2001.

Faraway, J. and Chatfield, C.: Times series forecasting with neural networks: a case study, Research report 95-06 of the statistics group, University of Bath, 1995.

Filizzola, C., Pergola, N., Pietrapertosa, C., and Tramutoli, V.: Robust satellite techniques for seismically active areas monitoring: a sensitivity analysis on September 7th 1999 Athens's earthquake, Phys. Chem. Earth, 29, 517–527, 2004.

Freund, F., Takeuchi, A., Lau, B. W. S., Al-Manaseer, A., Fu, C. C., Bryant, N. A., and Ouzounov, D.: Stimulated infrared emission from rocks: assessing a stress indicator, eEarth, 2, 1–10, 2007.

Freund, F., Kulahci, I. G., Cyr, G., Ling, J., Winnick, M., Tregloan-Reed, J., and Freund, M. M.: Air ionization at rock surfaces and pre-earthquake signals, J. Atmos. Solar-Terres. Phys., 71, 1824–1834, 2009.

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.

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.

Ouzounov, D. and Freund, T.: Mid-infrared emission prior to strong earthquakes analyzed remote sensing data, Adv. Space Res., 33, 268–273, 2004.

Pulinets, S. A., Ouzounov, D., Ciraolo, L., Singh, R., Cervone, G., Leyva, A., Dunajecka, M., Karelin, A. V., Boyarchuk, K. A., and Kotsarenko, A.: Thermal, atmospheric and ionospheric anomalies around the time of the Colima M7.8 earthquake of 21 January 2003, J. Ann. Geophys., 24, 835–849, 2006, http://www.ann-geophys.net/24/835/2006/.

Qiang, Z. J., Xu, X. D., and Dian, C. G.: Thermal infrared anomaly precursor of impending earthquakes, Chinese Sci. Bullet., 36, 319–323, 1991.

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.

Tronin, A. A.: Satellite thermal survey – a new tool for the studies of seismoactive regions, Int. J. Remote Sens., 17, 1439–1455, 1996.

Tronin, A. A.: Thermal IR Satellite sensor data application for earthquake research in China, Int. J. Remote Sens., 21, 3169–3177, 2000.

Zhang, G. P.: An investigation of neural networks for linear time-series forecasting, Comput. Operations Res., 28, 1183–1202, 2001.