Journal of Asian Earth Sciences 211 (2021) 104710

Contents lists available at ScienceDirect

Journal of Asian Earth Sciences

journal homepage: www.elsevier.com/locate/jseaes

Statistical framework for the evaluation of earthquake forecasting: A case study based on satellite surface temperature anomalies

Zhong-Hu Jiao *, Xinjian Shan

State Key Laboratory of Earthquake Dynamics, Institute of Geology, China Earthquake Administration, Beijing 100029, China

ARTICLE INFO

Keywords:
Global earthquakes
Earthquake precursors
Surface temperature anomalies
Atmospheric Infrared Sounder (AIRS)
Thermal infrared remote sensing
Seismic hazard

ABSTRACT

There are growing observational evidences that various geophysical anomalies precede large earthquakes. However, the reliability of these anomalies for earthquake forecasting is controversial, and therefore more consistent assessment of forecasting ability is required. A framework for investigating pre-seismic anomaly detection using essential statistical indicators before global earthquakes is proposed. Surface temperature (ST) data from the Atmospheric Infrared Sounder (AIRS) sensor were used to realize this framework. First, seismic-related ST anomalies were identified, and then the statistical characteristics of forecasting ability for three indicators (accuracy, missed detection, and false alarm) were calculated in retrospective and prospective ways. The ST anomalies displayed some aggregation effects. Negative anomalies mainly concentrated on epicenters and to the north, while positive anomalies were found on the periphery; neither were strongly dependent on earthquake magnitude. The temporal evolution of forecasting metrics was relatively stable for the period 2010–2018. Mean accuracy, missed detection, and false alarm ratios were 6.01%, 1.60%, and 92.39%, respectively. Accuracy and missed detection ratios showed some spatial correlation and both peaked in the same area (e.g., eastern Japan); however, most areas showed very high false alarm ratios. Based on our findings, the combination of AIRS ST data and the Z-score anomaly detection algorithm to predict short-term earthquakes is currently not practical; the possibility of earthquake forecasting based on satellite thermal infrared measurements remains a huge challenge. However, our results confirmed the efficiency of this framework for quantitatively evaluating earthquake forecasting ability. This approach could be applied to various geophysical parameters and anomaly detection methods.

1. Introduction

In view of growing global development and urbanization, the establishment of earthquake forecasting systems to mitigate the effects of earthquakes would be beneficial, especially for countries along tectonic plate boundaries (Bouchon et al., 2013). Although many successful cases of the prediction of large earthquakes in the intermediate or long terms have been reported (Ulukavak et al., 2020), reliable and consistent short-term earthquake forecasting remains a challenge (Ahmad et al., 2019; Sykes, 1996; Uyeda et al., 2009). Satellite remote sensing techniques potentially offer an effective approach to monitoring pre-seismic processes along active fault zones. Many retrospective statistical studies have used global earthquakes to identify statistically significant linkages between pre-seismic signals and upcoming earthquakes using a variety of anomaly analytical methods.

Over the last three decades, numerous studies have shown thermal anomalies prior to large earthquakes using satellite thermal infrared (TIR) observations; such anomalies have the potential for earthquake forecasting (Ahmad et al., 2019; Bhardwaj et al., 2017; Cicerone et al., 2009; Jiao et al., 2018; Tronin, 2010). Xiong and Shen (2017) analyzed the outgoing longwave radiation (OLR) anomalies of 3376 earthquakes using a statistical method based on the Robust Satellite technique (RST); the results demonstrate significant correlation with the earthquake preparation phase. Eleftheriou et al. (2016) used the RST approach and Robust Estimator of TIR Anomalies (RETIRA) index to identify TIR anomalies based on M $\geq$ 4 earthquakes in Greece between 2004 and 2013; significant sequences of TIR anomalies were statistically related to earthquakes and had a false positive rate of $<$ 7%.

Ulukavak et al. (2020) investigated ionospheric total electron content (TEC) anomalies based on global Mw $\geq$ 6 earthquakes from 2000 to 2019 using a 15-days moving median method; they found that numbers of positive anomalies in each magnitude group exceeded that of negative

* Corresponding author.
E-mail address: jzh@ies.ac.cn (Z.-H. Jiao).

https://doi.org/10.1016/j.jseaes.2021.104710
Received 20 August 2020; Received in revised form 21 January 2021; Accepted 6 February 2021
Available online 20 February 2021

anomalies. They conclude that TEC anomalies can be used for short-term earthquake prediction. De Santis et al. (2019) proposed a worldwide statistical correlation algorithm to examine the statistical significance of pre-seismic electron density and magnetic field anomalies from Swarm satellites. The significance is measured by 1) the exceeded level of the maximum quantity compared to a typical random maximum quantity, and 2) the standard deviation of a quantity far from the simulated ones. They confirmed earthquake related anomalies between a few days and 80 days before earthquakes; however, only ~ 40% cases showed significant anomalies. Liu and Xu (2017) observed significant TEC anomalies in 66.1% cases; they found that positive anomalies occur more often than negative ones, and that most anomalies are located south of epicenters, which is in contrast with the results of Zhu et al. (2016).

Despite the potential of these techniques for earthquake forecasting, controversy continues within the scientific community (Uyeda et al., 2009). Many studies have found no evidence to support the usability of pre-earthquake anomalous phenomena for earthquake prediction. Thermal anomalies are often deemed to occur over short temporal periods and within limited zones around epicenters (Pavlidou et al., 2018). However, the analysis of land surface temperature (LST) anomalies derived by the RST for earthquakes in Sichuan, China, from 2002 to 2018 indicates that anomalies have little correlation with impending earthquakes. Based on LST data from geostationary satellite with high temporal resolution, Pavlidou et al. (2018) inspected 20 large (Mw greater than 5.5) and shallow (<35 km) land-based earthquakes worldwide, and found no significant effect of earthquakes on LST anomalies. Thomas et al. (2017) showed that the TEC anomalies are statistically non-significant and useless for earthquake prediction. Wang and Burgmann (2019) deisgned a statistical hypothesis approach to rigorously evaluate the statistical significance of short-term gravity gradient anomalies; they found that detected changes in gravity were not statistically unique on spatial or temporal scales, and were more likely to represent time-dependent observational errors or signals related to hydrological, atmospheric, or oceanic processes.

No adequate justification is found for the synchronism or consistency of various anomalous manifestations, because all kinds of anomalous perturbations can be related to earthquake activity. Objective and quantitative assessment of the forecasting ability of these techniques is urgently needed to support and improve the reliable recognition of short-term pre-seismic anomalous features (Wang and Burgmann, 2019), which is critical for predicting the time, location, and magnitude of earthquakes. However, there is currently no consistent and quantitative way to evaluate the forecasting ability of geophysical parameters or of anomaly detection approaches with a uniform framework.

Previous studies have attempted to test possible statistical significance between pre-seismic anomalies in various geophysical parameters (Wu et al., 2012). However, the starting point of this work was how to quantify the earthquake forecasting ability of pre-seismic anomaly detection considering the current difficulties associated with these techniques. For this, quantitative assessment is required to give the maximum likelihood of forecasting future large earthquakes. Since earthquakes are statistical in nature (e.g., the Zipf-Mandelbrot law) (Abe and Suzuki, 2003; Kawamura et al., 2012; Rundle et al., 1997), we propose a statistical-based framework to systematically evaluate the forecasting ability of pre-seismic anomalies. This framework is based on the statistical significance (i.e., the higher accuracy while lower missed detection and false alarm ratios) of pre-seismic anomalies from a large dataset of global earthquakes.

The feasibility of using TIR data in earthquake prediction remains unclear (Jiao et al., 2018). Significant effort has been devoted to prove the existence of pre-earthquake thermal anomalies through different methods. The Z-score (ZS) anomaly detection method based on surface temperature (ST) data from the Atmospheric Infrared Sounder (AIRS) sensor was used; thus, a secondary objective of this study was to assess the short-term forecasting capability of ST in combination with the ZS method.

2. Data

2.1. Global earthquake catalog

Global earthquake cases were collected from the United States Geological Survey (USGS) Earthquake Hazards Program. Tectonic earthquakes occur from the Earth's surface down to ~ 700 km below the surface, among which shallow earthquakes (0–70 km) have the most devastating effects on human beings. In this study, earthquakes with magnitudes of ≥ 5 and at depths of ≤ 70 km from 2010 to 2018 were selected for analysis.

Earthquakes usually occur in close temporal and spatial proximity, and major earthquakes usually precede foreshocks with smaller magnitudes (Jordan et al., 2011). The effects are even more pronounced in seismically active areas, where earthquakes not only cluster together, but can occur within days or even hours. Therefore, it is important to consider the effects of spatially and temporally adjacent earthquakes. As such, we applied a filter procedure similar to that of Le et al. (2011) and De Santis et al. (2019). At adjoining positions ($<0.5^{\circ}$) and on the same day, only the earthquake with the largest magnitude was retained to avoid double counting. Finally, a total of 10,425 globally distributed events were retained for analysis (Fig. 1a); most are distributed along major active fault systems in various significantly different seismotectonics zones. Earthquake counts show an exponential decrease with magnitude from M5 to M9 (Fig. 1b). In terms of hypocenter depth (Fig. 1c), there are peaks of ~ 4200 and 800 cases at about 6–9 km and 35–38 km, respectively.

Earthquakes are not randomly spatially distributed (Jaumé and Sykes, 1999), and the generated earthquake dataset reflects global seismically active zones. From our dataset, earthquake counts within $1^{\circ}\times1^{\circ}$ grid-squares reach approximately 80 in the most tectonically active areas (Fig. 2). For grids containing more than one earthquake, we marked the $5^{\circ}\times5^{\circ}$ grid surrounding the pixel of the epicenter and repeated the same procedure to identify the most earthquake-prone regions of the world (gray regions in Fig. 2).

2.2. AIRS ST satellite data

AIRS is an innovative atmospheric sounder in the TIR spectrum aboard the polar-orbiting Aqua satellite. The AIRS/Aqua L3 daily standard physical retrieval (AIRS3STD) V006 product contains the standard retrieval of thermodynamic parameters, such as skin temperature, air temperature, surface emissivity, and profiles of water vapor (Susskind et al., 2014). These parameters have a temporal resolution of 12 h, and are averaged and binned into $1 \times 1^{\circ}$ grid cells on a global scale.

Nighttime ST in AIRS3STD data were used for anomaly analysis as it is a comprehensive parameter that measures the surface heat balance. The ST data represent land and sea surface temperatures and are considered a sensitive parameter for detecting pre-seismic anomalies (Jiao et al., 2018). Such anomalies result from the combination of deep tectonic activity and non-seismic factors such as solar radiation, topography, landform type, and meteorological factors. In fact, thermal energy from the deep subsurface only accounts for a small amount. AIRS3STD data from 2002 to 2018 were obtained, and the parameter dataset 'SurfSkinTemp' was extracted according to the global earthquake catalog. For each earthquake, data for 120 days before the earthquake were stored.

3. Methodology

3.1. Framework to assess earthquake forecasting ability

Despite the high complexity of the crust, earthquake occurrence has certain statistical characteristics that can be useful for prediction (Rundle et al., 2003). The key to the success of earthquake prediction is reliability, which depends on the establishment of a forecasting index

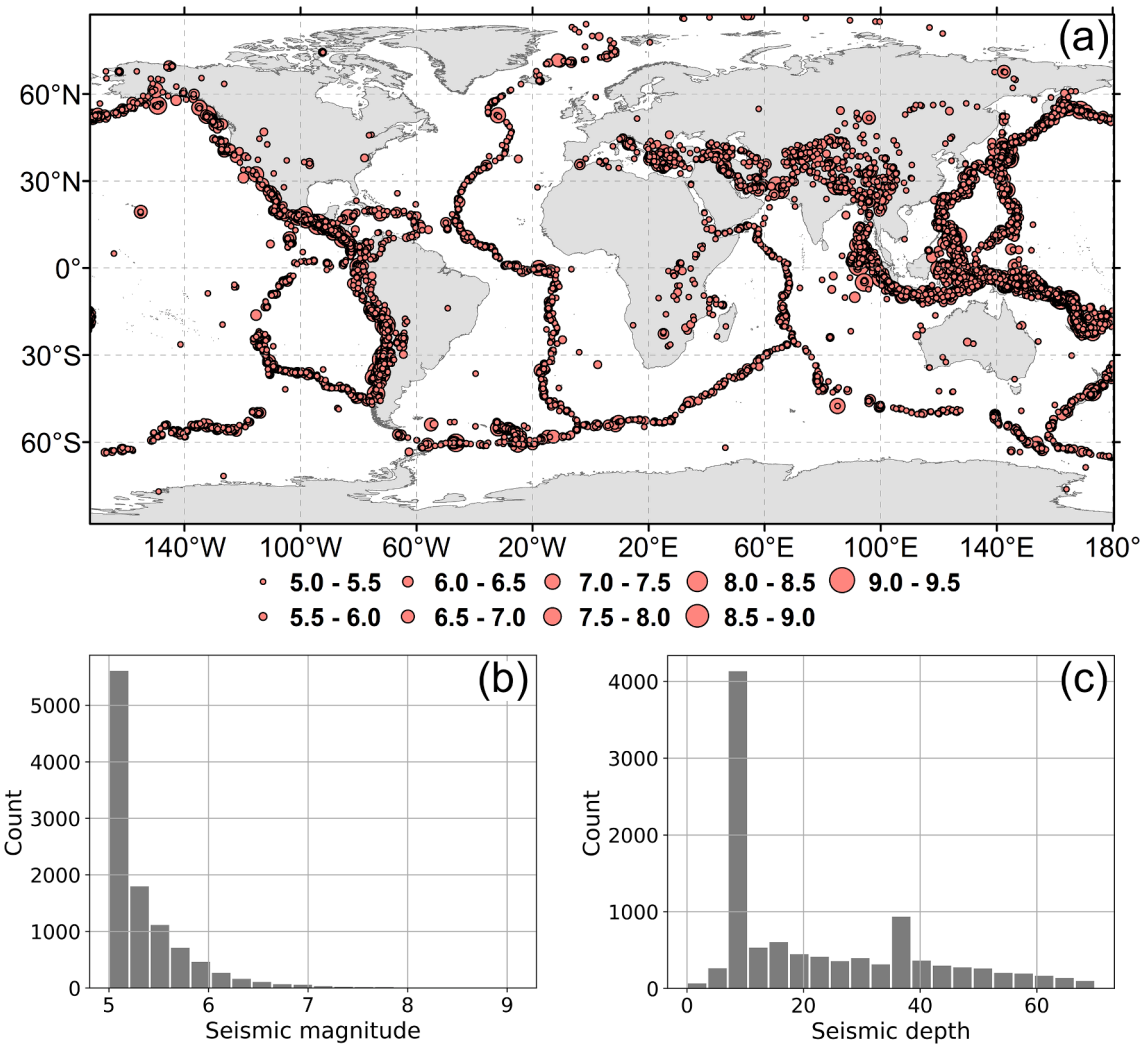

Fig. 1. Overview of global earthquakes (M ≥ 5) with shallow depths (≤70 km) from 2010 to 2018. (a) geospatial distributions of earthquakes; (b) histogram of seismic magnitude; (c) histogram of seismic depth.

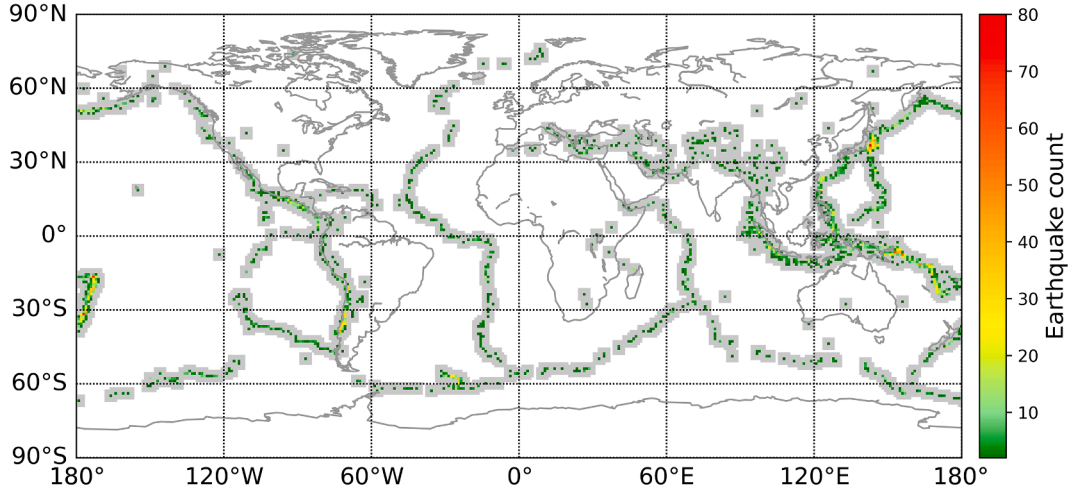

Fig. 2. Global earthquake counts summed in 1°×1° grid. Gray areas denote surrounding 5°×5° zones around the epicenters.

system composed of the following contents: (1) reliable and complete observation data; (2) a stable algorithm; (3) quantitative criteria; (4) operational forecasting rules; and (5) forecast effectiveness exceeding natural probability. To assess the earthquake forecasting ability of one geophysical parameter or anomaly detection approach, the proposed statistical framework consists of six procedures (Fig. 3).

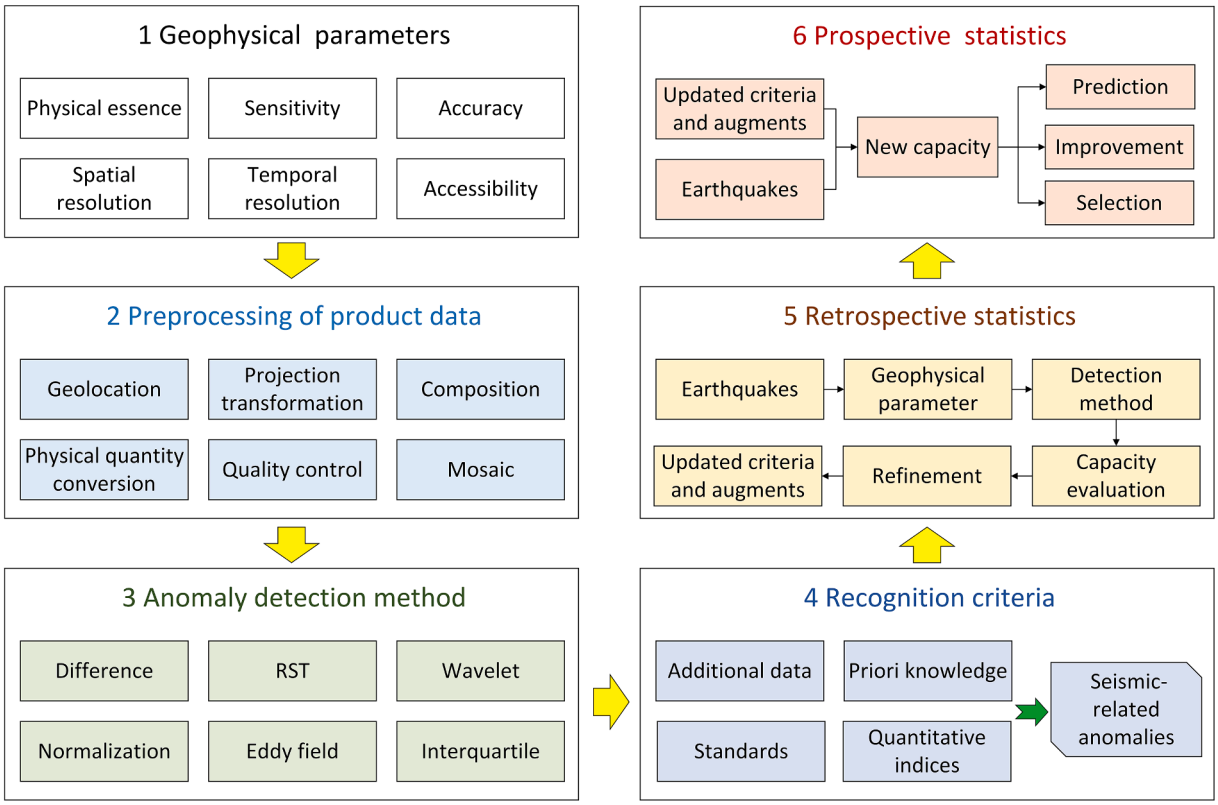

Fig. 3. Flow diagram of the proposed statistical framework for assessing the capacity of one geophysical parameter or anomaly detection method for predicting earthquakes. RST, Robust Satellite technique.

1) A geophysical parameter is selected according to its physical essence, sensitivity to seismogenic processes, retrieval accuracy from satellite observations, accessibility of data sources, and spatial and temporal resolutions. Such parameters include surface temperature, latent heat flux, air temperature, humidity, OLR, and brightness temperature of TIR measurements (Jiao et al., 2018). Appropriate parameters are the basis for recognizing pre-earthquake anomalies.

2) The preprocessing of product data is performed. These essential steps including geolocation, projection transformation, physical quantity conversion, mosaic, composition, and quality control. The estimated parameter values have different qualities, and only high-quality pixels should be retained. This procedure also includes any processing specific to anomaly analytical methods, like the normalization of neighborhood pixels (Pavlidou et al., 2018).

3) An anomaly detection method is chosen to calculate pre-seismic anomalous changes. There are a variety of methods that have been proposed to detect anomalies for seismic, volcanic, geothermal, and fire hazards. The chosen approach is one of the focuses for the assessment in this framework.

4) Criteria to recognize seismic-related anomalies is established. Anomaly detection methods can only generate anomalous values according to their mathematical formula. However, to judge whether these values are caused by earthquake activity or not requires additional data, priori knowledge, standards, and quantitative indices.

5) Global earthquake cases over a long time period are filtered based on their significant impacts. These events are used to calculate the forecasting ability based on quantitative metrics for a combination of one geophysical parameter and one anomaly detection method. The criteria or augments of the anomaly detection approach can be refined to maximize the correlation capacity in the retrospective procedure.

6) Based on the updated criteria or augments, the new forecasting ability is calculated in the prospective mode using the same earthquake cases in step 5. This capacity value is more objective and can be used to predict future possible large earthquakes and to guide the improvement of anomaly detection methods and the selection of geophysical parameters.

3.2. Anomaly detection using the ZS method

Numerous anomaly detection approaches for pre-seismic anomalies have been proposed (Jiao et al., 2018). In this study, we utilized a widely used anomaly detection procedure, the Z-score (ZS) method, to identify potentially significant pre-seismic perturbations (Ouzounov et al., 2011). Anomalies were calculated by the index through normalization via a Z-score calculated as follows:

$$\mathrm{ZS}(\mathrm{x},\mathrm{y},\mathrm{t}) = \frac{v(x,y,t) - \mu(x,y)}{\delta(x,y)}, \quad (1)$$

where $v(x,y,t)$ is the current value at position (x,y) and time t, $\mu(x,y)$ is the mean of the reference field that is calculated at the same position in the same or similar time slit of years using multi-year data, and $\delta(x,y)$ is the standard deviation of this reference field. The upper and lower boundaries of the envelope are expected to be $\mu \pm k\delta$ where k is an index number described in Section 3.3. Anomalous signals are detected if observations fall outside of the relevant upper and lower boundaries of that envelope. Moreover, the difference between the current value and the mean value of the reference field represents the signal, and the standard deviation of the reference field is the noise. The higher the signal-to-noise-ratio (SNR), the greater the possibility of thermal anomalies prior to an earthquake (Tramutoli et al., 2001).

3.3. Definition of an earthquake-related ST anomaly

The results of anomaly detection are not always connected with an earthquake. Therefore, criteria should be established to eliminate irrelevant anomalies to the extent possible (Eleftheriou et al., 2016; Qin

et al., 2013). Six criteria are proposed to recognize seismic-related ST anomalies (Table 1): anomaly intensity, count of anomaly occurrence, anomaly type, statistical period before the earthquake, grid size, and forecast period after anomalies occurred. These criteria take seismic parameters (e.g., time, epicenter, and magnitude) into account for earthquake forecasting.

The anomaly intensity was set to ≥ 2.2 for positive anomalies and ≤ -2.2 for negative anomalies which is a slightly stricter threshold than that of Thomas et al. (2017). In practical application of the ZS method, it is often assumed that a dataset has a probability distribution that approximates a normal distribution. Under this assumption, roughly 68.3%, 95.4%, and 99.7% of the values are distributed within 1, 2, and 3 standard deviations (STD) from the mean. Thus, if the index is 1, the current ST value is within a statistical bound ($\mu \pm \delta$) with a 68.3% confidence interval. The higher the index, the greater the likelihood of anomalies occurring prior to an earthquake. Then, the anomaly occurrence count is set to ≥ 2 (i.e., an anomaly that happens only once for a given pixel at any location and time is discarded)

The ZS method can calculate positive and negative anomalies. In terms of ST, a positive/negative anomaly denotes warming/cooling compared with the background mean temperature. The statistical period was set to 30, 60, 90, or 120 days before each earthquake. The spatial resolution of the AIRS3STD product is 1°; the grid size was set to 3×3 and 5×5 pixels. The statistical results in the $3^{\circ} \times 3^{\circ}$ and $5^{\circ} \times 5^{\circ}$ bins were used to improve the robustness of the statistics. The last criterion was the forecast period, which was set to 10, 20, or 30 days. When an ST anomaly exists on the current day, earthquakes within n days after and within a specific grid size were used as responsive events. The longer the forecast period, the greater the number of earthquakes included.

3.4. Statistical metrics to represent forecasting ability

Obtaining statistical indices is vital for quantitative earthquake monitoring and forecasting. Three indictors were used to represent the forecasting ability: accuracy, missed detection, and false alarm (Marchetti et al., 2020). For accuracy, when an anomaly occurs, the earthquake happened in an $n \times n$ grid where n is 1, 3 or 5. For missed detection, no anomaly appears but an earthquake occurred. For a false alarm, an anomaly occurred but there was no subsequent earthquake in an $n \times n$ grid. They are defined as follows:

$$R_A = \frac{C_A}{C_A + C_M + C_F}, \tag{2}$$

$$R_M = \frac{C_M}{C_A + C_M + C_F}, \tag{3}$$

$$R_F = \frac{C_F}{C_A + C_M + C_F}, \tag{4}$$

where R_A is the accuracy ratio, R_M is the missed detection ratio, R_F is the false alarm ratio, C_A is the accuracy count, C_M is the missed detection count, and C_F is the false alarm count. The sum of R_A, R_M, and R_F is equal to 1.

Table 1
Criteria for the recognition of earthquakes related surface temperature (ST) anomalies derived from ZS method.

No.	Criteria	Reference value
1	Anomaly intensity	≥ 2.2 for positive anomaly; ≤ -2.2 for negative anomaly
2	Count of anomaly occurrence	≥ 2
3	Anomaly type	Occurrence of positive or negative anomaly
4	Statistical period	120 days
5	Grid size	5×5
6	Forecast period	10 days

4. Statistical results

4.1. Retrospective statistics of ST anomalies

The basic assumption of retrospective anomaly analyses is to determine that an earthquake was preceded by an ST anomalous disturbance. In this study, retrospective analysis based on global earthquakes was performed based on the criteria defined in Table 1. Using data from only the 1×1 grid (i.e., 1 pixel of AIRS data) in which an earthquake was located, from 30 to 120 days before the event, the numbers of positive only, negative only, and both positive and negative anomalies all increased; overall 22.85% to 56.99% of 1×1 grid contained a positive anomaly, 39.41% to 77.88% contained a negative anomaly, and 6.86% to 42.41% contained both negative and positive anomalies (Table 2).

For the 3×3 and 5×5 grids surrounding the epicenter, a value that is equal to the count of anomaly occurrence multiplied by the mean anomaly was first calculated for each pixel. Then, the highest value for positive anomalies or lowest value for negative anomalies was chosen for that box. The trend was similar to that seen for the 1×1 grid; however, the larger zones contained higher percentages of pixels with positive only, negative only, and both positive and negative anomalies (e.g., negative anomalies increase from 39.41% to 68.95% for the 30-day period). Negative anomalies occurred more frequently than positive anomalies, which implies that negative ST anomalies are more important for earthquake recognition.

Neither positive nor negative anomalies were strongly correlated with magnitudes (Fig. 4). The mean anomaly count was ~ 2.5 for both positive and negative anomalies, while the negative anomaly count had a larger variation. The mean intensity of positive and negative anomalies was ~ 2.3 and -2.6, although the values for the 3×3 and 5×5 grid for earthquakes of M 8–9.1 were slightly higher and lower than those for the other cases, respectively. This variation is underrepresented owing to the increased error, which is caused by a decrease in the earthquake sample size with increasing earthquake magnitude (Thomas et al., 2017). In summary, the size of grid-box had little impact on the statistical results. Larger boxes contained more pixels, leading to greater anomaly count values; however, the 5×5 grid did not obviously increase the sample number as compared with the 3×3 grid. For the strongest anomaly intensities, the mean value for the 3×3 grid was the highest observed for positive anomalies, and for negative anomalies within the M6–7 magnitude range. These results demonstrate that ST anomalies aggregate around epicenters.

To further analyze the spatial characteristics of ST anomalies, anomaly counts were calculated within 5×5 grid-boxes surrounding the epicenters; in Fig. 5, the coordinate (0, 0) denotes the pixel of the epicenters. Overall, negative anomalies mainly concentrate on

Table 2
Percentage of pre-seismic positive and negative anomalies on different days within 1×1, 3×3, and 5×5 grids surrounding epicenters [a].

Day	Positive anomalies (%)	Negative anomalies (%)	Both (%)
1×1 box			
30	22.85	39.41	6.86
60	37.55	58.83	19.42
90	48.35	70.18	31.76
120	56.99	77.88	42.41
3×3 box			
30	25.94	49.05	9.98
60	47.84	72.93	30.38
90	61.08	81.05	45.95
120	68.40	85.34	55.33
5×5 box			
30	45.06	68.95	27.24
60	65.15	82.86	50.48
90	72.63	86.05	59.35
120	77.16	87.64	64.98

a. Pixel resolution = 1°

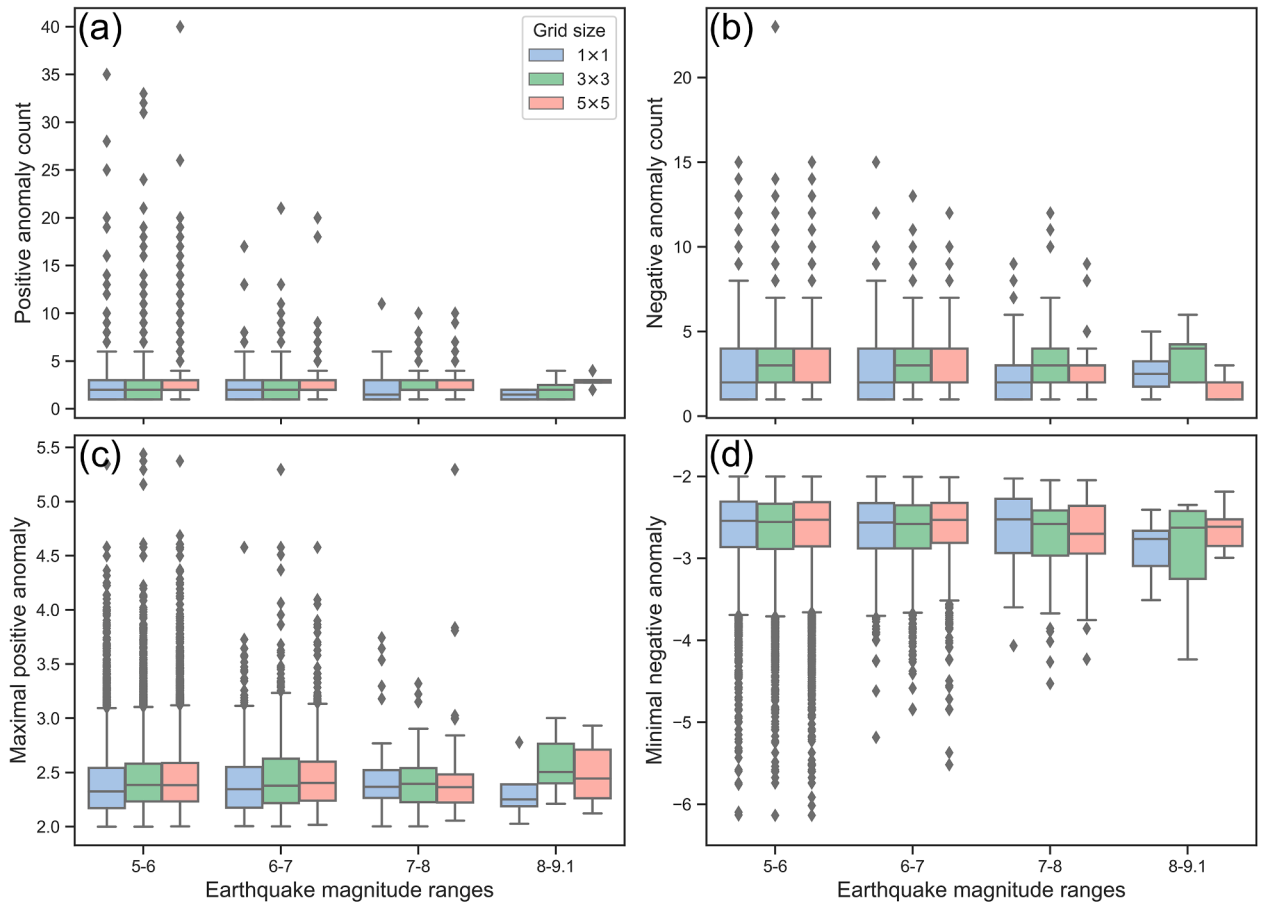

Fig. 4. Statistical box plots of ST anomalies in four magnitude range groups, and different grid sizes. (a) Positive and (b) negative anomaly counts, and (c) positive and (d) negative anomaly intensities in the 120-day period before earthquakes. The boxplot shows the three quartile values (lower, middle, and upper quartiles) of the distribution along with extreme values, the whiskers extend to show a 1.5 times interquartile range, and the rest of points are determined to be outliers that fall outside this range.

epicenters and to their north, while positive anomalies are found on the periphery. More importantly, the positive and negative anomalies are spatially complementary, analogous to the spatial distribution of ionospheric TEC (Zhu et al., 2016). The maximum count of negative anomalies is 3224, and that of positive anomalies is 1623 for the 60-day pre-earthquake periods, respectively. Fig. 5b shows the mean anomaly magnitudes for each grid corresponding to the anomaly counts (Fig. 5a). The negative anomalies show the analogical trends for both the count and mean magnitudes, but the positive anomalies have larger deviation. The relatively lower positive anomaly magnitudes mainly are in the south, and larger negative anomaly magnitudes are at the epicenter. The magnitude values suffixed with two asterisks indicate that the average anomaly magnitudes are statistically significant different from the value at the central pixel ($p < 0.05$) by the statistical *t*-test. The negative anomaly magnitude values at the surrounding pixels are very disparate with epicenter pixel, while only the south grids are significantly different from the epicenter grid for the positive anomaly magnitudes. Taken together, the results indicate a high degree of nonrandom coincident spatial distribution of strong earthquake epicenters and related positive and negative anomalies.

Furthermore, the earthquakes were divided into three classes according to their distances to the coastline: land, shore, and ocean events. The different spatial characteristics of positive and negative ST anomaly counts are shown in Fig. 6. Because the majority is the shore earthquakes, they show the similar features as shown in Fig. 5. For land earthquakes which have the minimal count, the more negative anomaly counts dominate the north, while the large positive anomaly counts approximately locate in the surrounding area. For ocean earthquakes, the more negative anomaly counts are mainly in the northwest parts, and the large positive anomaly counts can occur almost any grid. Due to the 1° spatial resolution of AIRS L3 product, different land covers are one of the reasons to explain these diversities.

The spatial statistical characteristics of earthquakes were also analyzed in different latitudinal zones (Fig. 7), and a latitude-dependent phenomenon can be found, and partly explains the cause of anomaly count distributions shown in Fig. 6. In the North Frigid Zone and North Temperate Zone, the large negative anomaly counts dominate the north and northeast, and the large positive anomaly counts most occurr in the northeast. The similar phenomenon can be found in the Torrid Zone for the negative anomaly, however more positive anomaly count was in the south because more earthquakes happened in the North Hemisphere contributes to this distribution. In the South Temperate Zone, the larger negative anomaly counts turn to the southeast, and the east grids have more positive anomaly counts. The spatial patters for both positive and negative anomaly counts are symmetrical in the North and South Temperate Zones. The same characteristic is also presented in the North and South Frigid Zones. However, only few earthquakes happened in the South Frigid Zone, therefore the statistics is not stable.

The impact of season factor on the positive and negative anomaly count was also considered, and shown in Fig. 8. There is no significant seasonal difference compared with the latitudinal effect (Fig. 7). Different seasons have similar spatial characteristics as the overall results presented in Fig. 5. The large negative anomaly counts are mainly located in the epicenter and its north, while the large positive anomaly count embrace the epicenter. Moreover, we divided the observation time with an interval of 30 days, and analyzed the differences between

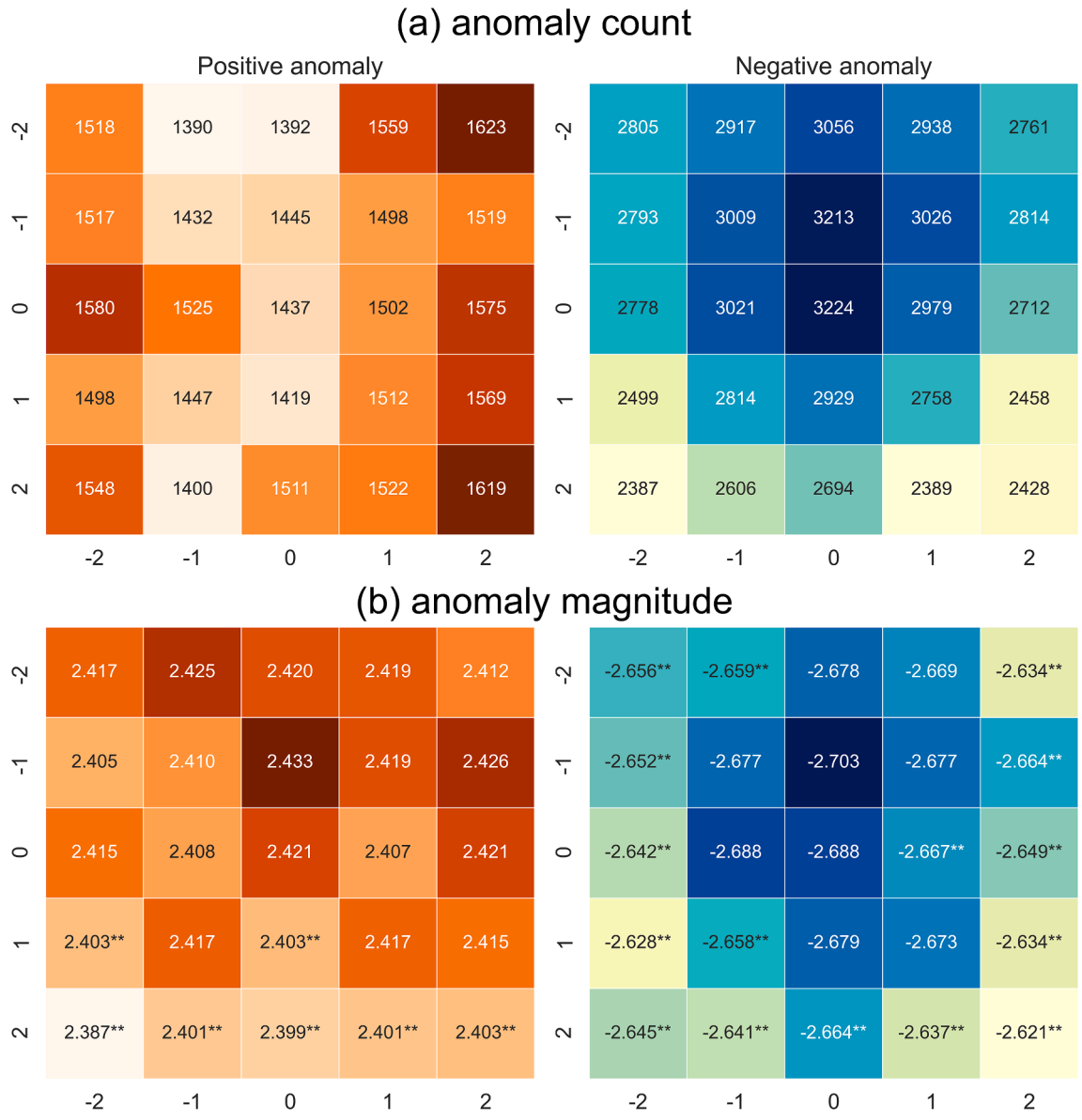

Fig. 5. (a) Positive and negative anomaly counts and (b) mean anomaly magnitude within 5 × 5 grid in 60-day pre-seismic periods. Two asterisks indicate that the average anomaly magnitudes are statistically significant different from the value at the central pixel ($p < 0.05$). The spatial resolution of the pixel is 1°.

different temporal spans. Again, the temporal factor cannot impose large influence on the spatial distribution of anomaly counts, and the anomaly count is relatively even for each span (Fig. 9). From above analysis, it is concluded that the spatial position is the more important factor than the temporal changes to impact the spatial characteristics of pre-earthquake anomaly counts.

4.2. Sensitivity analysis of recognition criteria

A prospective forecasting scenario using earthquakes in 2018 was established to inspect the forecasting ability of ST anomalies based on three indices: accuracy, missed detection, and false alarm ratios. A set of parameters (Table 1) was used as the baseline to recognize ST anomalies as earthquake-related ST anomalies (ASTA), and only one parameter one time was changed to show its impact on the anomaly indicators. The default values of these parameters were chosen according to dozens of experiments in current work. An anomaly within a grid was recognized when more than half of the pixels in that grid were anomalous (e.g., more than 12 pixels in a 5 × 5 grid). The interactive effects of spatially and temporally adjacent earthquakes was considered; when an earthquake happened on a given day, the ST anomalies within 5 × 5 pixels around the epicenter for n days before this day (where n is the value of the period in Table 1; e.g., 120 days) were deemed invalid in order to eliminate the influence of an earthquake event.

Parameters with customized values were compared objectively to explore the best possible solution. Table 3 shows the results for criteria parameters with different values. Case 2 (anomalous value ≥ 2) increased accuracy ratio and decreased missed detection ratio; however, the false alarm ratio rose significantly compared with the baseline (Case 1); Case 4 (count of anomaly occurrence ≥ 1) had analogous changes. In Case 5 (count of anomaly occurrence ≥ 3), missed detection ratio increased by 143.36%. In Case 8 (both positive and negative anomalies), missed detection ratio increased by 326.15%, whereas accuracy and false alarm ratios were improved compared with cases 6 and 7. As the statistical period was increased, accuracy ratio declined rapidly, while missed detection ratio increased from 54.58% to 305.16%. Contrary to our expectations, the false alarm ratio gradually fell. In Case 12 (grid size: 3 × 3), the accuracy ratio fell, while missed detection and false alarm ratios increased; compared with the 5 × 5 grid, this gave a poorer result. Increasing the forecast time from 10 to 30 days significantly

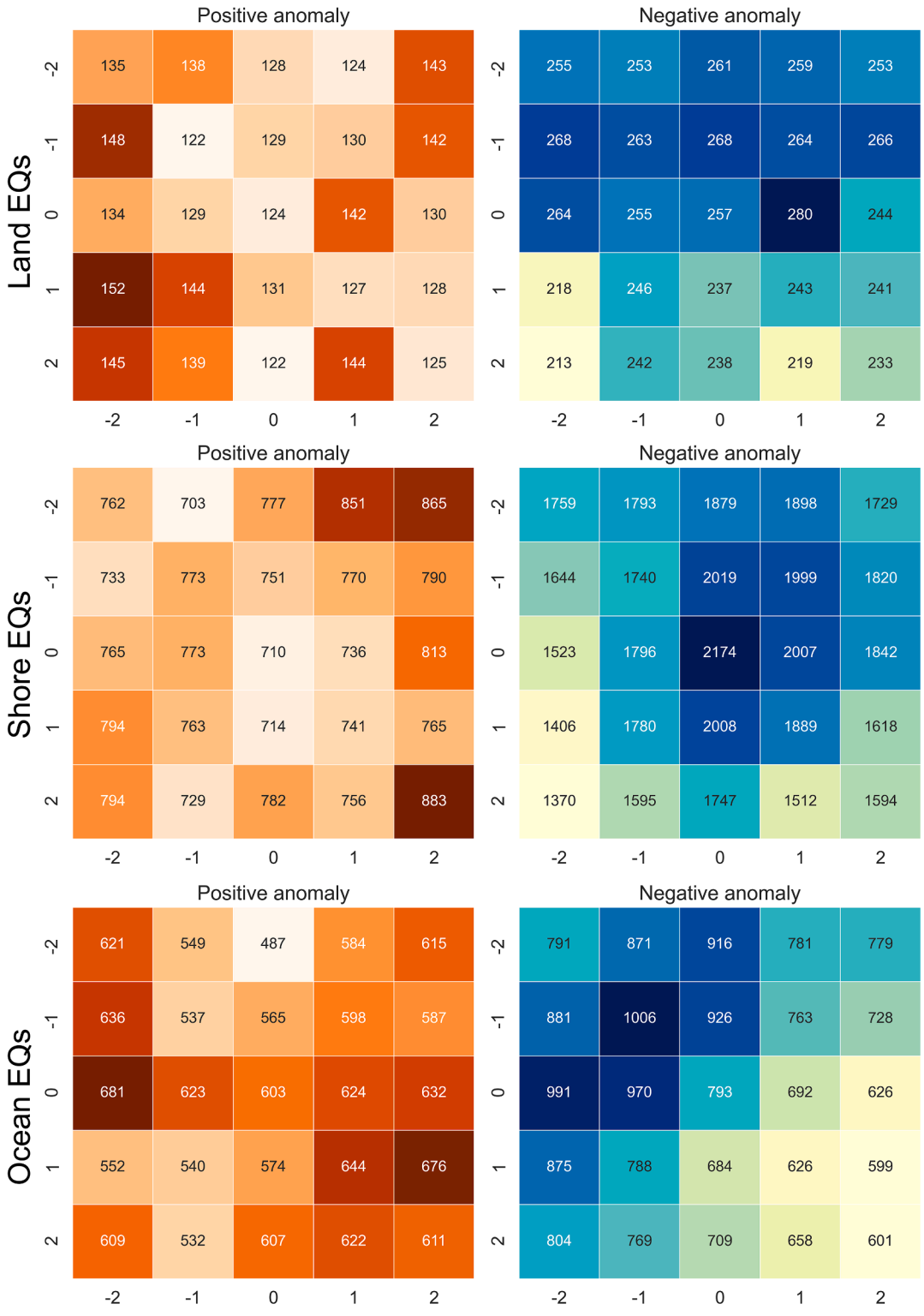

Fig. 6. Positive and negative anomaly counts within 5 × 5 grids within 60-day pre-seismic periods for land, shore, and ocean earthquakes (EQs).

increased the accuracy ratio, while missed detection ratio rapidly increased and false alarm ratio showed only minor changes. The results confirm that different parameters have important effects on forecasting ability; however, no single parameter dominates. Simultaneously increasing accuracy ratio while decreasing missed detection and false alarm ratios remains a challenge.

4.3. *Prospective statistics in seismically active regions*

The purpose of prospective statistics is to verify the forecasting ability of current techniques for future earthquakes. The above analyses provide insights into the criteria of recognition of ASTA using six parameters. Therefore, the parameters in Table 1 were utilized based on all of the global earthquakes from 2010 to 2018 to explore the spatial–temporal evolution of three indicators (accuracy, missed detection,

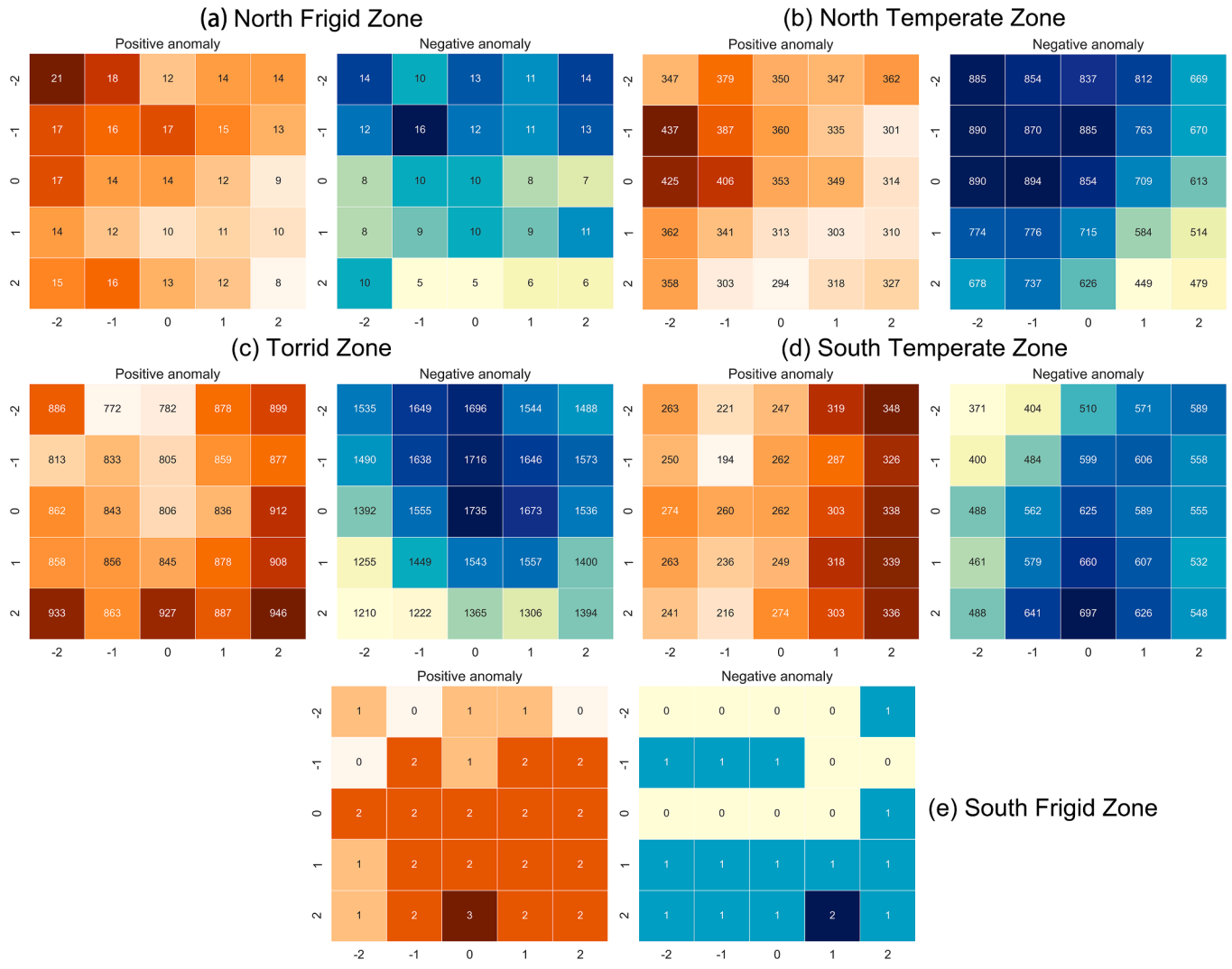

Fig. 7. Positive and negative anomaly counts within 5 × 5 grid within 60-day pre-seismic periods for earthquakes (EQs) in the five latitudinal zones, including (a) North Frigid Zone (NF, 90°N – 66.56°N), (b) North Temperate Zone (NT, 66.56°N – 23.44°), (c) Torrid Zone (TZ, 23.44°N – 23.44°S), (d) South Temperate Zone (ST, 23.44°S – 66.56°S), and (e) South Frigid Zone (SF, 66.56°S – 90°S).

and false alarm ratios) in 5 × 5 grid. To reduce significant false alarm count, only global seismically active regions (gray areas in Fig. 2) were considered. The three indicators showed good stability from 2010 to 2018 (Fig. 10a) with very low standard deviations and smooth trends after 360-day moving average. The mean accuracy, missed detection, and false alarm ratios were 6.01%, 1.60%, and 92.39%; standard deviations were 1.25%, 0.91%, and 1.70%, respectively.

The spatial characteristics of the forecasting metrics are shown in Fig. 10. The ranges of the accuracy, missed detection, and false alarm ratios were 0%–25%, 0%–6.2%, and 68%–100%, respectively. Spatially, accuracy and missed detection ratios were positively correlated to some extent; they both reached peak values in eastern Japan. As such, they become higher or lower at the same area. This means that it is difficult to find some approaches to increase accuracy ratio meanwhile decrease missed detection ratio. Other areas, like Coral Sea, west coast of South America, and the South Pacific Ocean, also show higher accuracy and missed detection ratios. However, most areas had high false alarm ratios (Fig. 10c). Such large uncertainties confirm that at present, this approach is not suitable for short-term earthquake forecasting (Uyeda et al., 2009).

5. Discussion

The primary goal of this work was to develop a framework to assess the forecasting ability of pre-earthquake anomaly detection. The framework was tested using global ST anomaly data, and the results confirm that the framework developed is appropriate and reliable. For successful earthquake prediction, the time, location, and magnitude of an earthquake must be determined (Jordan et al., 2011). However, we found that earthquake magnitude had little correlation with ST anomaly intensity or count. Furthermore, high false alarm ratios decreased the missed detection and accuracy ratios because ST anomalies can occur anywhere and spatial patterns constantly change with time as the results of gravity gradient anomalies (Wang and Burgmann, 2019). Additionally, anomalies in non-seismically active zones cannot be interpreted properly owing to the complex driving forces, such as topography, land cover, meteorology, and anthropogenic activities. These issues need to be addressed in future studies before their application in earthquake forecasting.

The use of remotely sensed observations to study pre-seismic anomalies on a global scale is a huge challenge. Processing large volumes of remote sensing data requires significant storage and computing resources. Therefore, we selected the AIRS's Daily Level 3 product; however, this product has gaps between satellite paths and issues with

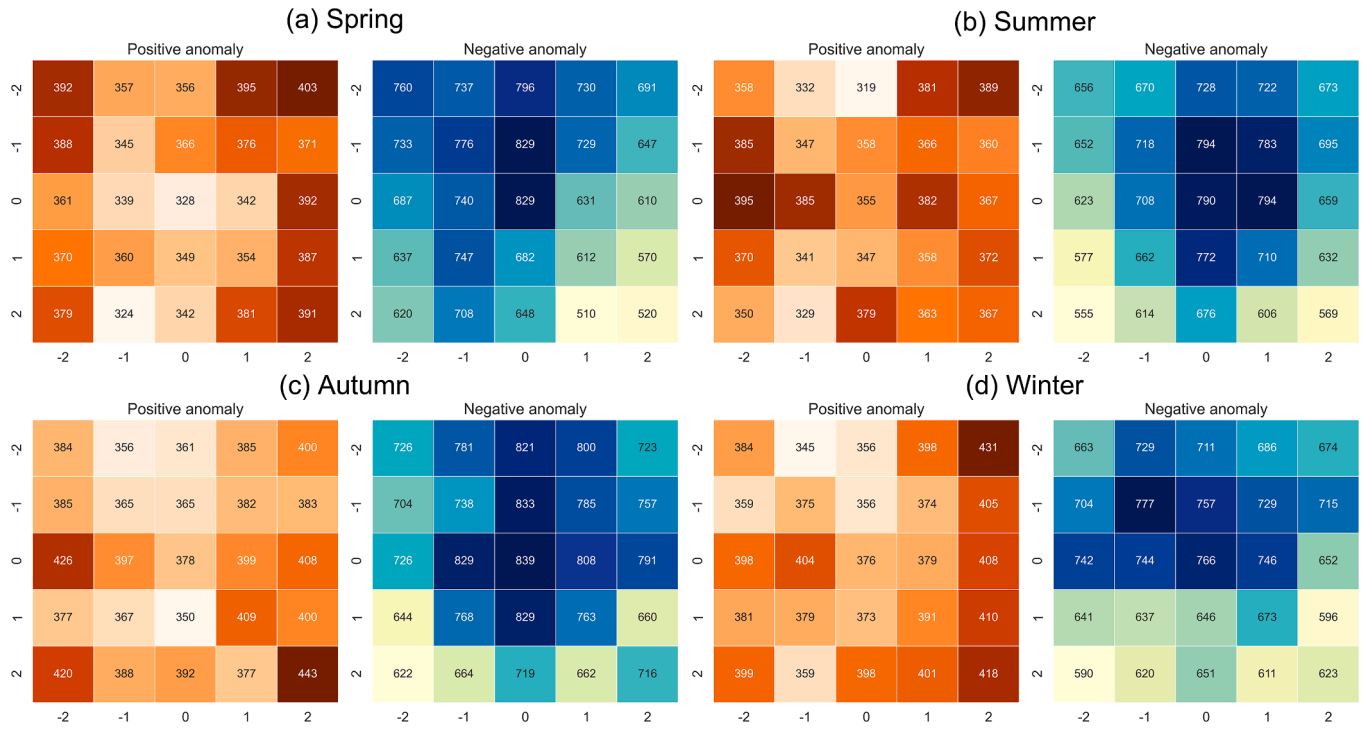

Fig. 8. Positive and negative anomaly counts within 5 × 5 grid within 60-day pre-seismic periods for earthquakes (EQs) in different seasons: (a) Sprint, (b) Summer, (c) Autumn, and (d) Winter. For the North Hemisphere, the Spring includes March, April, and May; the Summer include June, July, and August; the Autumn includes, September, October, and November; and the Winter includes December, January, and February. For the South Hemisphere, the Spring includes September, October, and November; the Summer include December, January, and February; the Autumn includes, March, April, and May; and the Winter includes June, July, and August. The earthquakes in the same season for both North and South Hemispheres were included.

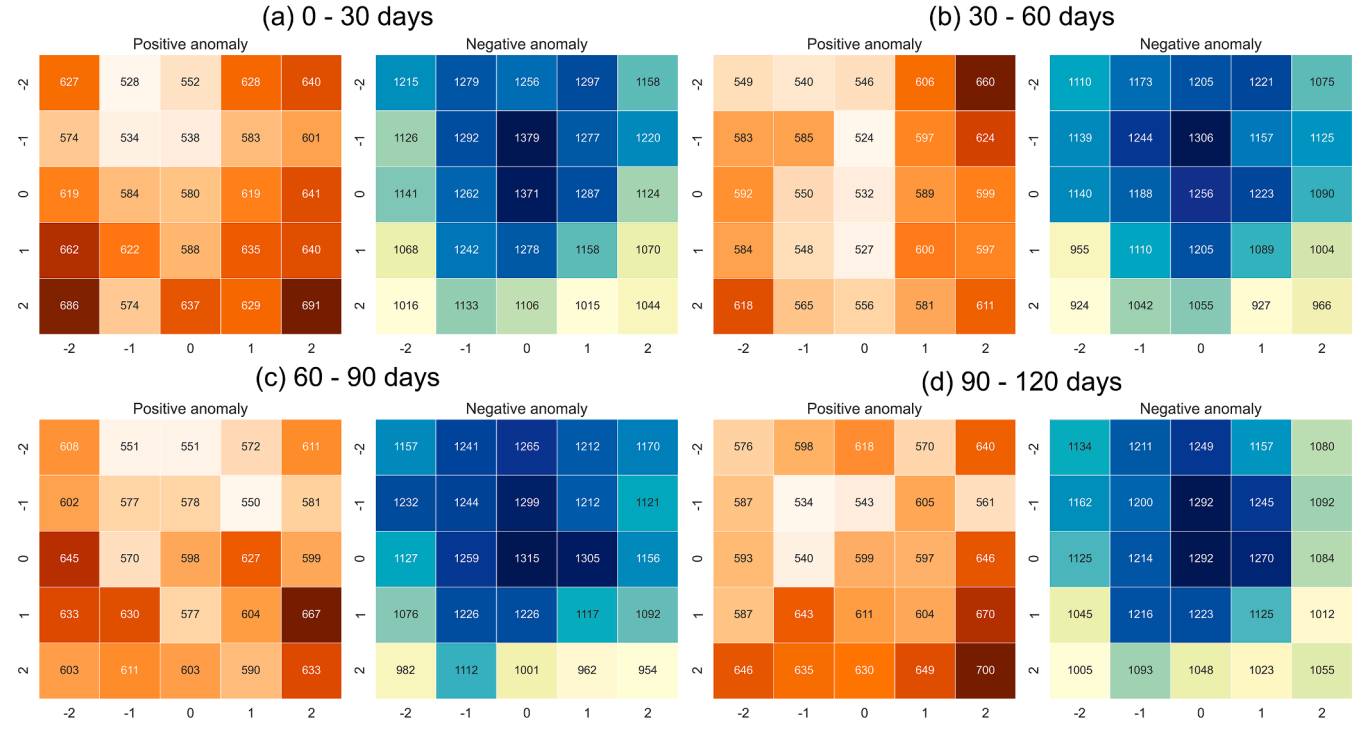

Fig. 9. Positive and negative anomaly counts within 5 × 5 grid within (a) 0–30 days, (b) 30–60 days, (c) 60–90 days, and (d) 90–120 days before the earthquakes.

cloud contamination. Such incomplete data limit the spatiotemporal analysis of ST anomalies for the purpose of earthquake prediction. Data without these gaps and with high spatiotemporal resolutions are needed in future studies.

However, regardless of the remotely sensed dataset used, the difficulty of earthquake forecasting is mainly due to the complexity of earthquake preparation processes (Freund, 2011; Ma et al., 2014). Tectonic activity may be manifested as long-term slow creep or short-

Table 3
Statistical metrics for different criteria parameters [a].

Case	Parameter	Accuracy count	Missed detection count	False alarm count
1	Baseline	9208	2772	525,014
2	Anomalous value ≥ 2	10,081 (9.48%)	1899 (−31.49%)	615,609 (17.26%)
3	Anomalous value ≥ 2.5	6290 (−31.69%)	5690 (105.27%)	299,052 (−43.04%)
4	Count of anomaly occurrence ≥ 1	10,605 (15.17%)	1375 (−50.40%)	674,016 (28.38%)
5	Count of anomaly occurrence ≥ 3	5234 (−43.16%)	6746 (143.36%)	292,254 (−44.33%)
6	Only positive anomalies	2434 (−73.57%)	9546 (244.37%)	211,694 (−59.68%)
7	Only negative anomalies	7217 (−21.62%)	4763 (71.83%)	337,882 (−35.64%)
8	Both positive and negative anomalies	167 (−98.19%)	11,813 (326.15%)	18,285 (−96.52%)
9	Statistics period: 90 days	7695 (−16.43%)	4285 (54.58%)	408,244 (−22.24%)
10	Statistics period: 60 days	3792 (−58.82%)	8188 (195.38%)	235,549 (−55.13%)
11	Statistics period: 30 days	749 (−91.87%)	11,231 (305.16%)	71,897 (−86.31%)
12	Grid size: 3 × 3	8915 (−3.18%)	3065 (10.57%)	1,326,303 (152.62%)
13	Forecast: 20 days	16,498 (79.17%)	6120 (120.78%)	523,642 (−0.26%)
14	Forecast: 30 days	23,110 (150.98%)	9817 (254.15%)	522,322 (−0.51%)

a. Values in brackets are the percentage difference relative to the baseline (Case 1).

term sudden displacement. This tectonic activity is accompanied by the movement of matter and energy, which inevitably changes the state of thermal radiation at the surface (Chen et al., 2015). There are significant thermal anomalies in the areas where earthquakes occur, but there are also thermal anomalies in areas where there are no earthquakes. Even along faults and at fault intersections there are areas without earthquakes; however, these areas play a role in the occurrence and development of earthquakes. Earthquake genesis accompanied by a thermal migration process (that is, an energy migration process); however, the energy released by the final earthquake is spatially concentrated.

To recognize subtle anomalous signals more reliably, the integration of more parameters related to different forms of energy release underground is needed. Since the Earth's atmospheric and ionospheric systems are related, anomalous changes in different geophysical parameters are not isolated (Marchetti et al., 2019). The development of multi-parameter methods to recognize anomalous distributions in the Earth's system presents a more promising approach than looking for individual precursors (De Santis et al., 2019). The ingestion of data from multiple observation sources with reliable algorithms would also contribute to our understanding of pre-seismic processes. The framework proposed in this study would allow for the assessment of such an approach in terms of short-term earthquake forecasting.

6. Conclusions

There has been much interest in the scientific community regarding the use of geophysical anomalies to forecast earthquakes. This study developed and implemented a statistical framework to assess earthquake forecasting ability; this framework is suitable for use with various geophysical parameters and anomaly detection methods, making consistent and quantitative assessment feasible.

ST from AIRS data was critically investigated to analyze the statistical characteristics of global moderate–major earthquake from 2010 to 2018. The ST anomalies showed some aggregation effects around epicenters; negative anomalies were found to be more important than positive anomalies in terms of earthquake recognition. However, neither positive nor negative anomalies significantly correlate with earthquake magnitude. Further analysis of the spatial characteristics of ST anomalies indicated that negative anomalies mainly concentrate at the epicenter and to its north; positive anomalies occur on the periphery. The positive and negative anomalies are spatially complementary. There is no significant temporal difference compared with the latitudinal effect for pre-earthquake anomaly counts, therefore the spatial position is a more important factor than the temporal changes to influence these

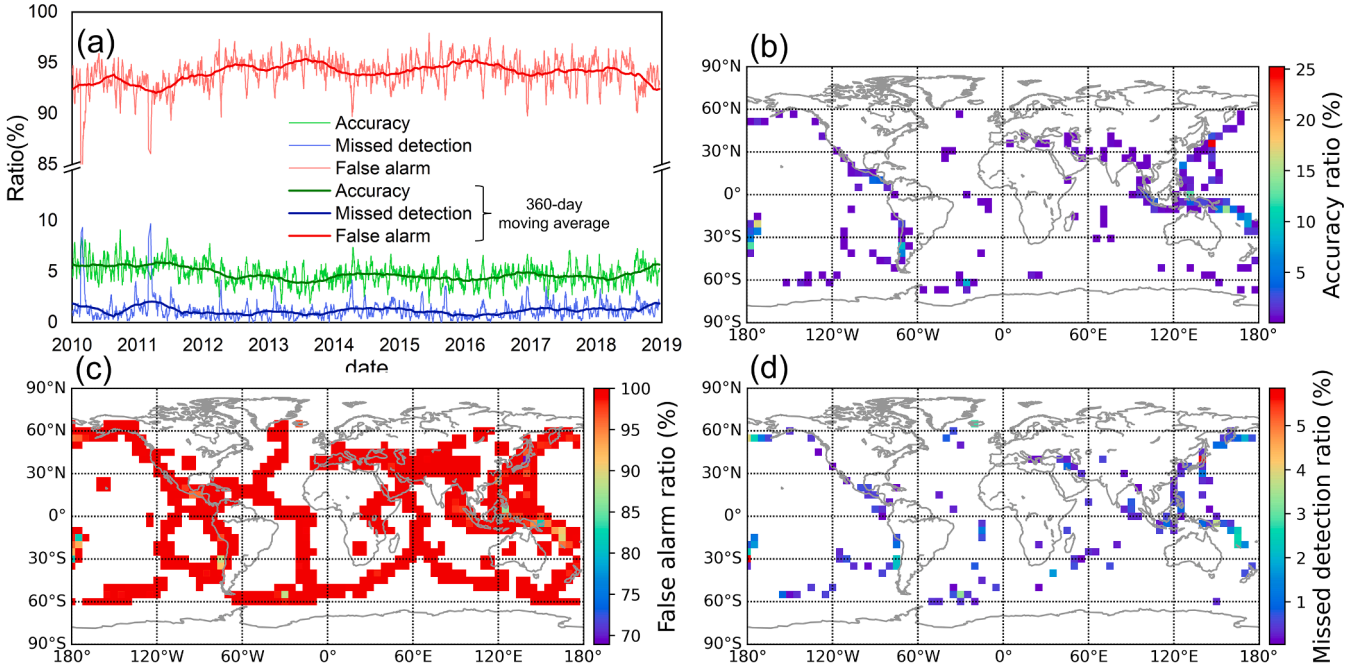

Fig. 10. Statistical characteristics of accuracy, missed detection, and false alarm ratios (as a percentage of the total). (a) Time series of three indicators each day from 2010 to 2018. Spatial distributions of (b) accuracy, and (c) false alarm, and (d) missed detection ratios for 5 × 5 grid in the seismically active regions worldwide.

spatial characteristics.

Different parameters have varying effects on forecasting ability; however, no parameter dominates and increasing accuracy ratio while decreasing missed detection and false alarm ratios remains a challenge. The temporal evolution of forecasting metrics is stable. The mean values for accuracy, missed detection, and false alarm ratios were found to be 6.01%, 1.60%, and 92.39%, respectively. To some extent, accuracy and missed detection ratios are spatially positively correlated; both show peak values in eastern Japan. Nevertheless, most areas illustrate high false alarm ratios. Based on our findings, currently, this approach for forecasting large earthquakes is not practical.

When using satellite TIR remote sensing techniques, new observations from ground and space need to be integrated into our framework to improve the statistical significance, repeatability, and universality; this will be a focus of future research. The results of this study demonstrate the strengths of the proposed statistical framework. We believe that this framework will facilitate the application of multiple data sources and anomaly detection methods to improve current knowledge about pre-seismic anomalous manifestations, and the potential establishment of earthquake forecasting systems.

CRediT authorship contribution statement

Zhong-Hu Jiao: Conceptualization, Methodology, Writing - original draft. **Xinjian Shan:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Key Research and Development Program of China [grant numbers 2018YFC1503602 and 2019YFC1509202]; China Seismic Experimental Site Project [grant number 2018CSES0205]. We thank Goddard Earth Sciences Data and Information Services Center (GES DISC) for providing AIRS AIRS3STD data (https://disc.gsfc.nasa.gov/datasets /AIRS3STD_006/summary), and United States Geological Survey (USGS) for providing earthquake catalog data (https://www.usgs.gov/natural-hazards/earthquake-hazards).

References

Abe, S., Suzuki, N., 2003. Law for the distance between successive earthquakes. Journal of Geophysical Research: Solid Earth 108, 2113.

Ahmad, N., Barkat, A., Ali, A., Sultan, M., Rasul, K., Iqbal, Z., Iqbal, T., 2019. Investigation of Spatio-temporal Satellite Thermal IR Anomalies Associated with the Awaran Earthquake (Sep 24, 2013; M 7.7). Pakistan. Pure and Applied Geophysics 176, 3533–3544.

Bhardwaj, A., Singh, S., Sam, L., Joshi, P.K., Bhardwaj, A., Martín-Torres, F.J., Kumar, R., 2017. A review on remotely sensed land surface temperature anomaly as an earthquake precursor. International Journal of Applied Earth Observation and Geoinformation 63, 158–166.

Bouchon, M., Durand, V., Marsan, D., Karabulut, H., Schmittbuhl, J., 2013. The long precursory phase of most large interplate earthquakes. Nature Geoscience 6, 299–302.

Chen, S., Liu, P., Guo, Y., Liu, L., Ma, J., 2015. An experiment on temperature variations in sandstone during biaxial loading. Physics and Chemistry of the Earth, Parts A/B/C 85–86, 3–8.

Cicerone, R.D., Ebel, J.E., Britton, J., 2009. A systematic compilation of earthquake precursors. Tectonophysics 476, 371–396.

De Santis, A., Marchetti, D., Pavon-Carrasco, F.J., Cianchini, G., Perrone, L., Abbattista, C., Alfonsi, L., Amoruso, L., Campuzano, S.A., Carbone, M., Cesaroni, C., De Franceschi, G., De Santis, A., Di Giovambattista, R., Ippolito, A., Piscini, A., Sabbagh, D., Soldani, M., Santoro, F., Spogli, L., Haagmans, R., 2019. Precursory worldwide signatures of earthquake occurrences on Swarm satellite data. Scientific Reports 9, 20287.

Eleftheriou, A., Filizzola, C., Genzano, N., Lacava, T., Lisi, M., Paciello, R., Pergola, N., Vallianatos, F., Tramutoli, V., 2016. Long-term RST analysis of anomalous TIR sequences in relation with earthquakes occurred in Greece in the period 2004–2013. Pure and Applied Geophysics 173, 285–303.

Freund, F., 2011. Pre-earthquake signals: Underlying physical processes. Journal of Asian Earth Sciences 41, 383–400.

Jaumé, S.C., Sykes, L.R., 1999. Evolving Towards a Critical Point: A Review of Accelerating Seismic Moment/Energy Release Prior to Large and Great Earthquakes. In: Wyss, M., Shimazaki, K., Ito, A. (Eds.), Seismicity Patterns, their Statistical Significance and Physical Meaning. Birkhäuser Basel, Basel, pp. 279–305.

Jiao, Z.-H., Zhao, J., Shan, X., 2018. Pre-seismic anomalies from optical satellite observations: a review. Natural Hazards and Earth System Sciences 18, 1013–1036.

Jordan, T.H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K., Zschau, J., 2011. Operational earthquake forecasting: State of Knowledge and Guidelines for Utilization. Annals of Geophysics 54, 316–391.

Kawamura, H., Hatano, T., Kato, N., Biswas, S., Chakrabarti, B.K., 2012. Statistical physics of fracture, friction, and earthquakes. Reviews of Modern Physics 84, 839–884.

Le, H., Liu, J.Y., Liu, L., 2011. A statistical analysis of ionospheric anomalies before 736 M6.0+ earthquakes during 2002–2010. Journal of Geophysical Research: Space Physics 116, A02303.

Liu, W., Xu, L., 2017. Statistical analysis of ionospheric TEC anomalies before global Mw $\geq$ 7.0 earthquakes using data of CODE GIM. Journal of Seismology 21, 759–775.

Ma, J., Guo, Y., Sherman, S.I., 2014. Accelerated synergism along a fault: A possible indicator for an impending major earthquake. Geodynamics & Tectonophysics 5, 387–399.

Marchetti, D., De Santis, A., D'Arcangelo, S., Poggio, F., Piscini, A., A. Campuzano, S., De Carvalho, W.V.J.O., 2019. Pre-earthquake chain processes detected from ground to satellite altitude in preparation of the 2016–2017 seismic sequence in Central Italy. Remote Sensing of Environment 229, 93-99.

Ouzounov, D., Pulinets, S., Romanov, A., Romanov, A., Tsybulya, K., Davidenko, D., Kafatos, M., Taylor, P., 2011. Atmosphere-ionosphere response to the M9 Tohoku earthquake revealed by multi-instrument space-borne and ground observations: Preliminary results. Earthquake Science 24, 557–564.

Pavlidou, E., van der Meijde, M., van der Werff, H., Hecker, C., 2018. Time Series Analysis of Land Surface Temperatures in 20 Earthquake Cases Worldwide. Remote Sensing 11, 61.

Qin, K., Wu, L., Zheng, S., Liu, S., 2013. A Deviation-Time-Space-Thermal (DTS-T) Method for Global Earth Observation System of Systems (GEOSS)-Based Earthquake Anomaly Recognition: Criterions and Quantify Indices. Remote Sensing 5, 5143–5151.

Rundle, J.B., Gross, S., Klein, W., Ferguson, C., Turcotte, D.L., 1997. The statistical mechanics of earthquakes. Tectonophysics 277, 147–164.

Rundle, J.B., Turcotte, D.L., Shcherbakov, R., Klein, W., Sammis, C., 2003. Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. Reviews of Geophysics 41, 1019.

Susskind, J., Blaisdell, J.M., Iredell, L., 2014. Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. Journal of Applied Remote Sensing 8, 084994.

Sykes, L.R., 1996. Intermediate- and long-term earthquake prediction. Proceedings of the National Academy of Sciences 93, 3732.

Thomas, J.N., Huard, J., Masci, F., 2017. A statistical study of global ionospheric map total electron content changes prior to occurrences of M $\geq$ 6.0 earthquakes during 2000–2014. Journal of Geophysical Research: Space Physics 122, 2151–2161.

Tramutoli, V., Di Bello, G., Pergola, N., Piscitelli, S., 2001. Robust satellite techniques for remote sensing of seismically active areas. Annals of Geophysics 44, 295–312.

Tronin, A.A., 2010. Satellite remote sensing in seismology. A review. Remote Sensing 2, 124–150.

Ulukavak, M., Yalçınkaya, M., Kayıkçı, E.T., Öztürk, S., Kandemir, R., Karslı, H., 2020. Analysis of ionospheric TEC anomalies for global earthquakes during 2000–2019 with respect to earthquake magnitude (Mw≥6.0). Journal of Geodynamics 135, 101721.

Uyeda, S., Nagao, T., Kamogawa, M., 2009. Short-term earthquake prediction: Current status of seismo-electromagnetics. Tectonophysics 470, 205–213.

Wang, L., Burgmann, R., 2019. Statistical Significance of Precursory Gravity Changes Before the 2011 Mw 9.0 Tohoku-Oki Earthquake. Geophysical Research Letters 46, 7323–7332.

Wu, L.X., Qin, K., Liu, S.J., 2012. GEOSS-based thermal parameters analysis for earthquake anomaly recognition. Proceedings of the IEEE 100, 2891–2907.

Xiong, P., Shen, X., 2017. Outgoing longwave radiation anomalies analysis associated with different types of seismic activity. Advances in Space Research 59, 1408–1415.

Zhu, F., Lin, J., Su, F., Zhou, Y., 2016. A spatial analysis of the ionospheric TEC anomalies prior to M7.0+ earthquakes during 2003–2014. Advances in Space Research 58, 1732–1738.